Volume 16, Number 2

Print ISSN: 1524-7252
Online ISSN: 1532-5806

ACADEMY OF INFORMATION AND MANAGEMENT SCIENCES JOURNAL

The *Academy of Information and Management Sciences Journal* is owned and published by Jordan Whitney Enterprises, Inc.. Editorial content is under the control of the Allied Academies, Inc., a non-profit association of scholars, whose purpose is to support and encourage research and the sharing and exchange of ideas and insights throughout the world.

The *Academy of Information and Management Sciences Journal* is owned and published by Jordan Whitney Enterprises, Inc. 51 Blake Drive,, Arden, NC 28704, USA. Those interested in communicating with the *Journal,* should contact the Executive Director of the Allied Academies at info@.alliedacademies.org.

EDITORIAL BOARD MEMBERS

EDITORIAL BOARD MEMBERS

EDITORIAL BOARD MEMBERS

ACADEMY OF INFORMATION AND MANAGEMENT SCIENCES JOURNAL

TABLE OF CONTENTS

EDITORIAL BOARD MEMBERS III

LETTER FROM THE EDITORS VIII

BENCHMARKING: AUTOMOTIVE FACILITY LAYOUT, JPH, AND OVER SPEED 1
Raed El Khalil, Lebanese American University
Leila Halawi, Lebanese American University

ISSUES IN FORECASTING INTERNATIONAL TOURIST TRAVEL 15
Steven Moss, Georgia Southern University
Jun Liu, Georgia Southern University
Janet Moss, Georgia Southern University

SCAPEGOATING HUMANS, SCAPEGOATING TECHNOLOGIES: EXAMINING ANOTHER SIDE OF INFORMATION SYSTEM PROJECT CONTROL 31
H. Kevin Fulk, Southern University at New Orleans
Obyung Kwun, Southern University at New Orleans
Ghasem S. Alijani, Southern University at New Orleans

EFFECT OF SAMPLE SIZE AND PROCESS CAPABILITY ON VARIANCE CONTROL IN A MULTIPLE FILL-HEAD PROCESS 49
Atul Agarwal, University of Illinois at Springfield
David Strubler, Champlain University

STUDY OF A FAST FACET MINING ALGORITHM USING DIFFERENT TYPES OF TEXT 61
Jose Perez-Carballo, California State University, Los Angeles

COORDINATING PRICING AND INVENTORY PURCHASING DECISIONS OF A SUPPLY CHAIN FOR AN E-TAILER IN FACE OF QUANTITY DISCOUNTS 75
Patrick Lee, Fairfield University

A MODEL FOR MANAGING RENTAL FLEETS IN THE NEW COMPETITIVE LANDSCAPE: MAINTENANCE, PRODUCTIVITY, CORPORATE BRANDING AND LEGAL IMPLICATIONS 85
Anthony Narsing, Macon State College
Terry Sanders, Macon State College
Mimi Ford, Macon State College

DATA MINING TECHNIQUES FOR PREDICTION OF DIFFERENT CATEGORIES OF DERMATOLOGY DISEASES 103
Dinesh K. Sharma, University of Maryland Eastern Shore
H. S. Hota, Guru Ghasidas University, India

LETTER FROM THE EDITORS

Welcome to the *Academy of Information and Management Sciences Journal,* the official journal of the Academy of Information and Management Sciences. The Academy is one of several academies which collectively comprise the Allied Academies. Allied Academies, Inc. is a non-profit association of scholars whose purpose is to encourage and support the advancement and exchange of knowledge throughout the world.

The editorial mission of the *AIMSJ* is to publish empirical and theoretical manuscripts which advance the disciplines of Information Systems and Management Science. All manuscripts are double blind refereed with an acceptance rate of approximately 25%. Diversity of thought will always be welcome.

Please visit the Allied Academies website at www.alliedacademies.org for continuing information about the Academy and the *Journal*, and about all the Allied Academies and our collective conferences. Submission instructions and guidelines for publication are also provided on the website.

Sharad K. Maheshwari
Hampton University

Gérard Fillion
Université de Monckton

www.alliedacademies.org

BENCHMARKING: AUTOMOTIVE FACILITY LAYOUT, JPH, AND OVER SPEED

Raed El Khalil, Lebanese American University
Leila Halawi, Lebanese American University

ABSTRACT

The process layout of any manufacturing facility plays a critical role in the level of effectiveness and efficiency that can be achieved. Process layout serves as a foundation for decisions taken by the manufacturer on issues such as competitive priorities, process strategy, quality, and capacity. The process layout varies from one manufacturer to the other based on the nature of work conducted and the products produced.

The following paper benchmarks the different process layouts utilized by automotive companies within North America. The paper will rank the layout based on Jobs per Hour (JPH) and over speed efficiency. Layout information and data utilized in this paper are obtained from each of the facilities during a benchmarking study conducted in 2010.

Keywords: Over speed, Assembly line, Jobs per Hour, Process layout, automotive industry.

INTRODUCTION

The current economic crisis presents great challenges for organizations in general and the auto industry in particular. In order to dodge financial crises, organizations typically take rigorous measures. These measures force these organizations to set priorities on investments, R & D, efficiency, quality, and others. It's long been understood and it's very well known, since the industrial revolution that manufacturing is the vehicle that drives economic development. Typically manufacturing generates wealth and reduces unemployment. Therefore, manufacturing is the foundation for economic power for any nation or organization in the past, present and future. The U.S. automotive manufacturing industry plays a significant part in the growth of the US economy. It accounts for almost 14% of the total value added for manufacturing making it one of the largest manufacturing sectors of the economy and one of the most significant export industries as well (www.bls.gov).

In 2010, the economy underwent modest growth, which aided the U.S. automotive industry to recuperate from one of the most challenging years in its history. For 2011, auto analysts did predict added growth with sales nearly 13 million units. All of the Detroit Big Three improved their financial performance in 2010. GM and Chrysler are recovering since arising from bankruptcy in 2009, both in terms of sales and their bottom lines. Ford grossed its greatest

profit in more than a decade and reoccupied its position as the second top-selling automaker in the United States. GM made its first profit since 2004, and its largest profit since 1999. (http://trade.gov). No other industry is more linked to the U.S. manufacturing or generates more retail business and employment than the auto industry. The increasingly competitive nature of the automotive industry globally, coupled with declining profits have prompted, manufacturers all over the world to examine ways to make the industry more efficient (Lansbury, Wright & Baird, 2006) while focusing on increasing productivity and improving quality.

All eyes are on manufacturing firms to design facilities that are lean and flexible enough to handle fluctuations in customer's demand. Therefore, manufacturers see layout as an opportunity to maximize efficiency and productivity, and to reduce cost. When designing the layout within the production facility, manufacturers take into consideration "placement of departments, work groups within the department, work stations, machines, and stock-holding points" (Olsen, et.al, 2003). These elements are arranged in a way to ensure a lean and smooth workflow. There are several different types of layouts: (1) Flexible-Flow Layout, where resources and equipment are organized by function (e.g. cutting, drilling, and welding departments), (2) Line-Flow Layout (also known as Product Layout) in which processes, systems, zones and/or departments are arranged in a linear way (e.g. automotive, and electronics industries), (3) Hybrid layout which is a mix of the previous two types where part of the process is Flexible-Flow and the other is Line-Flow , and (4) Fixed-Position Layout (static process) in which equipment and people come to the site to do their work and/or job (e.g. shipbuilding, aircraft, aerospace, farming, drilling for oil, home repair, and construction) (Meyers, 1993).

The following paper presents a benchmarking review of the process/line layouts utilized at several domestic automotive facilities in North America. The objective is to expose process layout efficiency and bottleneck (department and zones) based on JPH (job per hour) and over speed. The information analyzed will be utilized to determine the most efficient process layout.

LEAN MANUFACTURING

Lean production systems have long been adopted in manufacturing sector (Womack et al, 1990; Womack & Jones, 1996).Lean contains five primary elements: Manufacturing Flow, Organization, Process control, Metrics and Logistics (Feld, 2000). Process flow and layout are at the heart of lean manufacturing. The Lean story begins with Kiichiro Toyoda who opened the Toyota car manufacturing company (Bocock & Martin, 2011). Toyoda's vision was to create a company that can provide just in time while empowering workers to make any needed changes and adjustments without compromising quality. The Toyota Production System (TPS) enabled Toyota to compete against the American Mass production methods (Bocock & Martin, 2011). A Lean Manufacturing system is a comprehensive process or business model which centers on methodical identification and reducing waste that is embedded within the system/process (Motwani, 2003; Fathollah et al., 2004; Tapping, 2006; Narasimhan et al., 2006). Taj and Berro

(2005) claim that many manufacturing companies waste over 70 percent of their resources. Bhasin and Burcher (2006) claim that implementing lean can reduce waste by 40 percent. The lean manufacturing objectives are to: improve quality, reduce time, and reduce/improve cost. These objectives are established through a system that is based on a robust foundation that focuses on achieving a quality controlled system. The foundation of the lean system focuses on two main categories: People and Stability. Following the establishment of this foundation, lean focus next on two other categories: Just In Time (JIT) which is delivering the right part at the right time and in the right amount and Jidoka or built in quality (Preventing bad material from advancing within the system). Jidoka is the idea of building quality at the process (as opposed to inspecting it at the end of the production line) by separating normal and abnormal conditions, preventing production when identifying a problem, and calling attention to it to ensure that it's root causes are detected and eliminated. Each of the four categories comprises multiple tools such as: multiple skilled workers lead time reduction, standardized work, Kaizen, Leveling, and others, that are used to support achieving the three objectives. Taiichi Ohno (1988) defined seven common forms of waste as activities that add cost but no value. These forms of waste include over production, unnecessary stocks, producing quality defects, delays, unnecessary transport, inappropriate processing and unnecessary motion (Pascal, 2002)

Several research studies have shown that a lean strategy creates greater levels of quality and productivity and better customer responsiveness (Krafcik, 1998; Nicholas, 1998). The bearing on lean strategy is mostly based on empirical evidence that it advances the company's competitiveness (Oliver et al, 1996; Doolen & Hacker, 2005)

PROCESS LAYOUT

The process layout in the manufacturing facility affects the efficiency and utilization of a company. The objective of any process layout is to organize the company's physical facility in a manner that promotes an efficient use of people, equipment, material, and energy.

Process layout can be defined as the physical arrangement of machines, equipment, and people involved in a manufacturing and assembly process in order to produce a certain product. The foundation for making decisions on the physical arrangement of the process layout is guided by the nature and location of work required within the facility (relative and absolute location), the space required for each process and/or system, the capacity for each process, the physical shape of each process (start and finish stages), and the physical location of each stage within the sequence of processes to insure continuous and/or lean process flow (El-Khalil, 2009).

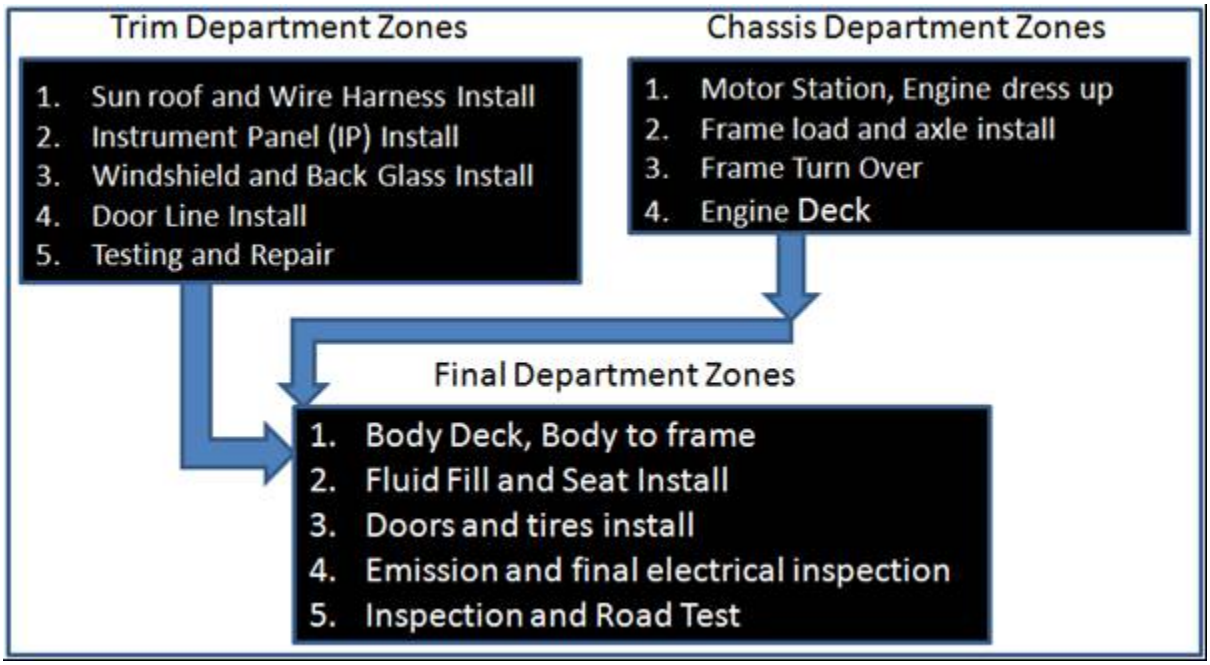
The manufacturing and assembly process flow at the automotive facilities is designed in a sequence flow pattern (line flow), as illustrated in Figure 1. There are three main stages/departments; each department is divided into several zones. The sequences and /or departments are as follows:

Stage one is the body shop department (Body in White, BIW). This is the automobile body assembly without the hang on panels such as doors, hood and lift gate. BIW is the structural frame and the outer "skin"with important aesthetic requirements (Ceglarek, 1994) and is comprised of numerous major components: underbody left and right body frames, shelf, rear end, and roof panel (Jang & Kai, 2001). BIW is divided into 7 to 9 zones. Stamping parts come to the facility from the stamping plant and they are welded together in order to form the shell of the vehicles. This department is highly automated; most of the work is done by welding robots.

Stage two is the paint shop; it is divided into 6 to 7 zones. After the body shop constructs the shell, it is shipped to the paint shop where the shell is washed, coated, and painted. The paint department is divided into the following zones (in sequential order): 1) Sand Deck and Sealer, 2) Phosphate and E-coat, 3) Powder, 4) Color Booth 1 and 2, 5) Polish Deck and Black Out Deck, and 5) Final Inspection and Repair. This department is less automated than the body shop department,

Stage three is the assembly department. It is divided into three departments; each department contains several zones as illustrated in Figure 1 (in sequential order):

Figure 1: Assembly Department: process layout by departments and zones

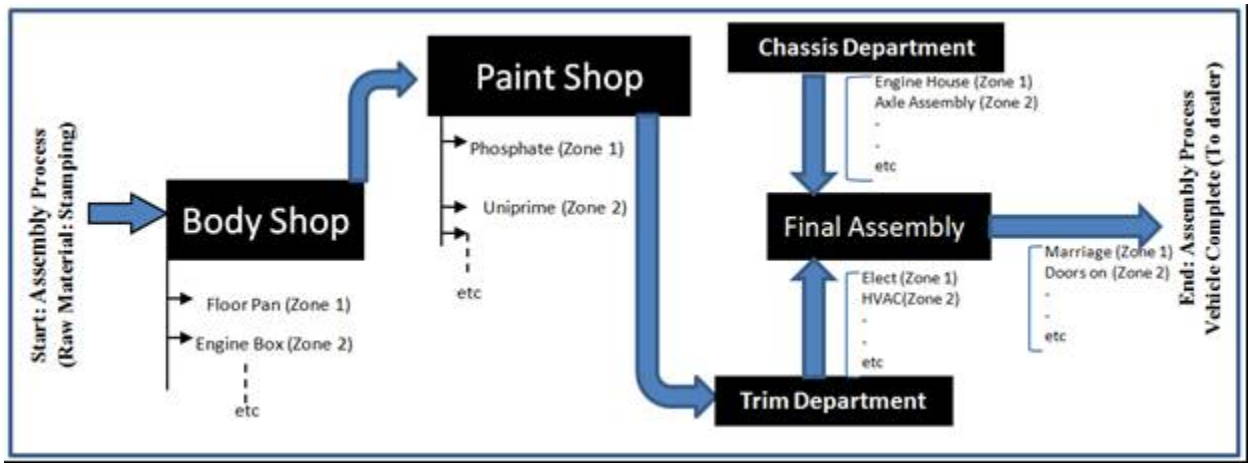


Each department contains 5 to 7 zones. At the trim department installation of the interior parts (e.g. wires, HVAC, and carpet) takes place. At the chassis department, installation of under body parts (e.g. Axle, engine, prop shaft, fuel tank) takes place. At the final assembly the

marriage of the trim to chassis portions takes place in addition to the installation of other parts (e.g. wheels, doors, moldings).

After the final line, the vehicles are inspected, tested, and shipped to customers and/ or dealers as shown in figure 2.

Figure 2: Automotive facility: manufacturing and assembly process layout



The advantages and disadvantages of the process/line flow layout are as follows (Meyers, 1993).

Advantages include:

- High volume at the lowest time (in comparison to the other types of layouts),
- High volume results in lower unit cost. Line layout generates/trains technicians that are specialized in certain operations (skilled and non-skilled). In definition specialization of workers reduces training time and cost. This layout requires less supervision and management due to the connection between the different stages (any problem and/or stoppage will shut down the entire line); therefore, it reduces labor costs for supervision.
- High equipment and/or machine utilization.

Disadvantages include:

- Low Motivation. The repetitive aspect of the job can be very stressful on the employees. The nature of the layout prevents the supervisors and managers from paying close attention to each individual on the assembly line,

- The line flow layout in concept is inflexible. The product produced cannot be easily switched especially if the product's design is somewhat different. When changing products, high setup time is required which in turn leads to higher cost,
- The system is very sensitive to breakdowns, absenteeism, and downtime. Each and every issue that results in shutting down any station in the process will result in a lost production time at the end of the production line.

Several measures such as implementing lean management processes, or flexible manufacturing systems can be used to overcome the disadvantages. However, each one of these processes requires a high initial investment (El-Khalil, 2009). On the other hand, to reduce the shortfalls in the current layouts, companies typically create a buffer (decoupling) between the stations, zones, machines, and departments. This buffer means that more space is required to stock material, and more material will exist between the stations. Therefore, cost will increase. In addition, increasing flexibility means investing in new equipments that require less setup and adjust faster. The initial cost of such equipments is very high in comparison to the traditional equipments. A flexible framing station in the body shop is 250% more expensive than the traditional framing station. Note that the return on investment in the long run for utilizing a flexible framing station can reach up to 350 % (EL-Khalil, Taraman, & Taraman. ,2010). The basic elements of the system are vehicles, robots, machines, conveyors, buffer and operators. The robots and automatic machines can be busy, down for repair, blocked or in starvation mode.

The paper focuses on two domestic companies addressed in the paper as D1 (Lansing, MI) and D2 (Detroit, MI) visited in 2010; the two facilities produce similar vehicle segments. These 2 facilities are the best in class for this benchmarking study. The following variables will be used to measure the performance and effectiveness of the process: JPH (actual/forecasted) and overspeed.

DOMESTIC AUTOMOTIVE FACILITIES (D1 AND D2)

The domestic facilities D1 and D2 are identical from process layout aspect, as illustrated in Figure 1. Each of the two facilities is capable of producing 2 different vehicles, 3 models each. The most significant differences are the following:

- Shift pattern, number of lines, and products produced are illustrated in Table 1.
- JPH for the two facilities shows significant difference, as illustrated in Table 2.
- Body Shop Department: D1 facility performs single stage framing while D2 layered framing
- Assembly Department: D1 facility performs IP, Exhaust system, and engine transmission assemblies while D2 does not.

Table 1: D1 and D2 number of products, shifts and capability						
Facility	Number of Lines			Number of	Shift	Capacity
	BIW	Paint	Assembly	Products	Pattern	Jobs Per Hour (JPH)
D1	1	1	1	3	2	68
D2	1	1	1	3	2	37

Table 2: D1 and D2 JPH by department Forecasted Vs Actual						
Facility	Jobs Per Hour (JPH): Forecasted (F) Vs Actual (A)					
	BIW (F)	*BIW (A)*	*Paint (F)*	*Paint (A)*	*Assembly (F)*	*Assembly (A)*
D1	80	69.4	74	65.3	68	60.7
D2	46	42.1	42	37.4	37	32.3

The overspeed gap between the two domestic facilities shows a 1 to 2 % difference, as illustrated in table 3. Overspeed calculation is determined as follows:

$$Overspeed = \frac{Gross\ JPH}{Net\ JPH} - 1 \qquad \text{---------- ----------} \quad (1)$$

Gross JPH = The JPH rate of a system when it is not down, blocked, or starved. Net JPH = the JPH rate calculated after a period of time, including the effects of down, blocked, and starved time.

Table 3: Overspeed D1 and D2 facilities			
Facility	Overspeed % (Gross Vs Net)		
	BIW	*Paint*	*Assembly*
D1	13%	12%	11%
D2	8%	11%	13%

The overspeed for D2 facility is more efficient in BIW and Paint departments and D1 is more efficient than D2 in the Assembly department.

The detailed analysis for D1 facility, illustrated in Figures 3, 4, and 5 reveals the following:

1. The number one bottleneck in the Body shop is the Frame line at 35%,
2. The number one bottleneck in the paint shop is the Inspection station followed by color booth,
3. The number one bottleneck stage at the assembly department is the trim department at 55%.

Figure 3: D1 Body Shop Department % of JPH losses by zone

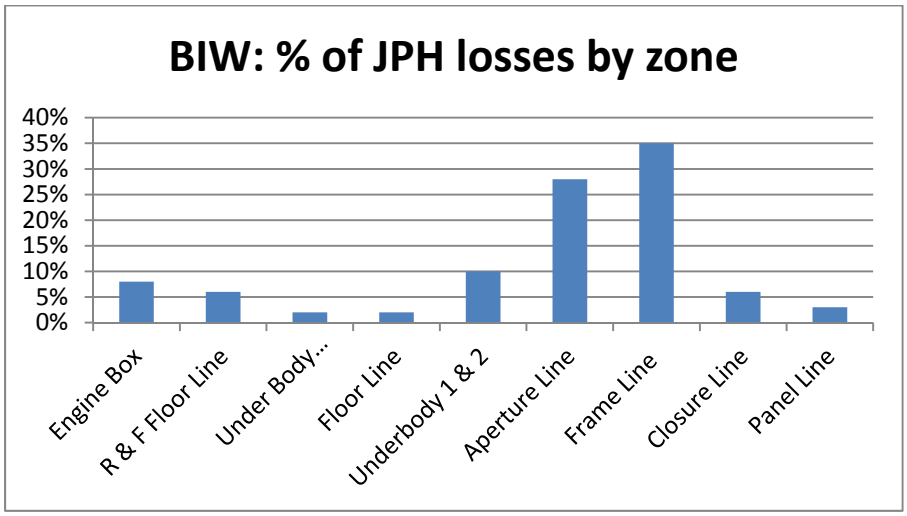


Figure 4: D1 Paint Department % of JPH losses by zone

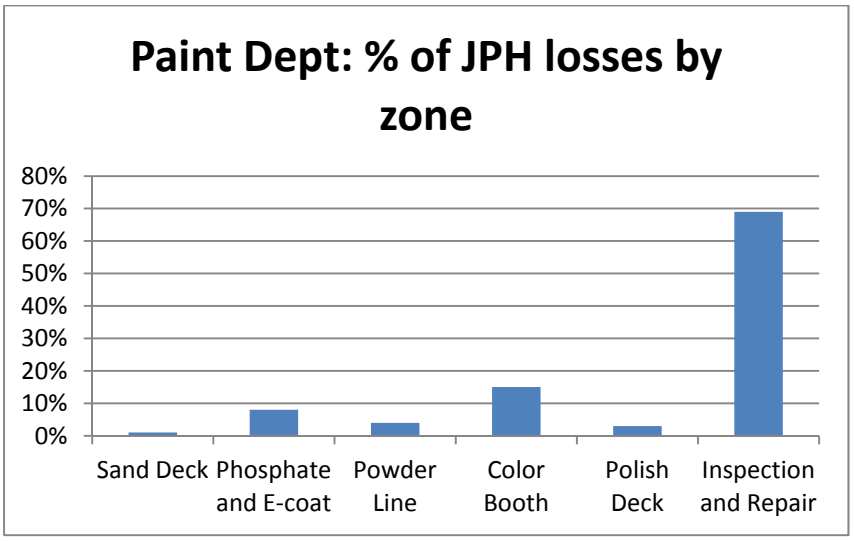
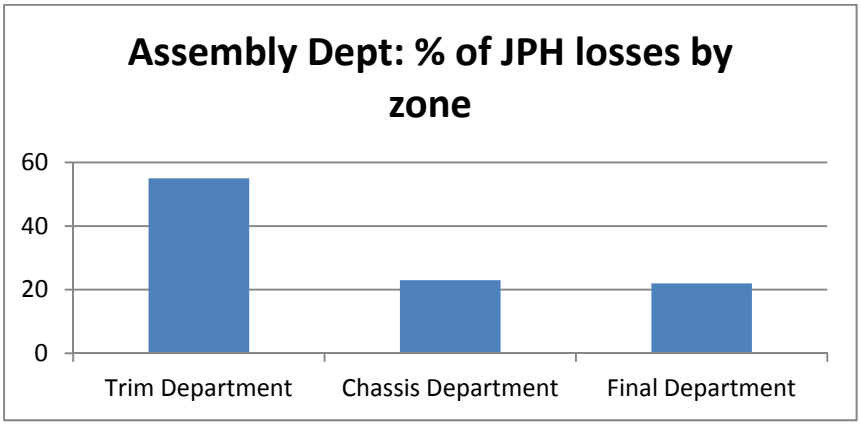


Figure 5: D1 Assembly Department % of JPH losses by department



The detailed analysis for D2 facility, illustrated in Figures 6, 7, and 8 reveals the following:

- The number one bottleneck in the Body shop is the Frame line 37%,
- The number one bottleneck in the paint shop is the Inspection station followed by color booth,
- The number one bottleneck stage at the assembly department is the trim department at 42%.

Figure 6: D2 Body Shop Department % of JPH losses by zone

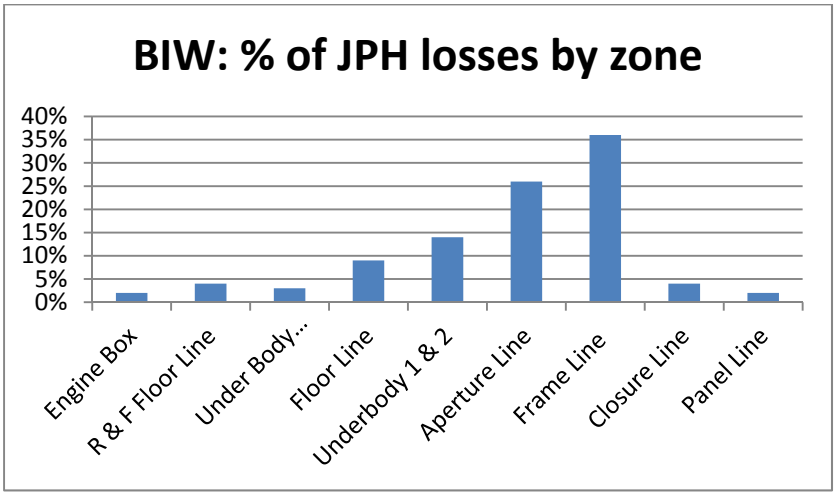


Figure 7: D2 Paint Department % of JPH losses by zone

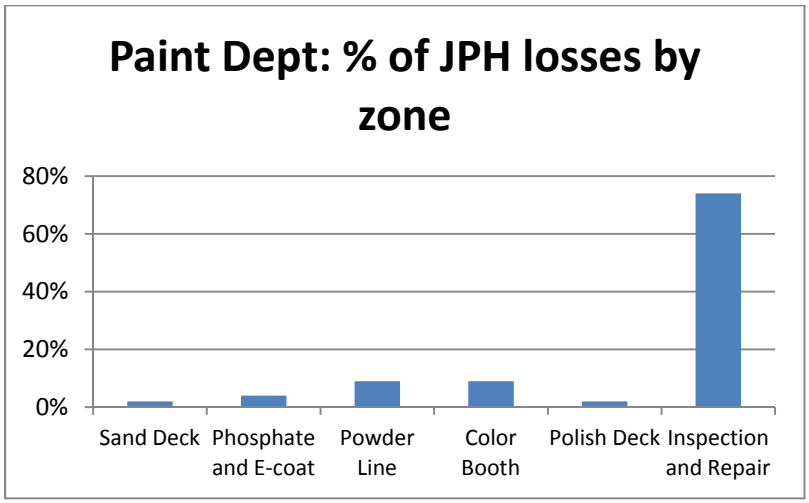
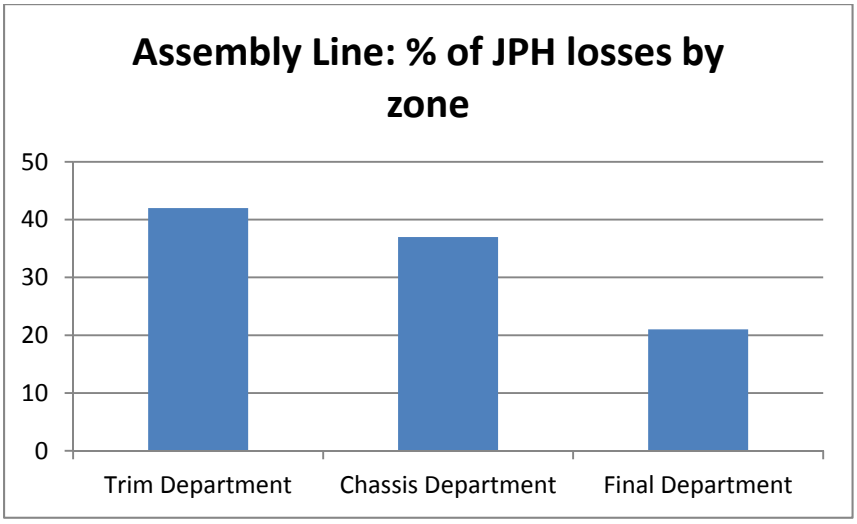


Figure 8: D2 Assembly Department % of JPH losses by department



The domestic facilities bottleneck ranking is very similar (or identical) from the stage perspective of the assembly process.

CONCLUSION

From a gap standpoint based on table 2, the D1 facility is less efficient than the D2 facility. Based on the forecasted to actual production figures, the individual gaps between the two facilities favored the D2 facility by an average of 3%. JPH losses across all departments are driven by zones/areas that are highly labor intensive and in particular in the assembly departments as depicted in figures 3 through 8.Based on overspeed data in table 3, D2 is more efficient than D1 by an average of 2%.

Other variables not considered in this paper contribute to the inefficiencies in the facilities (e.g. Union, economical down turn, lack of liquidity, aging facilities, and high initial investment).

The future of the US automotive industry lies in the abilities of these companies to implement new manufacturing culture driven by improving quality, eliminating waste and reducing costs. The manufacturing facilities at the automotive industry (theoretically) should operate as a team. Management serves as the controllers of the process. They are responsible for organizing their equipment and technicians (skilled and non-skilled) and/or employees to provide consistent work flow through the process.

To improve the current systems, the automotive industry has to absorb significant cost to improve its efficiency and quality. Improving labor efficiency and making high initial investment in manufacturing will payback significantly in the long run (http://trade.gov/static/2011RApt1FINAL.pdf). The most important investment would start with hiring and training qualified labor force on new technologies/systems and in particular embracing the lean culture. Only when the work force is properly trained and committed to the process, every new system that the company may adopt will in turn be successful. Future research will focus on gap analysis that benchmarks the best in class foreign companies such as Toyota and Honda.

ACKNOWLEDGEMENTS

The work presented was partially funded by Chrysler LLC. I would like to thank Chrysler LLC and Mr. Frank Ewasyshy Executive Vice President at Chrysler LLC for supporting this project.

BIOGRAPHIES

Raed EL-Khalil is an assistant professor at the Lebanese American University, school of business and management since 2010. He received his Doctorate degree from Lawrence Technological University., MI, U.S.A. He holds three Masters in Engineering and Management from the University of Michigan, His practical experience includes working in the field of Industrial and Advance Manufacturing Engineering at Chrysler LLC for

over 11 years. Currently serves as a consultant in the field of Flexible Manufacturing For Ford Motors and Chrysler LLC. His research and teaching interest focus on optimization and management of manufacturing processes. He is the author of several publications in the area of Flexible Manufacturing Systems and lean manufacturing.

Leila A. Halawi (DBA, Nova Southeastern University, Fort Lauderdale, Fl., U.S.A) is an assistant professor of business administration in the school of business at the Lebanese American University. Dr Halawi is also an adjunct professor at the University of Tampa and Embry Riddle University. Prior to this, Dr. Halawi was an assistant professor of management information systems at Quinnipiac University. She has 6 year experience as the operations manager for Avis in Brandon, Florida. She has authored several journal articles and contributed to the textbook Decision Support and Business Intelligence Systems. She is frequently invited to present her research at national and international conferences. Her current research interests include knowledge management, business intelligence, information systems success and strategy, and ethical impacts of information technology.

REFERENCES

Bhasin, S. and Burcher, P. (2006). Lean viewed as a philosophy. *Journal of Manufacturing Technology Management*, 17 (1), 56-72.

Bocock, L. and Martin, A. (2011). There's something about lean: A case study. *IEEE Computer Society*, Agile Conference, Salt Lake City, Utah, 10-19.. 10.1109/AGILE.2011.44

Ceglarek, D.J. (1994). Knowledge Based Diagnosis for automotive body assembly: Methodology and implementation. PhD Dissertation, (Ann Arbor, MI.: University of Michigan).

Doolen, T.L and Hacker, M.E. (2005). A review of lean assessment in organizations: an exploratory study of lean practices by electronics manufacturers. *Journal of Manufacturing Systems*, 24 (1), 55-67.

EL-Khalil., R. (2011, November 1). Pour on the efficiency by applying work study analysis to car production. *Industrial Engineering*, 30-35.

El-Khalil, R. (2009). Optimization of Flexible Body System. Doctoral Dissertation, Retrieved from ProQuest Dissertation and Theses (Accession No 3356289).

EL-Khalil, R., Taraman, K., & Taraman. S. (2010, July 21). Body shop flexibility: Impact of framing system on overall performance. *Asian International Journal of Science and Technology*, 3(2).

Fathollah, M., Sabzehparvar, M. and Noorosana, R. (2004). Lean six sigma methodology: A strategy for creating competitive advantage in the third millennium. *Proceedings of the 4th Industrial Engineering Conference, Amirkabir University of Technology, Tehran, Iran*, 408-19.

Feld, W.M. (2000). *Lean Manufacturing: Tools, Techniques and How to Use Them*. Baco Raton, FL: St. Lucie Press.

Jang, K.Y. & Kai, Y. (2001). Improving principal component analysis (PCA) in automotive body assembly using artifical neural networks. Journal of Manufacturing Systems, 20(3), 188-197.

Krafcik, J.F. (1998). Triumph of the lean production system. *Sloan Management Review*, 30 (1), 41-52.

Lansbury, R., Wright, C. & Baird, M. (2006). Decentralized bargaining in a globalizing industry: The automotive assembly industry in Australia. *Industrial Relations*, 61 (1), 70-174.

Meyers, F. (1993). *Plant Layout and Material Handling*, Prentice Hall, Englewood Cliff, NJ.

Motwani, J.G. (2003). A business process change framework for examining lean manufacturing: A case study. *Industrial Management & Data Systems*, 103 (5), 339-346.

Narasimhan, R., Swink, M. & Kim, S.W. (2006). Disentangling leanness and agility: An empirical investigation. *Journal of Operations Management*, 24 (5), 440-57.

Nicholas, J. (1998). *Competitive Manufacturing Management: Continuous Improvement, Lean Production, Customer-Focused Quality*. Irwin/Mc Graw-Hill, New York, NY.

Ohno, T. (1998). *Toyota Production System- Beyond Large Scale Production*. Productivity Press, Cambridge, MA.

Oliver, N., Delbridge, R. & Lowe, J. (1996). The European auto components industry: Manufacturing performance and practice. *International Journal of Operations & Production Management*, 16 (11), 85-98.

Olsen, P. R., Sasser, W. E., and Wyckoff, D. D. (2003). *Management of Service Operations: Text, Cases, and Readings*, available at: www.ateneonline.it/chase2e/studenti/tn/6184-7_tn05.pdf

Pascal, D. (2002). *Lean Production Simplified*. Productivity Press, New York, NY.

Rother, M. (2003). *Toyota Kata: Managing People for Improvement, Adaptiveness and Superior Results,* McGraw-Hill, New York, NY.

Tapping, D. (2006). *The Lean Pocket Guide XL- Tools for the Elimination of Waste*. Kindle ed., Kindle Book, New York, NY.

The Oliver Wyman Group. (2009, August). *The Harbour Report*, Marsh & McLennan Companies, New York, NY.

US Department of Commerce (2011).The Road Ahead. Retrieved January 16 from (http://trade.gov/static/2011RApt1FINAL.pdf).

Womack, J., Daniel, T.J & Daniel R. (1990). *The Machine that Changed the World*. New York: Rawson Associates.

ISSUES IN FORECASTING INTERNATIONAL TOURIST TRAVEL

Steven Moss, Georgia Southern University
Jun Liu, Georgia Southern University
Janet Moss, Georgia Southern University

ABSTRACT

In this paper two popular time series methods for modeling seasonality in tourism forecasts are compared. The first uses a decomposition methodology to estimate seasonal variation. In this method seasonal variation is estimated with a ratio-to-centered moving average approach. Three different approaches in calculating the seasonal indices are analyzed. The deseasonalized series are then forecast using an ARIMA model. The second methodology uses a multiplicative seasonal ARIMA (SARIMA) approach to simultaneously model trend and seasonal variations. The two methodologies are compared and the accuracy and managerial advantages of each are discussed.

INTRODUCTION

Tourism forecasting plays a significant role in estimating economic impacts, allocation of resources, investments, and tourism management (Chen, Chang and Chang, 2009; Li et al., 2006; Karlaftis et al., 1996; Preez & Witt, 2003). Forecasting and measuring the impact of major tourism events are important planning tools for government and private sector businesses. How to measure the impacts of events such as the Olympics on tourism is central to the debate over the economic importance of tourism events (Preuss, 2006; Kasimati, 2003). Additionally, seasonality is an important and often overlooked part of tourism forecasting (Butler, 1994). Due to its importance, there has been a strong, ever-growing interest in tourism forecasting and various models have been developed. A (slightly dated) review of time series models used to forecast tourism can be found in Sheldon and Var (1985).

The purpose of this paper is to compare two popular time series methods for forecasting tourism. The first methodology uses a decomposition model where seasonal indices are estimated with a ratio-to-centered moving average approach. The series are then deseasonalized with the resulting seasonal indices. The deseasonalized series is then modeled with an Autoregressive Integrated Moving Average (ARIMA) model. The second methodology uses a multiplicative seasonal ARIMA (SARIMA) model to simultaneously model trend and seasonal variations. The two methodologies are compared and the advantages of each are discussed.

LITERATURE REVIEW

Li et al. (2006) compared ARIMA and time varying parameter models to forecast United Kingdom tourism expenditures in France, Greece, Italy, Portugal and Spain, using annual data for the period 1980 to 2000. Li et al. found the more complex time varying parameter models to be more accurate. However, they noted future research should incorporate seasonality in the analysis. Li et al. also found a significant intervention from Gulf War I (January-February 1991) in the series.

Multivariate state space models, such as multivariate ARMA, have also been compared to ARIMA models to forecast tourism demand (Preez & Witt, 2003). The data analyzed were tourist arrivals in Seychelles from 1980 to 1994. Preez and Witt concluded that in terms of forecasting performance, ARIMA models are as good as or better than more complex multivariate models.

Tourism at the Milwaukee County Zoo was used to compare forecasts generated with a single moving average with classical decomposition and a SARIMA for the period 1981 to 1999 (Chen, Bloomfield and Fu, 2003). Using the mean absolute percentage error (MAPE) and the root mean square error (RMSE) as measures of accuracy, they found the procedures to be equivalent and preferable to other forecasting models tested. The single moving average with classical decomposition resulted in a MAPE of 18.26% and the SARIMA resulted in a MAPE of 19.15%.

Chen et al. (2009) compared Holt-Winters, GM(1,1) and SARIMA(2,1,2)(1,0,0) models to forecast inbound air traffic to Taiwan between 1996 and 2007. They concluded the SARIMA model was the most accurate for forecasting this series. Chen et al. were able to show the United States was the third largest source of inbound air passengers to Taiwan. Turning point analysis (TPA) and the MAPE were used to compare models. TPA involves predicting the direction of the movement of the series (up or down) using the estimated model. Chen et al. found that the TPA for passengers from the US was 91.67% for a one period ahead forecast, and the MAPE was 3.24%. The longest forecast into the future tested was three periods where the TPA dropped to 83.33% with the SARIMA model.

Econometric models have also been successful in modeling tourism. For example, Song, Witt and Jensen (2003) evaluated several econometric models in forecasting inbound tourism to Denmark from six major origin countries. Song and Witt (2003) used the general-to-specific approach to model the number of tourists to South Korea from four different origin countries. Song, Romilly and Liu (2000) used the general-to-specific approach to study the outbound tourism demand in the UK to twelve overseas destinations. They found that error correction models are supported for most of the destinations.

Smith and Carmichael (2005) studied the affects of 9/11 on Canadian tourism. Smith and Carmichael found that 9/11 did have a significant impact on Canadian tourism. They found changes in destination selection, mode of travel and seasonal patterns comparing pre- and post-9/11 periods. Canadian tourists also showed an increased preference for domestic tourism destinations versus international tourism destinations after 9/11.

Putra and Hitchcock (2006) examined other interventions in air passenger traffic. They found the Bali bombings (October 2002), the 1997 Asian financial crisis, the wars in Iraq, and

the outbreak of severe acute respiratory syndrome (SARS) in 2003 in Asia to all have a significant impact on tourist travel to Bali.

Song and Witt (2003) found events including 1974 worldwide oil crisis, student riots in 1980, the 1988 Seoul Olympics, the establishment of the Tourism Policy Council in 1965 and the subsequent relaxation of foreign travel control, to have significant impact on the number of visitors to South Korea. In studying Demark tourism, Song, Witt and Jensen (2003) identified the following events that significantly affected tourism to Demark from six foreign countries: the two worldwide oil crises (1974 and 1979), the first Gulf war, German reunification, the Chernobyl accident, and 1986 US bombing on Libya.

Moss et al. (2008b) showed the impact of hurricane Katrina on the Mississippi casino industry. In this study a time series decomposition model with a ratio-to-centered moving average was used to estimate seasonality. An ARIMA model was then used to forecast the deseasonalized series. Moss et al. (2008a) used the same methodology when comparing tourism in Las Vegas, NV and Mallorca, Spain. In this study 9/11 was shown to have a significant impact on tourism. Liu et al. (2010) used the same approach along with a more complex SARIMA model to estimate the effects of SARS on air passenger volume from the US to Taiwan, China, and Hong Kong. SARS was shown to have a significant impact on US air travel to these destinations. In Moss et al. (2008a), Moss et al. (2008b) and Liu et al. (2010), it was argued that the ratio-to-centered moving average approach to modeling seasonality was preferable to a SARIMA model, however, no empirical evidence of this was presented.

Although many events, e.g. 9/11, SARS, the Gulf Wars, have a negative effect on tourism, other events, such as the Olympics, may have a positive effect on tourism. How to model the affects of such events is an important element in the debate over how large the benefits are from such events. Often public expenditures to host major events such as the Olympics or the World Cup are justified in part by promises of increased post event tourism revenues. Many of the studies in this area lack ex-post impact results. An ex-post analysis examines the pre- and post- time periods around the event (Kasimati, 2003). Other studies attempt to use control cases to compare trends between supposedly comparable tourism destinations. This ignores external factors not under consideration that can have effects on different destinations (Preuss, 2006). Thus, various approaches to forecast tourism travel have been studied, some more accurate and simpler than others. This paper will compare two of the more popular time series methods for forecasting tourism that have the ability to model ex-post impact results.

DATA

The data for this research was collected from the Research and Innovative Technology Administration coordinating the U.S. Department of Transportation Research Programs (RITA). Monthly passenger air traffic from the US to China, Taiwan, Great Britain, Spain and France for the time periods January 1990 through July 2010 is used in this research. The time periods used were determined by the availability of the data at the time of the research. The overall series length of approximately 20 years is more than adequate for modeling purposes, and far exceeds most prior research in this area. The data is monthly versus annual used in previous studies.

Seat capacity of flights going in either direction, inbound to the US or outbound from the US, is not a constraining factor. The minimum airplane seat occupancy by month for the periods studied between all destinations was 24.1% and the maximum seat occupancy was 98.7%. The average seat occupancy was 74.8% for the time periods studied.

Monthly international air travel volume has been shown to be highly seasonal (Moss et al., 2008a; Liu et al., 2010). For example, Figures 1 and 2 (see results) show the monthly passenger volume from the US to France. The travel volume repeats a similar pattern each year, with low passenger volume in the beginning of the year, reaching a peak during the summer months and dropping to the lowest point at the end of the year.

In this paper, monthly passenger air traffic data from January 1990 to July 2009 are used to estimate the forecasting models. The last 12 observations (August 2009 to July 2010) are reserved to evaluate the forecasting performance of the models when applied to future time periods.

METHODOLOGY

There is a large body of work supporting the concept that tourist destinations have a cyclical, trend and seasonal patterns over time (Butler, 1994; Cooper, 1994; Haywood 1986; Weaver, 1990; Tooman, 1997; Liu et al., 2010; Moss et al., 2003, 2008a&b). The model selected for forecasting tourism, therefore, should incorporate these components. In addition, external events such as 9/11 and SARS have been shown to impact tourism (Liu et al., 2010; Moss et al., 2003, 2008a&b). Therefore the models used must also be capable of incorporating external interventions. Two types of models were selected and compared in this research. The first type uses a decomposition approach where seasonality is modeled with a ratio-to-centered moving average. Three different approaches to calculating the seasonal index are used. The deseasonalized time series are then modeled with an ARIMA model. The second approach uses a SARIMA model to simultaneously model seasonal and trend components in the series.

Methodology 1: Decomposition Models

As noted by Butler (1994), seasonality is extremely important and often overlooked in tourism research. Removing seasonality can provide a more accurate estimation of the trend portion of a forecasting model (Bowerman and O'Connell, 1993). By using the decomposition method, an estimate of seasonal variation can be obtained in the form of seasonal indices and then used to remove the seasonal variation from the time series; thus, isolating variation attributable to long term trends and interventions (Bowerman & O'Connell, 1993; Makridakis et al., 1998; Moss et al., 2003; 2008a; 2008b; Liu et al., 2010). In the decomposition approach, the seasonal indices represent the average percentage of annual passengers for each month of the year. In this paper, three approaches to calculating the seasonal indices are compared. The seasonal index for month *j* (*j=1,…,12*) is calculated as in Equation 1.

$$index_j = \frac{1}{n_j}\sum_{i=1}^{n_j} S_{ij}\,, \qquad (1)$$

In equation 1 j is the month of the year, n_j is the number of the j-th month in the series, S_{ij} is the i-th "raw" seasonal index for month j , $S_t = Y_t/CMA_t$, $CMA_t$is the centered moving average at time t.

The first decomposition model uses all available data to calculate the seasonal indices. Makridakis et al. (1998) showed that seasonal indices are improved by removing the maximum and minimum raw seasonal indices prior to averaging. In this research, the second approach to calculating seasonal indices removes raw seasonal indices more than two standard deviations from the mean. This is done to remove the effects of interventions such as 9/11 and SARS have on the seasonal indices.

In the third decomposition approach, separate seasonal indices are calculated pre- and post- 9/11, in addition to removing outliers in the raw seasonal indices as in the second approach. Table 1 below shows the seasonal indices calculated using all data (Model 1) and the seasonal indices calculated separately before and after 9/11 (Model 3). From this table, significant shifts in the seasonal patterns of air passenger travel after 9/11 can be seen.

Table 1: Seasonal Indices Using Ratio-to-Centered Moving Average %
Model 1 – Uses all months in the series to calculate seasonal indices.
Model 3 – Removes outliers and calculates separate seasonal indices pre- and post- 9/11

	China			Spain			France			Great Britain			Taiwan		
Model	1	3	3	1	3	3	1	3	3	1	3	3	1	3	3
Month		Pre	Post		Pre	Post		Pre	Post		Pre	Post		Pre	Post
Jan	76	71	84	75	77	73	76	73	79	81	78	84	107	105	108
Feb	66	59	78	65	67	66	65	66	68	67	66	72	96	96	96
Mar	92	87	100	95	100	94	96	98	97	91	91	94	89	82	100
Apr	91	94	98	94	98	91	97	100	94	98	97	102	87	89	91
May	109	108	115	113	106	118	118	120	115	109	108	110	105	108	111
Jun	119	126	118	124	118	128	123	123	124	123	123	122	108	110	111
Jul	106	108	105	124	126	120	118	120	115	116	119	112	111	113	107
Aug	120	121	107	135	140	127	130	137	119	125	130	117	115	119	109
Sep	113	117	100	119	120	121	115	119	113	113	119	108	90	95	84
Oct	111	108	115	97	96	103	97	98	100	107	112	102	92	92	97
Nov	88	88	89	77	77	78	81	79	87	88	86	94	93	90	96
Dec	104	106	104	82	82	83	83	76	93	81	78	92	107	109	104

Each deseasonalized series is then modeled with a Box-Jenkins ARIMA model with intervention factors. Each deseasonalized series is first differenced to achieve stationary series prior to estimation of the AR and MA components. Intervention factors are used to model the effects of 9/11, SARS, Gulf War I, and other significant events.

Methodology 2: SARIMA Model

The decomposition model is compared to the multiplicative Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The SARIMA is a parsimonious model for seasonal time series (Box and Jenkins, 1976). Additive indicator variables are used to model intervention effects. The models are of the form of Equation (2) below.

$$Y_t = \sum_{i=1}^{k} \omega_i d_{it} + \frac{(1-\theta_1 B)(1-\theta_{12} B^{12})}{(1-B)(1-B^{12})} \varepsilon_t \qquad (2)$$

Some of the series studied in this paper exhibit increasing variability as travel volume rises over time. In order to achieve constant variance, a log transformation of the original travel volume is used. In Equation 2, Y_t is the logarithm of the original series, B is the "backshift operator" defined as $B^i Y_t = Y_{t-i}$. The indicator variables (d_{it}) are used to model possible intervention effects, i.e., d_{it}=1 for $t \geq t_i$, 0 otherwise. t_i is the time when the i-th intervention event occurs. The noise term ε_t is assumed to be iid normal with mean zero and constant variance σ. To see if the series is stationary, the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) is used and the test results show some series are non-stationary. To transform the data to a stationary series, differences of orders 1 and 12 are taken by using the differencing term $(1-B)(1-B^{12})$. The ADF test is applied on the differenced series and test results show the series are stationary. The seasonal correlations are modeled with the MA component $(1-\theta_1 B)(1-\theta_{12} B^{12})$. For different series studied in this paper, the ARIMA models identified are slightly different (for example, AR terms instead of seasonal differencing are used in some of the models), for details of the identified models, please see Tables 2-6 below.

RESULTS

For convenience, we use Models 1-3 to denote the three decomposition-based models.

Model 1 – Seasonality calculated with a ratio-to-centered moving average using all data.
Model 2 – Seasonality calculated with a ratio-to-centered moving average with adjustments for outliers.
Model 3– Seasonality calculated with a ratio-to-centered moving average with adjustments for outliers. Seasonal indexes are allowed to be different pre- and post- 9/11 to account for shifts in seasonal patterns.

The performance of the models for US air travel to China, Spain, Great Britain, France and Taiwan, are presented in Tables 2-6 below. The number of outliers identified during model building, as well as the number of indicator variables used to account for the intervention events, is shown. In the tables, "Lag Structure" shows the detailed ARIMA model structure, for Models 1-3, it is the lags included in the ARIMA models, for the SARIMA model, it is the structure of the multiplicative model. For example,

$$(0,1,1) \times (0,1,1)_{12} \text{ is } (1 - B)(1 - B^{12})Y_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})\varepsilon_t \quad (3)$$

$$(1,0,1) \times (1,0,1)_{12} \text{ is } (1 - \phi_1 B)(1 - \phi_{12} B^{12})Y_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12})\varepsilon_t \quad (4)$$

For details of this notation, please refer to Box and Jenkins (1976).

In addition to R-square, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Deviation (MAD) are used to measure the goodness of fit of the models, where

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}} \quad (5)$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} \frac{|e_t|}{Y_t} \quad (6)$$

$$\text{MAD} = \frac{1}{n}\sum_{t=1}^{n} |et| \quad (7)$$

In equations 5-7, e_t is the residual at time *t*, Y_t is the observation at time *t*, and *n* is the length of the series. RMSE, MAPE, and MAD are also used to evaluate post-sample forecasting performance, with residuals e_t replaced by forecast errors. The independence of the residuals is checked using the Ljung-Box Q, as well as the sample Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF). Turning point analysis is conducted both within- and post-sample and the percentage of time that the direction of the movement is correctly predicted is reported in the tables.

Table 2: Summary Results: US to China

	Model 1	Model 2	Model 3	SARIMA
Outliers	7	7	5	5
Lag Structure	1,2,5,9	1,2,5,6,9	1,2,5,9	(0,1,0)x(0,1,1)_12
Interventions	6	6	6	4
R^2	98.8%	98.8%	99.0%	98.8%
In-sample RMSE	4,354	4,356	3,408	3,775
In-sample MAPE	9.85%	10.00%	8.90%	9.47%
In-sample MAD	2,577	2,670	2,198	2,404
Ljung-Box Q(12)	0.12	0.23	0.92	0.91
ACF & PACF 1-11	Pass	Pass	Pass	Pass
ACF & PACF 1-12	Fail	Fail	Pass	Pass
In-sample TPA	80.2%	80.6%	81.5%	79.5%
Post-sample RMSE	7,741	7,970	4,753	12,317
Post-sample MAPE	6.92%	7.38%	4.34%	9.99%

Table 2: Summary Results: US to China				
	Model 1	Model 2	Model 3	SARIMA
Post-sample MAD	6,064	6,576	4,180	10,239
Post-sample TPA	83.3%	83.3%	91.7%	100.0%

Table 3: Summary Results: US to Spain				
	Model 1	Model 2	Model 3	SARIMA
Outliers	1	2	3	2
Lag Structure	2,3,9	2,3,9	2,3,5,6,9	(1,0,0)x(0,1,1)_12
Interventions	5	5	5	2
R^2	91.4%	91.5%	92.1%	95.5%
In-sample RMSE	4,578	4,537	6,958	4,836
In-sample MAPE	5.60%	5.62%	8.72%	5.37%
In-sample MAD	3,587	3,500	5,486	3,485
Ljung-Box Q(12)	0.00	0.00	0.44	0.77
ACF & PACF 1-11	Pass	Pass	Pass	Pass
ACF & PACF 1-12	Fail	Fail	Fail	Pass
In-sample TPA	85.1%	81.5%	67.1%	85.0%
Post-sample RMSE	8,883	8,676	10,165	11,202
Post-sample MAPE	6.17%	5.80%	8.96%	5.92%
Post-sample MAD	6,677	6,115	8,824	7,514
Post-sample TPA	75.0%	91.7%	83.3%	100.0%

Table 4: Summary Results: US to Great Britain				
	Model 1	Model 2	Model 3	SARIMA
Outliers	3	2	3	5
Lag Structure	1-4,6-8,10	1-4,6-8,10	1,2,6-8	(3,0,1)x(0,1,0)_12
Interventions	4	4	4	5
R^2	94.3%	94.6%	96.1%	97.58%
In-sample RMSE	50,555	34,883	28,019	27,567
In-sample MAPE	5.93%	4.18%	3.32%	3.40%
In-sample MAD	37,704	26,298	20,926	20,430
Ljung-Box Q(12)	0.02	-	0.08	0.3582
ACF & PACF 1-11	Fail	Pass	Pass	Pass
ACF & PACF 1-12	Fail	Fail	Fail	Pass
In-sample TPA	85.1%	86.5%	91.9%	89.32%
Post-sample RMSE	47,927	42,926	43,241	41,333
Post-sample MAPE	6.56%	5.63%	5.05%	4.47%
Post-sample MAD	41,875	36,273	32,588	26,460
Post-sample TPA	66.7%	75.0%	75.0%	83.33%

Table 5: Summary Results: US to France				
	Model 1	Model 2	Model 3	SARIMA
Outliers	1	2	4	6
Lag Structure	1-4,6,8	1-4,8	1-4,6,9	(0,1,1)x(0,1,1)_12
Interventions	4	4	4	4
R^2	93.6%	92.9%	94.8%	97.4%
In-sample RMSE	12,965	13,557	11,175	10,963
In-sample MAPE	5.15%	5.30%	4.14%	4.21%
In-sample MAD	10,313	10,548	8,360	8,496
Ljung-Box Q(12)	-	-	0.33	0.34
ACF & PACF 1-11	Pass	Pass	Pass	Pass
ACF & PACF 1-12	Fail	Fail	Fail	Pass
In-sample TPA	83.3%	85.1%	87.8%	85.0%
Post-sample RMSE	20,237	21,259	22,413	20,361
Post-sample MAPE	6.99%	7.60%	7.06%	6.76%
Post-sample MAD	16,131	17,450	16,884	14,761
Post-sample TPA	75.0%	75.0%	83.3%	91.7%

Table 6: Summary Results: US to Taiwan				
	Model 1	Model 2	Model 3	SARIMA
Outliers	3	3	4	6
Lag Structure	1,2,6	1,2,6	1,2,6	(1,0,1)x(1,0,1)_12
Interventions	5	5	5	6
R^2	84.2%	84.8%	87.5%	83.1%
In-sample RMSE	7,956	7,893	7,378	5,697
In-sample MAPE	7.63%	7.57%	6.70%	5.80%
In-sample MAD	5,868	5,706	5,048	4,306
Ljung-Box Q(12)	0.03	0.07	0.85	0.50
ACF & PACF 1-11	Pass	Pass	Pass	Pass
ACF & PACF 1-12	Fail	Fail	Pass	Pass
In-sample TPA	64.0%	64.9%	74.8%	72.2%
Post-sample RMSE	7,019	6,649	5,026	5,919
Post-sample MAPE	7.66%	7.02%	4.66%	5.77%
Post-sample MAD	5,753	5,669	3,756	4,682
Post-sample TPA	75.0%	83.3%	91.7%	91.7%

All models examined in this paper have MAPE values below 10% both within- and the 12 month post-sample period. Model 3 and the SARIMA model consistently outperform the first two decomposition approaches. It has been proposed that a MAPE of less than 10% is highly accurate, 11-20% is good, 21-50% is reasonable, and more than 51% is inaccurate (Chen,

Bloomfield and Fu, 2003). The models forecast the direction of movement correctly 75-100% of the time.

As shown in Tables 2-6, Model 2 outperforms Model 1 by most of the measures. This is expected because the calculation of seasonal index as a monthly average of the "raw" ratios of the observations to the centered moving averages, is sensitive to outliers in the time series. By removing extreme ratio-to-centered moving averages before calculating the seasonal indices, the impact of intervention events is mitigated and more accurate estimates are obtained.

By allowing the seasonal indexes to shift after 9/11, Model 3 consistently outperforms Models 1 and 2 by most measurements. The results confirm that the 9/11 terrorist attacks had a significant impact on US passengers international air travel, and the seasonality shifted post-9/11. This observation also demonstrates one important (yet often overlooked) aspect in using the decomposition method: the decomposition method is based on the assumption that there is no significant change in the structure of the series during the period of study. However, when the times series spans over a long time period, this assumption is often not met due to changes in the general environment, or due to intervention events such as 9/11. Song, Witt, and Jensen (2003) also discovered structural change in tourism time series. In this case, it is not reasonable to assume the same seasonal index applies to the entire series and it is important to adjust the seasonal indices accordingly.

The estimated equations of decomposition Model 3 and the SARIMA models are shown in Equations 8 through 17. In Equation 8, Yt is the first differences of the log of number of passengers; in Equations 9-12 Yt is the first difference of the number of the passengers. In Equations 8-12, the indicator variables for the intervention effects are denoted by the month and year of the event, for example, 9/01 denotes the terrorist attacks happened September 2001. In Equations 13-17, d_{it} is an indicator variable used to model the effect of the *i*-th event, more specifically, $d_{it} = 0$ before the event occurs, and $d_{it} = 1$ thereafter.

Decomposition Model 3: US to China

$$Y_t = .018 - .5Y_{t-1} - .28Y_{t-2} - .14Y_{t-5} - .08Y_{t-9} - .25\times 9/01 - 1.19\times 4/03 - .69\times 5/03 + .64\times 6/03 + .73\times 7/03 + .38\times 8/03 \quad (8)$$

Decomposition Model 3: US to Great Britain

$$Y_t = 2656 - .24Y_{t-1} - .22Y_{t-2} - .16Y_{t-6} - .24Y_{t-7} - .19Y_{t-8} - 87{,}460\times 2/91 - 160{,}523\times 9/01 - 425{,}226\times 12/07 + 384{,}265\times 1/08 \quad (9)$$

Decomposition Model 3: US to France

$$Y_t = 942 - .16Y_{t-1} - .18Y_{t-2} - .15Y_{t-3} - .12Y_{t-4} - .24Y_{t-6} - .16Y_{t-9} - 41{,}661\times 2/91 + 28{,}134\times 2/00 - 53{,}398\times 9/01 - 27{,}209\times 3/03 \quad (10)$$

Decomposition Model 3: US to Spain

$$Y_t = 300 - .23Y_{t-2} - .24Y_{t-3} - .15Y_{t-5} - .12Y_{t-6} - .23Y_{t-9} - 12{,}408\times 2/91 - 15{,}563\times 9/01 - 7{,}591\times 3/03 + 9{,}268\times 5/08 + 21{,}559\times 4/09 \quad (11)$$

Decomposition Model 3: US to Taiwan

$$Y_t = 241 - .61Y_{t-1} - .38Y_{t-2} - .13Y_{t-6} - 18{,}392\times 9/01 - 33{,}555\times 4/03 - 15{,}881\times 5/03 + 21{,}387\times 6/03 + 30{,}714\times 7/03 \quad (12)$$

SARIMA Model: US to China

$$Y_t = 2.57d_{1t} + 2.97d_{2t} - 6.56d_{3t} - 3.56d_{4t} + \frac{1+.71B^{12}}{(1-B)(1-B^{12})}\varepsilon_t \quad (13)$$

Where d_{1t} and d_{2t} are indicator variables used to account for the effect of the 1996 Taiwan Strait Crisis (February and May 1996, respectively), in which the US responded militarily. d_{3t} and d_{4t} are used to account for the effects of SARS in East Asia (April and May 2003, respectively).

SARIMA Model: US to Spain

$$Y_t = -.44d_{1t} - .28d_{3t} + \frac{1+.52B^{12}}{(1-.93B)(1-B^{12})}\varepsilon_t \quad (14)$$

Where d_{1t} is used to account for the effect of the first Gulf War (February 1991), d_{2t} is used to account for the effect of the 9/11 terrorist attack (September 2001).

SARIMA Model: US to Great Britain

$$Y_t = -.30d_{1t} - .14d_{2t} + .40d_{3t} - .29d_{4t} - .83d_{5t} + .82d_{6t} + \frac{(1-.43B)}{(1-.19B-.38B^2-.26B^3)(1-.97B^{12})}\varepsilon_t \quad (15)$$

Where d_{1t}, d_{2t}, d_{3t} are used to model the effect of the first Gulf War (February, March, and April 2001, respectively); d_{4t} is used to model the 9/11 effect (9/2001). The model also shows that US-Great Britain air travel dropped significantly in December 2007 and bounced back in January 2008. d_{5t} and d_{6t} are used to model this effect.

SARIMA Model: US to France

$$Y_t = -.54d_{1t} + .31d_{2t} + .20d_{3t} - .42d_{4t} - .18d_{5t} + .20d_{6t} + \frac{(1+.48B)(1+.41B^{12})}{(1-B)(1-B^{12})}\varepsilon_t \quad (16)$$

Where d_{1t}, d_{2t}, d_{3t} are used to model the effect of the first Gulf War (February, March, and May 2001, respectively); d_{4t} is used to account for the 9/11 effect; d_{5t} and d_{6t} are used to model the effect of SARS (April and June 2003).

SARIMA Model: US to Taiwan

$$Y_t = -.13d_{1t} - .46d_{2t} - .93d_{3t} - .45d_{4t} - .43d_{5t} - .59d_{6t} + \frac{1+.58B^{12}}{(1-.82B)(1-.96B^{12})}\varepsilon_t \quad (17)$$

Where d_{1t} is used to account for the 9/11 effect (9/2001); d_{2t}, d_{3t} and d_{4t} are used to model the effect of SARS (April, May and June 2003, respectively). The model also shows that

US-Taiwan air travel started to drop significantly in September 2007 and the drop lasted for a few months. d_{5t} and d_{6t} are used to model the effect of this event.

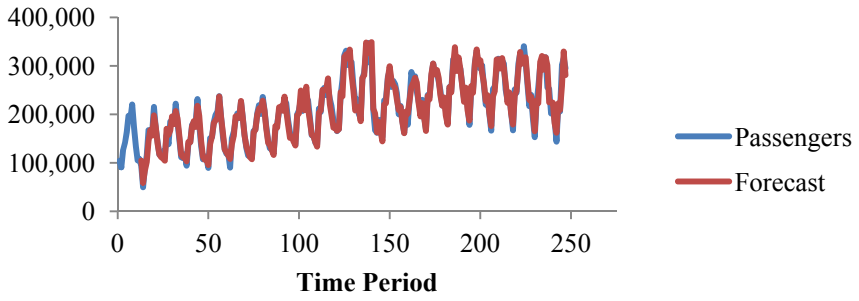
The SARIMA model is more parsimonious than the decomposition models entertained in this paper. In addition to the indicator variables used to model the intervention effects (which remains roughly the same in number across the models), for most of the series the SARIMA model uses only two AR and/or MA parameters (one exception is US to Great Britain, which uses five parameters). The decomposition-based models require additional lag variables.

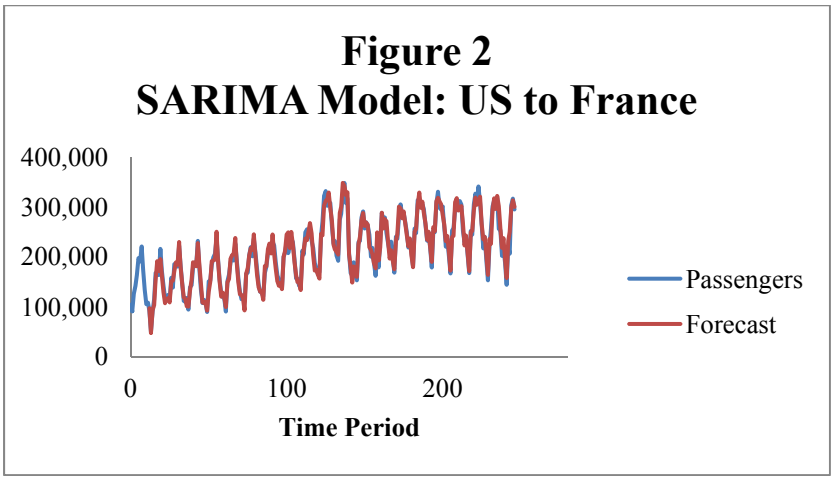
The use of a multiplicative model of the form shown in Equation 2 also allows the lag-12 serial correlation to be accounted for parsimoniously. The residual ACF and PACF indicate that the SARIMA models are free of lag-12 serial correlation in all series, while for decomposition model 3, the lag-12 serial correlation is found significant in three out of five series. This indicates the decomposition approach does not fully account for all seasonal variation. It is interesting to note that Model 3 is free of lag 1-11 residual serial correlation for all five series. This observation shows the importance of modeling the lag-12 serial correlation in highly seasonal monthly time series.

It is possible to extend Models 1-3 to account for the lag-12 correlation by including additional lag terms, but this approach adds complexity to the models and does not always lead to improved forecasting performance. Model 3 is based on the well-known decomposition method, therefore it has the advantage of being relatively easy to implement and should be more appealing to the practitioners in the field. The decomposition approach also estimates seasonal fluctuation explicitly via the seasonal indices. The seasonal indices are useful information by themselves for the practitioner. The SARIMA model, while more parsimonious, does not explicitly estimate seasonal effects and, in general, is more difficult to interpret.

As an illustration of the performance of the models, the estimated number of passengers from the US to France, along with the post-sample forecast by Model 3 and the SARIMA model, are plotted in Figures 1 and 2 below. Both figures show monthly passenger volume from the US to France. Also shown are the forecasts by decomposition model 3 and the SARIMA model.

Figure 1
Decomposition Model 3: US to France





CONCLUSION

Both the decomposition model 3 and the SARIMA model were found to be highly accurate in forecasting air passenger volume to the five locations selected for this research. Forecasting performance is based on the MAD, MAPE, RMSE and TPA. All of these measures were within acceptable to excellent ranges. The model that performs the best by each performance measure and series modeled is summarized in Table 7. A blank cell in Table 7 indicates the two models performed equally.

Table 7: Summary of Model Performance, Comparison of Decomposition Model 3 and SARIMA

	US to CN	US to Spain	US to GB	US to FR	US to TW
Outliers		SARIMA	Decomposition	Decomposition	Decomposition
R^2	Decomposition	SARIMA	SARIMA	SARIMA	Decomposition
In-sample RMSE	Decomposition	SARIMA	Decomposition	SARIMA	SARIMA
In-sample MAPE	Decomposition	SARIMA	Decomposition	Decomposition	SARIMA
In-sample MAD	Decomposition	SARIMA	Decomposition	Decomposition	SARIMA
ACF & PACF 1-11			Decomposition		
ACF & PACF 1-12		SARIMA		SARIMA	
TPA In Sample	Decomposition	SARIMA	Decomposition	Decomposition	Decomposition
Post-sample RMSE	Decomposition	Decomposition	SARIMA	SARIMA	Decomposition
Post-sample MAPE	Decomposition	SARIMA	SARIMA	SARIMA	Decomposition
Post-sample MAD	Decomposition	SARIMA	SARIMA	SARIMA	Decomposition
TPA Post-Sample	SARIMA	SARIMA	SARIMA	SARIMA	Decomposition

The decomposition models show the sensitivity the seasonal index calculations have regarding outliers such as 9/11. The decomposition models also confirm prior research that showed the seasonal travel patterns may have shifted post 9/11 (Smith and Carmichael, 2005). The shift in seasonal patterns post 9/11 may be a result of changes in destination selection, mode

of travel selected or percentage of air travelers which are tourists. Adjusting for outliers and the post 9/11 shift in seasonal patterns greatly increases forecasting accuracy when using a decomposition model. However, even with these adjustments the decomposition models still do not fully remove seasonal variations as well as the SARIMA approach as evidenced by the significant lag 12 autocorrelation in 3 of the 5 series when using a decomposition model.

The SARIMA model also has the advantage of being more parsimonious. However, the decomposition models are simpler to estimate and require a shorter series length. Additionally, the decomposition models explicitly show the seasonal patterns that are hidden in the SARIMA model. The clear presentation of the seasonal patterns via seasonal indices in the decomposition model is valuable information for the practitioner. The seasonal patterns are also useful when modeling the effects of events such as 9/11 or the Olympics. Without accounting for normal seasonal variation an observed shift in the series could simply be normal seasonal variation or an effect resulting from the event.

When modeling events such as 9/11, SARS or the Olympics both models are preferable to methodologies that simply compare same month last year, monthly percentage changes or try to compare the series of interest to comparable destinations (Kasimati, 2003; Preuss, 2006). Both the SARIMA and decomposition models can also be used to estimate the magnitude, duration, and shape of these interventions. This is accomplished by comparing actual to forecast data. The decomposition approach has the advantage of being able to model the effects of the intervention on a seasonally adjusted basis. The decomposition approach is simpler and requires less data to estimate than the SARIMA model. For most applied research or for practitioners, the gains that maybe achieved using the SARIMA model do not always outweigh the increased complexity. As shown in this analysis, the simpler decomposition approach can be highly accurate if carefully applied, e.g., making adjustments for outliers and shifts in seasonal patterns over time.

REFERENCES

Bowerman, B.L. & R.T. O'Connell (1993). *Forecasting and time series: An applied approach (3rd ed.)*. Pacific Grove, CA: Duxbury.

Box, G. E. P. & Jenkins G. (1976). *Time series analysis: forecasting and control*. San Francisco, CA: Holden-Day.

Butler, R.W. (1994). Seasonality in tourism: Issues and problems In Seaton, A.V. et al.(Eds). *Tourism: The State of the Art* (pp. 332-339), Chichester, West Sussex, UK: John Wiley and Sons.

Chen, R.J.C., P. Bloomfield & J. Fu (2003). An Evaluation of Alternative Forecasting Methods to Recreation Visitation. *Journal of Leisure Research*, 35(4), 441-454.

Chen, C., Y. Chang & Y. Chang (2009). Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. *Transportmetrica*, 5(2), 125-140.

Cooper, C. (1994). The destination life cycle: an update In Seaton, A.V. et al.(Eds). *Tourism: The State of the Art* (pp. 340-346), Chichester, West Sussex, UK: John Wiley and Sons.

Dickey, D.A. & W.A. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427–431.

Haywood, K. M. (1986). Can the tourist-area life cycle be made operational? *Tourism Management*, 7(2), 154-167.

Karlaftis, M. G., Z. G. Konstanttinos, P. D. Jason, J. D. Papastavrou & J. M. Charnes (1996). Methodological Framework for Air-Travel Demand Forecasting. *Journal of Transportation Engineering*, 122(2), 96-105.

Kasimati, E. (2003). Economic Aspects and the Summer Olympics: a Review of Related Research. *International Journal of Tourism Research*, 5, 433-444.

Li, G., K. K. F. Wong, H. Song & S. Witt (2006). Tourism Demand Forecasting: A Time Varying Parameter Error Correction Model. *Journal of Travel Research*, 45(2), 175-185.

Liu, J., S.E. Moss & J. Zhang, (2010). The Life Cycle of a Pandemic Crisis: SARS Impact on Air Travel. *Journal of International Business Research*, (2nd Quarter/Spring)

Makridakis, S., S.C. Wheelwright & R.J. Hyndman (1998). *Forecasting Methods and Applications 3rd ed.* New York, NY: John Wiley & Sons Inc.

Moss, S, E., C. Ryan & C. Wagoner (2003). An Empirical Test of Butler's Resort Product Life Cycle: Forecasting Casino Winnings. *Journal of Travel Research*, 41(4), 393-399.

Moss, S.E., C. Ryan, & J. Moss (2008a). The Life Cycle of a Terrorism Crisis: Impact on Tourist Travel. *Tourism Analysis*, 13(1), 33-42.

Moss, S.E., C. Ryan & J. Moss (2008b). The Impact of Hurricane Katrina on Mississippi's Casino Industry. *Academy of Strategic Management Journal*, 7(special issue), 11-16.

Preez, J. d. & S. F. Witt (2003). Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, 19(3), 435-451.

Preuss, H. (2006). Lasting Effects of Major Sporting Events. Retrieved February 1, 2012, from www.idrottsforum.org.

Putra, I. N. D. & M. Hitchcock (2006). The Bali bombs and the tourist development cycle. *Progress in Development Studies*, 6(2), 157-166.

Sheldon, P. and T. Var (1985). Tourism Forecasting: a Review of Empirical Research. *Journal of Forecasting*, 4(2), 183-195.

Song, H., P. Romilly and X. Liu (2000). An Empirical Study of Outbound Demand in the UK. *Applied Economics*, 32(5), 611-624.

Song, H., S. F. Witt and T. C. Jensen (2003). Tourism Forecasting: Accuracy of Alternative Econometric Models. *Journal of International Forecasting*, 19, 123-141.

Song, H. and S. F. Witt (2003). Tourism Forecasting: The General-to-Specific Approach. *Journal of Travel Research*, 42, 65-74.

Smith, W. W. & B. A.Carmichael (2005). Canadian Seasonality and Domestic Travel Patterns: Regularities and Dislocations as a Result of the Events of 9/11. *Journal of Travel & Tourism Marketing*, 19(2/3), 61-76.

Tooman, L.A. (1997). Multipliers and Life Cycles: A Comparison of Methods for Evaluating Tourism and Its Impacts. *Journal of Economic Issues*, XXXI(4), 917-932.

Weaver, D. B. (1990). Grand Cayman Island and the Resort Life Cycle Concept. *Journal of Travel Research*, 29(fall), 9-15.

SCAPEGOATING HUMANS, SCAPEGOATING TECHNOLOGIES: EXAMINING ANOTHER SIDE OF INFORMATION SYSTEM PROJECT CONTROL

H. Kevin Fulk, Southern University at New Orleans
Obyung Kwun, Southern University at New Orleans
Ghasem S. Alijani, Southern University at New Orleans

ABSTRACT

A large number of information system (IS) projects are considered to be failures for not achieving the objectives set for them. One means to improve success of these projects is adequate control. While a growing body of studies have examined IS project control, researchers have given little attention to two issues, project stakeholders' understanding of: 1) the roles of scapegoating in IS project control and 2) the roles of information technologies involved in IS project control. Using an interpretive case study of a series of related IS projects, this study investigated these issues. Results show that project stakeholders may exercise and interfere with IS project control by drawing upon the practice of scapegoating. Furthermore, information technologies may play multiple roles in IS project control beyond their mere use as control mechanisms. These technologies may serve as objects of scapegoating, while interfering with project stakeholders' efforts to exercise control through the practice of scapegoating.

INTRODUCTION

Information systems (IS) projects, intended to develop and/or implement these information technologies, often fail to achieve the objectives that management in business organizations intend them to accomplish. Indeed, a recent Standish Group report finds that only 32% of IS projects deliver their required functions and features on-time and on-budget, with most projects failing to achieve such goals to varying degrees (Standish Group, 2009). Moreover, IS project failure rates have increased in recent years (Standish Group, 2009). To improve the ability of project stakeholders (i.e., those individuals with an interest in a project and its outcomes, such as a project manager) to address the high failure rates of IS projects, increasing numbers of IS researchers (e.g., Kirsch, 1997, 2004; Kirsch, Sambamurthy, Ko, & Purvis, 2002; Mähring, 2002; Soh, Chua, & Singh, 2011) have investigated the control of these projects.

As defined in this study, control consists of any measures to align the interests and actions of individuals or groups to increase the likelihood of their compliance with the objectives established for an activity (e.g., a project) (Law, 1986, 1987, 1992). Adequate project control can enhance the ability of those overseeing an IS project to ensure that project stakeholders direct their efforts to accomplishing the project's objectives (Kirsch, 2004).

Existing studies provide valuable insights about IS project control. These studies suggest that project stakeholders exercising control (or controllers) select one or more project control strategies (Mähring, 2002). Following strategy selection, stakeholders engage in decision-making processes about the control of their projects (e.g., Kirsch, 2004; Soh et al 2011). These decision-making processes may vary over the course of a project (e.g., Kirsch, 2004) and may be targeted at particular stakeholder groups (Soh et al 2011). Stakeholders' decisions about control may be influenced by a number of factors (e.g., task characteristics) (e.g., Kirsch, 1997). To implement their decisions, project stakeholders often employ portfolios of control mechanisms (i.e., related collections of tools for exercising control) (e.g., Kirsch, 1997, 2004; Kirsch et al 2002; Soh et al 2011).

According to existing studies, the portfolios of control mechanisms used by project stakeholders often include formal types of controls (also known as control modes), in combination with informal control modes. Formal control modes are formally documented (i.e., written) and initiated by managers (e.g., Jaworski, 1988). In contrast, informal control modes are often not written down. Formal control modes are commonly viewed as including outcome and behavior control (e.g., Eisenhardt, 1985; Ouchi, 1979; Wiredu & Sørensen, 2006). Outcome control involves specifying outcomes, assessing outcomes, and providing rewards if the desired outcomes are met (Ouchi, 1979). Behavior control involves specifying desired behaviors, evaluating behaviors, and providing rewards if those being controlled (or controlees) exhibit the desired behaviors (Eisenhardt, 1985).

In contrast, informal control modes include clan control and self-control (Choudhury & Sabherwal, 2003). Clan control entails engendering shared goals and values among employees by selection for employment and activities to socialize employees. This is done to minimize preference differences between controllers and controlees. To exercise self-control, the person to be controlled also serves as the controller, establishing the desired standards for goals and behavior and engaging in self-regulation to achieve the desired outcomes and behavior (Henderson & Lee, 1992). The distinction between formal and informal controls is not always a clear one. In recent years, several control researchers have suggested that the formal/informal control modes distinction should be viewed as a continuum, rather than as a dichotomy (e.g., Choudhury & Sabherwal, 2003; Jaworski, 1988; Kirsch, 2004).

While existing studies have enhanced researchers' and practitioners' understanding of IS project control, these studies tend to give little attention to two issues that may merit further examination. The first of these is the relationship between scapegoating and IS project control. Scapegoating is the action of transferring blame for negative outcomes from one actor to another (e.g., Douglas, 1995). Scapegoating is a common occurrence in IS project settings (Runge, 2009). The targets of scapegoating need not be limited to human beings and may, at times, include information technologies (e.g., Nah, Mao, & Benbasat, 1999). While scapegoating may be used merely to avoid blame (e.g., Douglas, 1995), a number of researchers have considered the use of scapegoating for other purposes.

Bonazzi (1983) introduces the concept of instrumental scapegoating. Instrumental scapegoating is actions of assigning blame, used by those in positions of power for maintaining their influence over others. Other researchers (e.g., Cooke, 2007; Levy, 1998) also point to the use of scapegoating to maintain or enhance positions of influence over groups. However,

scapegoating is not only used by those in powerful positions. It may also be used by group members to resist actions of those in more powerful positions (e.g., managers) (Ellwardt, 2011).

This concept of the instrumental use of scapegoating fits well with the activities that form part of the concept of control (e.g., Law, 1986) used by this study (please see a description of this control concept below). Thus, work of researchers from other fields suggests the importance of examining the possible role of scapegoating in efforts to exercise control in IS projects. However, existing studies of IS project control (e.g., Soh et al 2011) appear to have given little, if any, attention to this issue. For this reason, this study seeks to complement existing IS project control research by examining the following research question:

1) What roles do project stakeholders interpret scapegoating as playing in efforts to exercise and impede control in an IS project?

The second issue that has received little attention in existing IS project control research is the roles of information technologies. Typically, these studies examine the roles of information as tools for exercising control (e.g., Kirsch, 2004). According to these studies, these technologies may be used to monitor project status (e.g., Choudhury & Sabherwal, 2003; Kirsch, 2004), and to structure project stakeholders' behavior or the outcomes that result in part from that behavior (e.g., Choudhury & Sabherwal, 2003; Kirsch, 2004). Information technologies may be used in a complementary fashion as parts of portfolios of control mechanisms (Kirsch, 2004). Thus, in these studies information technologies tend to be viewed as playing stable roles as tools (Orlikowski & Iacono, 2001), the use of which leads to predictable outcomes. An open question that remains is whether these technologies play other roles in IS project control, such as impeding efforts to exercise control. Building upon the first research question, this study investigates the following second research question:

2) What roles do project stakeholders interpret information technologies as playing in the relationship between scapegoating and control of an IS project?

The wording of these research questions is consistent with the interpretive research method that is employed by this study. This method is described later below.

The remainder of this paper is organized as follows. In the next section, the theoretical perspective of control used by this study is presented. Afterwards, the methodology employed to understand how project stakeholders interpret the issues presented in our research questions is described. Then, it presents the narrative of the case study. An analysis of the case in terms of the issues presented by the research questions follows. It concludes by discussing implications of this study for research and practice, limitations of this study, and possible future research directions.

THIS STUDY'S THEORETICAL PERSPECTIVE ON CONTROL

This section presents the theoretical perspective on control used by this study. As described later, this perspective guided this study's data collection and analysis. Prior to doing so, it considers the behavioral control perspective, a concept of control used by most existing IS project control research (e.g., Kirsch, 2004).

According to the behavioral perspective, control consists of activities of measurement, evaluation, and corrective action (Eisenhardt, 1985). This approach allows for multiple ways to exercise control through the regulation of human behavior in non-routine (e.g., IS implementation), as well as routine, situations. The multiple approaches may be categorized into the formal and informal modes, or a combination of the two (Soh et al 2011), as described above. The theoretical perspective adopted by this study provides a useful alternative to the dominant behavioral control perspective found in IS project research.

As an alternative to the behavioral control perspective, this study draws upon a perspective developed by sociologist John Law in a series of works (Law, 1986, 1987, 1992, 1994). Like the commonly-used behavioral control perspective, control for Law is concerned with evaluating actors' compliance with the objectives for an activity, and the use of corrective action to increase the likelihood that they will act as intended by controllers. However, it possesses a number of distinctive characteristics give it the potential to provide a richer view of control in IS projects. Before considering those distinctive characteristics, it is appropriate to consider the concepts that make up Law's perspective on control.

To begin, Law's perspective incorporates a number of concepts from Actor-Network Theory (e.g., Latour, 2005). The first of these is that of actors. Actors may be defined as entities, not limited to human beings, which take actions capable of influencing the activities or other actors or events' outcomes (e.g., Law, 1992). This perspective's symmetrical treatment of nonhuman entities (e.g., machines), which gives equal potential influence to humans as well as nonhumans, has been criticized (e.g., Munir & Jones, 2004). However, Pentland and Feldman (2007) note that treating humans and nonhumans in this way reflects the empirical reality that humans and machines often can serve as substitutes for one another. A concept that builds upon that of the actor is that of the actor-network.

An actor-network is an aligned collection of actors that a controller attempts to assemble and maintain to accomplish particular objectives (e.g., Law, 1992). Consistent with the concept of actors described above, this related collection of actors may include humans, as well as nonhumans. Control is exercised through these activities of creating and maintaining the actor-network in which actors are supposed to play roles that further the control objectives (e.g., Law, 1986). Translation is the process used to achieve the needed alignment of efforts of actors making up the actor-network. Given space limitations, an extensive discussion of translation is not included in this paper. Readers wanting to learn more about translation are advised to consult Callon (1986)'s discussion of it. In addition, Law's perspective incorporates the concept of inscriptions. Inscriptions consist of previously-developed knowledge converted into an easily-reproducible form to influence the work of other actors (e.g., Law, 1992). An example is a manager's previous experience converted into policies placed on a company's intranet to be downloaded by employees.

As noted previously, Law's control perspective is conceptually similar to the commonly-used behavioral perspective. Two differences make Law's perspective a useful alternative to the behavioral perspective. First, Law's perspective is a strategic view of control, in which power relations have an important place (Law, 1987). Power relations are relevant to understanding how particular actors gain and maintain controlling positions, how they define control objectives, and how they maintain other actors' faithful compliance with the control objectives. Furthermore, Law sees control as potentially marked by struggles (Law, 1987). Specifically, actors may resist efforts to exercise control by failing to play the roles assigned to them by the controlling actor (Law, 1987). The important role of power relations is not explicitly considered in the behavioral control perspective.

Second, Law's approach of viewing actors symmetrically (i.e., that both humans and nonhumans may influence operations of control) contrasts with that of the behavioral perspective. Instead, the behavioral perspective privileges human influence over control (e.g., Eisenhardt, 1985). Law's perspective's attention to possible roles of nonhuman actors in control makes it well-suited to guiding research to better understand the possible involvement of nonhuman actors (e.g., information technologies) in IS projects. Indeed, it reminds researchers to consider these actors' possible involvement beyond their use as tools for control. Taken together, these differences make Law's perspective a compelling alternative to the behavioral perspective for investigating. It offers the potential to provide a richer view of IS project control than would be possible using the behavioral perspective.

METHODOLOGY

This study employs the empirical method of interpretive case research (e.g., Walsham, 2006). This choice was an appropriate one, given the subject of control. Issues of power and control may be interpreted in multiple ways by study participants and other actors (Law, 1994; Silva, 2007). Interpretive case research focuses on these meanings that actors assign to their experiences. This form of research is based on the double hermeneutic (e.g., Law, 1994). In this way researchers develop interpretations of the phenomenon under study, based in part on other actors' interpretations. Researchers convey their interpretations as written narratives. In this way, our research approach was well-aligned with the subject of our study. Concepts making up Law's perspective on control served as the theoretical lens (e.g., Klein & Myers, 1999) guiding the data collection and data analysis for this study.

The focus of this study was a series of related IS projects to implement and upgrade an enterprise resource planning (ERP) system within a university system. An ERP system is a configurable off-the-shelf software package comprising an integrated set of systems and information resources, used to coordinate a wide range of operational and management activities (Daniel & White, 2005, citing Davenport, 1998; Al-Mudimigh, Zairi, & Al-Mashari, 2001; and Akkermans & van Helden, 2002). These projects were aimed at implementing versions of the ERP system, ERP-Plus, in the Diverse University System (or DUS), a system of universities located in the southern United States.

Four universities comprise the Diverse University System and served as sites for the ERP-Plus projects. Alpha is a comprehensive undergraduate, doctoral, and professional degree-

granting university. Beta is an upper level university with more limited graduate programs. Gamma is an open-enrollment university with preparatory programs, in addition to undergraduate and graduate programs. Finally, Delta is also an upper level university offering undergraduate and graduate programs.

DUS is overseen by a system-level administration, headed by a chancellor, charged with managing day-to-day operations of the system. The system-level-administrators report to a board of regents whose members make decisions about the overall direction of DUS, as well as approving major changes (e.g., adding a new degree program). In spite of this centralized management structure, the DUS member universities, headed by their individual university presidents, enjoy considerable discretion in managing their operations.

Data Collection

Data collection for this study was aimed at developing an in-depth understanding of IS project control and actors' involvement in it, including their roles in actions of blame. Data were collected over a five year period. As this study investigated a phenomenon involving actors that express interpretations verbally (e.g., managers) and those who do not (e.g., software), data collection included interviewing project stakeholders and collecting relevant documents (e.g., emails). This use of multiple sources of data allowed us to attain theoretical saturation, or the confidence that the data provide a comprehensive view of the phenomenon under study (Eisenhardt, 1989).

Seventy-two (72) initial interviews and 30 follow-up interviews were conducted for this study. Interview participants were recruited with the intent of interviewing a wide range of project stakeholders. These interviews were semi-structured, so as to give participants freedom to express their interpretations. Initial interviews, guided by Law's perspective on control, were used to explore the issues presented by our research questions. Follow-up interviews made it possible to clarify participants' responses to questions asked earlier and to ask additional questions to improve our understanding of issues. Interviews were recorded, to facilitate verbatim transcriptions, when permitted by participants. In any case, written notes were taken to record the interviews and/or supplement audio recordings.

Relevant documents were collected to complement interviews of participants. These documents were a useful tool for understanding additional actors' interpretations of project control, as well as the projects' context. Document collection began with resources mentioned by interview participants. This list served as a starting point. Participants were asked to suggest additional documents relevant to understanding the research issues. Document collection, as was the case with the interviews, was guided by Law's perspective on control and our research questions.

Data Analysis

Analysis of the case data proceeded through three stages. In the first stage, the sources of data (including interview transcripts and organizational documents) were read multiple times to develop an understanding of the issues related to Law's perspective on control and our research

questions. An initial interpretation of the case data was prepared in the second stage. During the third stage, an initial case narrative was written. This narrative described the creation of the actor-network associated with the control of the IS projects, and actors' roles in the network, including their roles in connection with the use of blaming actions. This initial narrative was provided to study participants. Where appropriate, their feedback was incorporated to clarify issues relating to our study, allowing us to complete the Hermeneutic Circle (Klein & Myers, 1999). This provided the foundation for preparing the final version of the case narrative and its analysis.

The findings of this study take the form of analytical generalizations (Lee & Baskerville, 2003). As analytical generalizations, the findings are generalizable to the theoretical logic of Law's perspective of control (and to other cases through additional studies), rather than being generalizable to a statistical population (Lee & Baskerville, 2003). Prior to presenting results of this study's analysis, the narrative is presented in the next section.

Research Validation

Klein and Myers (1999) proposed a set of principles appropriate for validating interpretive case research (Walsham, 2006), such as this study. Table 1, found below, presents how these principles were applied to validate this study. The narrative of the case follows this table.

Table 1. Application of Klein and Myers' (1999) Criteria to this Case Study

Klein and Myers' Criterion	Application to This Study
1. The Fundamental Principle of the Hermeneutic Circle: Researchers conducting case studies in the hermeneutic interpretive tradition should engage in a process based on the principle that all human understanding occurs through the act of iterating between considering the related parts of a whole and that whole.	The researchers developed an initial narrative and presented it to participants. The researchers revised this initial narrative by incorporating participants' feedback when appropriate. This allowed the researchers to clarify and validate the relationship between the theoretical framework and the data, by revising language and confirming initial interpretations of concepts that were emerging from the analysis.
2. Contextualization: In order for interpretations to make sense, they need to account for the historical and social context of the case under study.	The historical and social contexts, including those relating to the emergence of the Diverse University System's ERP-Plus projects, are incorporated into the study's case narrative and analysis. One of the researchers was not actively involved in the projects, but was a student at the Diverse University System.
3. Interaction between the Researcher and Subjects: The participants play an active role in research by offering their interpretations of events. Researchers must critically reflect upon the way in which the data were socially constructed as a result of the participants' interactions with the researchers.	The researchers selected documents that they believed would complement the data of the interview transcripts. The interview questions may have introduced the participants to concepts (e.g., actor-networks) shaping to some extent responses that they provided.

Table 1. Application of Klein and Myers' (1999) Criteria to this Case Study	
Klein and Myers' Criterion	**Application to This Study**
4. Abstraction and Generalization: This entails generalizing particulars to abstract categories and social theories. These general concepts describe social action and humans' understanding.	Theoretical concepts were utilized in the field work. These concepts are also reflected in the analysis and discussion. The results of the field work provide illustrations (and in some instances, extensions) of existing theoretical concepts. The application of the framework facilitated the development of findings (taking the form of analytical generalizations) about the uses of blaming actions to exercise and impede control of the ERP-Plus projects, as well as roles of information technologies in this relationship between blaming and IS project control.
5. Dialogical Reasoning: This requires attention to possible contradictions between the theoretical concepts used in the study and the study's actual findings.	The researchers revised their initial conceptions of blaming in efforts to further and impede IS project control, and the roles of information technologies in these actions.
6. Multiple Interpretations: This principle requires that the researchers be sensitive to differences in interpretations of case events provided by participants.	This study utilized findings from different classes of participants (e.g., project team members and faculty), complementing the interviews with supplemental data (e.g., official documents produced by Diverse University System administrators).
7. Suspicion: The researchers must be sensitive to participants' biases and systematic distortions of events stemming from their positions and desired objectives within the organizational context.	This research examines the different interpretations of participants that reflected their particular desired objectives (e.g., those associated with being a university administrator and wanting an IS to enhance control over business processes).

NARRATIVE

In 1995, the Board of Regents, together with the system level administration, reached the decision to implement a new information system. The chosen system, ERP-Plus, was comprised of three applications: student administration, human resource management system (HRMS), and finance. The student administration application encompassed modules to support student administration functions, including admissions, student records, student financials, financial aid, and academic advisement. The HRMS application encompassed modules supporting the functions of human resources, benefits administration, payroll, and time and labor. Finally, the finance application encompassed modules to support general ledger, accounts payable, purchasing, project costing, assets management, accounts receivable, billing, and inventory.

The system-level administrators and board of regents believed that ERP-Plus would solve problems faced by DUS by enabling them to enhance their control over DUS business processes (through increased standardization of those processes), while increasing their ability to meet the reporting demands of governmental agencies. ERP-Plus would eventually replace a number of existing information systems, such as the student administration system shared by Alpha and Delta, known as ManageR. The actors championing the implementation of ERP-Plus considered existing systems such as ManageR to deficient in their abilities to enable adequate control of

operations and reporting, too expensive to maintain, and responsible for undesirable situations faced by DUS (e.g., decreased enrollment at some DUS universities).

Following the decision to implement ERP-Plus, the system-level administrators were appointed to serve as controllers of what would be a series of projects to implement this new IS at the DUS universities. A member of the project team formed to undertake the implementations and upgrades succinctly described the administrators' role: "…the administrators say how we will do it [undertake the implementations and upgrades]." In this case, the approach taken by the administrators was one of centralized control the projects, rather than implementing the applications at each university independently. At times this approach would clash with the culture of independence of member universities that had long existed at DUS.

In this leading role, the system-level administrators turned in 1999 to assembling the varied collection of actors whose involvement would be necessary for the series of projects to achieve their objectives. This group of actors was to work to ensure that ERP-Plus was implemented and used in a cost-effective manner. Aside from the team assembled to undertake the projects, another critical group of actors was the end users (a group of managers, faculty, and staff at the DUS universities), who were expected to assist with the projects and to use the implemented ERP-Plus. A system-level administrator explained why these users were important to the administrators' network: "Well, no system is going to be successful unless you have user buy-in…if the customers are not somehow supportive, you are going to have major problems." It was not only human actors that were supposed to play a key role in the projects.

ERP-Plus itself was, as suggested above, to play a key role in meeting the administrators' and board of regents' objectives. It was enable these actors to enhance their control of operations and improve their ability to provide required reports. However, this software-based information system was also dependent upon a complex collection of additional technological elements (e.g., servers for fulfilling users' reporting requests) to operate. The role of this complexity as it relates to the issue of scapegoating will be discussed in greater detail in the analysis section below.

The process proceeded, and the first versions of the ERP-Plus applications were implemented at member universities by 2001. The first versions of the applications required client software to be installed on users' desktops for interacting with ERP-Plus. These were later replaced by web-enabled versions of the applications, after the software's vendor, ERP-Corp, announced that it would no longer support the earlier versions. This announcement motivated the system-level administrators' decision to upgrade, as they were concerned that the vendor's decision could lead to greatly increased maintenance costs.

Over the course of the process, the system-level administrators and their subordinates made use of control mechanisms found in the existing control research (e.g., Kirsch, 2004). One such tool was what study participants called the "dashboard." The dashboard was a scheduling tool based on an Excel spreadsheet that was used to track progress that actors (e.g., those testing components of ERP-Plus) were making. One administrator described its value: "[The dashboard] was a successful tool because it quantified progress, signaled at least some risks, and flagged issues." Some actors familiar with its use did not share this evaluation of the dashboard.

In contrast, a Beta manager, recruited to work on the implementation described the dashboard's shortcomings: "It [the dashboard] was something that, if you could get everybody

on the same page, it was – it probably would work good…[but] then somebody would work and a couple of weeks later nobody had updated it in a week and there was another spreadsheet that had things in priority order." Thus, for this manager, the dashboard's usefulness could be compromised when it was not updated and when other actors made use of competing scheduling documents.

The use of tools such as the dashboard comprised only one of the dimensions of control used during the ERP-Plus projects. There was another element of control, characterized by struggles among actors involved. Multiple actors failed to play the roles intended for them by the controllers. Project stakeholders did not unanimously share a commitment to the objectives being pursued by the system-level administrators. For example, a faculty member offered her view of the process "To me [implementing ERP-Plus] is pretty senseless – the system that we had was working fine." Over the course of the process, there were ongoing struggles between users and those championing the implementation of ERP-Plus that remained largely unresolved by the end. Users resisted in both covert and overt ways. For example, a project manager noted: "One woman in human resources quit."

The system-level administrators responded with exercises of power to quell resistance. One such measure was their decision in 2004 to form the change management team. A member of this team described her view of its role: "It [the implementation of ERP-Plus] was gonna happen. But it [the change management team] was sort of to strategically deal with resistance as it was emerging…" Other responses to users' resistance were less subtle. Some users were reassigned to new positions; others were fired. ERP-Plus also contributed to the struggles going on in the process.

Aspects of ERP-Plus experienced by users appeared to heighten the struggles going on during the implementation and upgrade process. As noted above, ERP-Plus was intended to standardize business processes used at DUS. However, it also contained inscriptions intended to structure users' work that ran counter to how these employees believed they should do their work. Originally designed for business organizations, it required usage of terminology unfamiliar to users. This language could be found in interfaces that users interacted with, as well as in reports produced by ERP-Plus. Rather than terms familiar to users in a higher educational setting, such as majors, ERP-Plus used different terms. One faculty member recalled: "It [ERPlus] also requires that users deal with terminology not used in higher education. [It makes] use of plans and sub-plans." As the projects continued and actors had chances to use it, this aspect of ERP-Plus appeared to increase their frustration.

Another facet of ERP-Plus that many users found frustrating was the unreliable functioning of some of its components. An especially problematic element of ERPlus was a feature for cancelling students' enrollment. A staff member described this feature's inconsistent functioning: "One day the enrollment cancellation process works. The next day it doesn't…We never know from one day to the next if it will even run – much less run correctly." Problems with reports produced by this feature required staff that used them to devote considerable time to manually correcting errors. Problems of reliability issues were sufficiently great that system-level administrators took care to pick only certain components of ERP-Plus for demonstration to the board of regents for fear of losing the board's continued support of the implementation. Switching to the new versions of ERP-Plus in 2007 helped to reduce, but not eliminate,

reliability concerns among project stakeholders. This implementation of the newer versions of ERP-Plus marks the end of this narrative. However, the analysis presented next delves more deeply into an aspect of the struggles present during the process, scapegoating, and information technologies' roles in this practice.

ANALYSIS

The use of scapegoating was a salient aspect of control during the ERP-Plus projects at DUS. These attributions of responsibility appeared to be a part of efforts to exercise control through gaining actors' support for the projects' objectives. Scapegoating also appeared to be used to undermine control of the implementation and upgrade process. When discussing scapegoating and project control at DUS, this analysis considers the connection of scapegoating with each of these categories of activities. The objects of scapegoating in this case encompassed a diverse array of actors, not limited to human beings. This analysis concludes by considering the identities of these varied actors, efforts to define the boundaries of these actors' identities as part of scapegoating, and challenges of doing this.

The Relationship of Scapegoating with Exercising and Impeding Project Control

As noted above, efforts to exercise control during the implementation and upgrade process entailed defining a solution centered on ERP-Plus to problems identified to be faced by DUS, securing the involvement of particular actors so that they would play roles in furtherance of that solution, and protecting that solution against threats to undermine it. Blaming actors was woven into each of these aspects of control during the process. During the process, actors supporting the move to new system-wide information systems turned to assigning blame as a way to define dimensions of the problems to be solved by implementing ERP-Plus.

A frequent object of scapegoating, identified as a part of the problems to be solved, was the existing student administration information system, ManageR. A former Alpha manager pointed to an instance in which ManageR was defined early on as part of the problem:

> "There was a period when the enrollment of the university went way, way up. Uh, they made some changes – in fact (garbled), then they got concerned that they tightened and then it leveled off and started dropping. And I think the Regents got concerned why is the enrollment dropping?...And the system got blamed for…where is the information…I think a lot of times the system [ManageR] ended up getting blamed for enrollment drops that had nothing with system…."

Assigning blame to ManageR as part of defining the problem continued as the ERP-Plus Student Administration system was implemented. Alpha's associate registrar, whose office helped manage the implementation, described his office's efforts to define the existing system, rather than an existing policy, as part of the problem:

> "Now one [policy] change we did implement that was a huge difference for student financials was that if you were paid…by the time we got to the first day of classes, the right people were in the right seats. Our goal was, frankly, to blame it on the software. Our goal was to dramatically change how we did business in that, if students who hadn't paid by X amount of time, we need those seats available to the other students who could pay….We'd all be in a healthier place financially. And, interestingly enough, that actually worked. And we did it because [since] we are changing software, we might as well do it now. And, yeah, we did actually blame that on the software."

Here, the office of the registrar began to enforce an old policy, but blamed negative consequences for students on ERP-Plus. As suggested above, assigning blame as part of exercising control was not limited to defining the problem to be solved. It also extended to positioning new software as the solution to that problem.

Scapegoating also appeared to be used as a means of defining the actors that were (and were not) a part of the solution to address the problem for which the existing information systems were defined to be a part. Defining the solution in this way appears to have been a part of efforts to induce and secure the compliance of other actors deemed to be a part of the solution. As an illustration of this, consider a former DUS chancellor's efforts to secure the compliance of Alpha faculty members. The faculty members appeared to be regarded by those undertaking the implementation to be a part of the solution, as ERP-Plus was. As noted previously, they were expected to use ERP-Plus after its implementation. As such, exercises of control by those undertaking the implementation seemed to include efforts to secure faculty members' compliance.

Speaking to the Alpha faculty senate, this former chancellor addressed a senator's question about the need to move to ERP-Plus:

> The faculty senator asked: "So, what does ERP-Plus give DUS over what is currently in place?'
>
> The chancellor replied: "The current system [ManageR] is something that has been cobbled together over the decades, adding a piece at a time. It has become increasingly expensive to maintain. The technology that it uses it is outmoded….What DUS needs is a vendor-supported system where annually the vendor will be investing the resources necessary to keep it current so that annually DUS can get upgrades along with other universities that are part of that system. One has to compare that cost to DUS trying to maintain its increasingly outmoded, increasingly expensive and increasingly dysfunctional system."

In this way, the chancellor, by criticizing ManageR, excluded it from being a part of a solution of which ERP-Plus would be part. The use of this tactic appears to have been successful in gaining the support of some faculty members for the ERP-Plus-based solution, rather than a ManageR-based solution.

Finally, scapegoating appears to have been used by those undertaking the projects in an effort to safeguard the process against threats that might undermine it. These actors seemed to take pains to avoid blaming particular actors or groups of actors, blaming others instead. To begin, these actors sought to avoid blaming ERP-Plus. A project manager of the later implementations of the ERP-Plus described the purpose of emails sent to users' departments in response to some users' concerns about the newly-implemented system: "these are the issues that are coming up, no, that wasn't broken, someone had had a bad episode, but...don't go dissing [blaming] the system [ERP-Plus]."

Scapegoating did not appear to be a part of the repertoire of only those who sought to control the implementation and upgrade process at DUS. ERP-Plus applications were objects of blame for multiple actors at DUS. The dean of a school of education within DUS pointed to ERP-Plus: "We will probably never know the true impact of ERP-Plus on the school, the university or system. How many students left the program because they tired of having to correct mistakes the system made." Likewise, a Beta manager blamed ERP-Plus for an enrollment decrease experienced in 2002: "…we are of the opinion that ERP-Plus may be at fault for the enrollment decrease that we experienced in the Spring 2002 semester since we were experiencing similar problems at that time as well."

Scapegoating also appeared to be a tool of those who did not favor the implementation of ERP-Plus at DUS. For example, a technician who worked during the first implementation of the ERP-Plus Student Administration application at Beta claimed: "Academic advisors [employees at Beta] used their powers for evil, stirring up negativity towards ERP-Plus." Instances of assigning blame such as this appeared to be aimed at disrupting other actors' relationships with the implementation and upgrade process. As noted earlier, exercises of control (including those argued here to be based upon scapegoating) had been employed to establish and/or maintain such relationships. Thus, scapegoating appeared to serve as a means for both exercising and interfering with efforts to exercise control, consistent with Law's (1986, 1987, 1992) view of control marked by struggle.

Scapegoating through Defining Actors' Boundaries and Challenges of Doing So

What comprised or did not comprise actors blamed for problems at DUS was a point of controversy among stakeholders. In seeking support for the move to a new system, actors drew the boundaries of the previous system, ManageR, wide. It was this system that was argued to be inadequate for meeting the needs of DUS. In contrast, a user of ManageR argued that the fault lay outside of the boundaries of what they considered to be this system:

> "When the biggest problem that we had on the legacy system [ManageR] was that when we got close to registration…at any given time, you couldn't serve more than 64 students. But it wasn't the software that wasn't working. It was the servers."

Thus, for this ManageR user, the blame lay with another technology (i.e., servers). As the implementation proceeded and applications of ERP-Plus were implemented at DUS institutions,

the issue of whether or not to blame the new system for experienced problems was also a source of controversy.

Just as boundaries of actors can be drawn widely in order to assign blame, they can also be drawn narrowly. More than one actor appeared to follow this approach to avoid assigning blame to ERP-Plus. Here are multiple examples to illustrate this practice. An academic advisor recalled his conversation with a member of the change management team: "…she said 'It's not the software [ERP-Plus] [that caused the problem]; it's a modification [to the software].'" In addition, the project manager seemed to try to shift the blame from ERP-Plus using this approach: "…there were a lot of rumors about the system's doing this, the system's doing that. I can't get in, therefore the system's bad. Well, could it be that maybe your computer wasn't working right?"

The term Ontological Scapegoating is proposed for this practice of actors, through verbal communication or other actions, drawing the boundaries of other actors for the apparent purpose of assigning blame. This term, combining 'ontological' and 'scapegoating,' appears to be appropriate. Ontological may be defined as: "Of or relating to essence or the nature of being" (Merriam Webster Online Dictionary, 2012). Thus, the concept of ontological scapegoating expands upon the concept of instrumental scapegoating found in existing research (Bonazzi, 1983).

The combination of technologies that allowed ERP-Plus to operate seems to have contributed to the difficulty of settling controversies around blame for issues experienced at DUS. Some indication of the complex array of technologies involved was provided by a DUS manager charged with maintaining ERP-Plus. She described it thusly:

> "One of the things with this – the whole ERP-Plus – it is very complicated with lots of moving pieces. You know running a database server that has an Augur database sitting on it, but up in front of it, it's got – this is just the Student/HR system – because they're all on one system – it's got four physical machines called AP servers on it but on each one of those we run multiple instances of this Suit server, which is a transaction cue manager and up in front of that are four different web servers that run about eight instances of a web server, web logic, and there's software deployed in each one of those. There's lots of moving parts between all of these things that, we you get them out of sync and they all have their own kind of cache that can get corrupted – you can run into quite a few problems."

This complex combination of technologies had the ability to mask to a degree the cause or causes of problems experienced by users.

The aforementioned manager explained the sometimes difficult nature of determining causes of problems: "So, yep, the technology can certainly play havoc with what's going on. You know, and it manifests itself in different ways and in the way people think that things are not working…you know you'll have a person who they don't have a single problem whereas the person sitting at the desk next to them…is having problems. So, they're having problems and

this person's not. So, it gets very complicated trying to figure out where the actual problem lies. Some people are working fine and some are not."

As noted above, users were one of the targets of assigning blame. In spite of the effect of the technologies' complexity, it was possible on occasion to determine at least that blame should not be assigned to a particular user. For example, those in charge of overseeing the functional aspects of the HRMS application argued that an Alpha compensation manager was at fault for the resetting of dates connected with individuals' employment relationship with DUS. In this case, ERP-Plus provided the manager with the means to demonstrate that he was not to blame. The compensation manager explained the means of demonstrating that he should not be blamed:

"Then somebody got the bright idea to run a [ERP-Plus] query. (laughter) We realized that there was no way that I could have done them [changes to the data] all! (laughter)…There – audit records didn't prove that I had touched all of them, so there's no way that I could have done it…So, turned out that…it was a different process that when the record was touched, those dates were recalculated." In this way, the collection of technologies that was sometimes viewed as ERP-Plus enabled a user to persuade others that he should not be blamed.

Table 2. Summary of Findings

Research Question	Findings
1. What roles do project stakeholders interpret scapegoating as playing in efforts to exercise and impede control in an IS project?	Project stakeholders may employ the practice of scapegoating as part of exercises of IS project control. Project stakeholders may also use scapegoating to interfere with efforts to exercise project control.
2. What roles do project stakeholders interpret information technologies as playing in the relationship between scapegoating and control of an IS project?	Information technologies may serve as objects of scapegoating to defend other actors against attempts to blame them for organizational problems. Information technologies may interfere with efforts to use scapegoating for IS project control.

As shown in Table 2, considering stakeholders' interpretations of the roles of scapegoating and information technologies illuminated another side of IS project control than that seen in existing studies (e.g., Kirsch, 2004). Drawing upon Law's perspective on control made it possible to develop an in-depth understanding of stakeholders' struggles in connection with the ERP-Plus projects. Analyzing their interpretations, as well as supplemental data, made it possible understand how these stakeholders employed scapegoating to build support for the projects' objectives, as well as challenge them. A key element was the practice (ontological scapegoating) of actors drawing the boundaries of other actors to either build support or opposition. Lastly, information technologies had prominent roles in actors' scapegoating practices. Though often serving as scapegoats themselves, these technologies also appeared to play disruptive roles in other actors' efforts to assign blame. These roles included confounding other actors' efforts to assign blame and allowing actors to defend themselves against accusations.

CONCLUSION

This concluding section considers the implications of this research, its limitations, and possible directions for future research. This study's findings have implications for both IS researchers and stakeholders involved in IS projects. This study contributed to IS project control research by considering two largely neglected issues: the roles of scapegoating in project control and the roles of information technologies in this process. To investigate these issues, it employed a theory of control well-suited to guiding a study of struggles as a dimension of project control and the roles played by information technologies in this process. The findings provide an interesting complement to the growing body of IS control research. They do so by providing insights into how control is exercised, not limited to techniques identified in existing studies (e.g., Kirsch, 2004). Project stakeholders may use scapegoating both to realize their objectives for control, as well as to interfere with project control. The instrumental use of scapegoating, consistent with Law's (1986, 1987, 1992, 1994) strategic concept of control, adds to the view of control commonly held in IS control research by considering the role of power relations in control. In addition, this study also contributes to IS researchers' understanding of information technologies' roles in IS project control. Rather than merely confirming previous studies' findings of the use of these technologies as tools for control, this study also found that they may play disruptive roles as well. In this way, this study contributes to a richer understanding of IS project control.

Likewise, project stakeholders may also benefit from this study's findings. Project managers considering our findings can better understand the possible consequences of employing scapegoating as part of project control. In addition, they can gain a greater awareness of how information technologies may influence their efforts to exercise control. Other stakeholders can also gain a greater understanding of possible approaches that those leading projects may use to gain their compliance (e.g., ontological scapegoating), and how they can respond.

This study has three limitations. To begin, this study only investigated projects on one type of organizational setting, that of a university system. In addition, it did not investigate the involvement of all kinds of information technologies in control. For example, it did not examine the roles of intranets. Finally, it would have been interesting to see what happened at DUS after this study concluded in 2008.

This study also presents a number of opportunities for future research. First, a future study could test the findings of this research in another university system setting. A variant of this study could also examine this study's research questions in another organizational setting (e.g., a corporation). Thus, this research provides a foundation upon which multiple future studies can be built.

REFERENCES

Bonazzi, G. (1983). Scapegoating in complex organizations: The results of a comparative study of symbolic blame-giving in Italian and French public administration. *Organization Studies*, 4(1), 1-18.

Choudhury, V. & R. Sabherwal (2003). Portfolios of control in outsourced software development projects. *Information Systems Research* 14(3), 291-314.

Cooke, H.F. (2007). Scapegoating the unpopular nurse. *Nurse Education Today*, 27(3), 177-184.

Daniel, E.M. & A. White (2005). The future of inter-organizational system linkages: Findings of an international Delphi study. *European Journal of Information Systems*, 14(2), 188-203.

Douglas, T. (1995). *Scapegoats: Transferring blame*. New York, NY: Routledge.

Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review*, 14(4), 532-550.

Eisenhardt, K.M. (1985). Control: Organization and economic approaches. *Management Science*, 31(2), 134-149.

Ellwardt, L. (2011). *Gossip in organizations: A social network study.* (Unpublished doctoral dissertation). University of Groningen, Groningen, Netherlands.

Henderson, J. & S. Lee (1992). Managing I/S design teams: A control theories perspective. *Management Science*, 38(6), 757-777.

Jaworski, B. (1988). Toward a theory of marketing control: Environmental context, control types, and consequences," *Journal of Marketing* 52(3), 23-39.

Kirsch, L. (2004). Deploying common systems globally: The dynamics of control. *Information Systems Research,* 15(4), 374-395.

Kirsch, L., V. Sambamurthy, D-G. Ko, & R. Purvis. (2002). Controlling information systems development projects: The view from the client. *Management Science*, 48(4), 484-498.

Klein, H.K. & M.D. Myers (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, 23(1), 67-94.

Latour, B. (2005). *Reassembling the social*. Oxford, United Kingdom: Oxford University Press.

Law, J. (1994). *Organizing modernity*. Cambridge, Massachusetts: Blackwell.

Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systemic Practice and Action Research*, 5(4), 379-393.

Law, J. (1987). Technology and heterogeneous engineering: The case of Portuguese expansion. In W. Bijiker, T. Hughes, & T. Pinch (Eds) *The Social Construction of Technological Systems, New Directions in the Sociology and History of Technology*, (pp. 111-134). Cambridge, Massachusetts: MIT Press.

Law, J. (1986). On the methods of long-distance control: Vessels, navigation, and the Portuguese route to India. In J. Law (Ed) *Power, Action and Belief: A New Sociology of Knowledge*, (pp. 234-263). London, England: Routledge and Kegan Paul.

Lee, A. & R. Baskerville (2003). Generalizing generalizability in information systems research. *Information Systems Research*, 14(3), 221-243.

Levy, J.S. (1998). The causes of war and the conditions of peace. *Annual Review of Political Science*, 1, 139-165.

Mähring, M. (2002). *IT project governance*. (Unpublished doctoral dissertation). Stockholm School of Economics, Stockholm, Sweden.

Merchant, K.A. (1988). Progressing toward a theory of marketing control: A comment. *Journal of Marketing*, 52(3), 40-44.

Munir, K.A. & M. Jones (2004). Discontinuity and after: The social dynamics of technology evolution and dominance. *Organization Studies*, 25(4), 561-581.

Nah, F.H., J. Mao & I. Benbasat (1999). The effectiveness of expert support technology for decision making: Individuals versus small groups. *Journal of Information Technology*, 14(2), 137-147.

Orlikowski, W. & C.S. Iacono (2001). Research commentary: Desperately seeking the "IT" in IT research – a call to theorizing the IT artifact. *Information Systems Research*, 12(2), 121-134.

Ouchi, W. (1979). A conceptual framework for the design of organizational control mechanisms. *Management Science*, 25(9), 833-848.

Pentland, B. T., & M. S. Feldman (2007). Narrative Networks: Patterns of Technology and Organization. *Organization Science*, 18(5), 781-795.

ontological. 2012. In *Merriam-Webster.com*. Retrieved on February 26, 2012 from http://www.merriam-webster.com/dictionary/ontological

Runge, S. (2009). The art of scapegoating in IT projects. Retrieved on February 27, 2012 from http://www.pmhut.com/the-art-of-scapegoating-in-it-projects

Silva, L. (2007). Epistemological and theoretical challenges for studying power and politics in information systems. *Information Systems Journal*, 17(2), 165-183.

Soh, C., C.E.H. Chua, & H. Singh (2011). Managing diverse stakeholders in enterprise systems projects: A control portfolio approach. *Journal of Information Technology*, 26(1), 16-31.

Standish Group (2009). New Standish group report shows more project failing and less successful projects [Press release]. Retrieved on March 4, 2012 from http://www1.standishgroup.com/newsroom/chaos_2009.php

Walsham, G. (2006) Doing interpretive research. *European Journal of Information Systems*, 15, 320-330.

Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems*, 4, 74-81.

Wiredu, G. & C. Sørensen (2006). The dynamics of control and mobile computing in distributed activities. *European Journal of Information Systems*, 15(3), 307-319.

EFFECT OF SAMPLE SIZE AND PROCESS CAPABILITY ON VARIANCE CONTROL IN A MULTIPLE FILL-HEAD PROCESS

Atul Agarwal, University of Illinois at Springfield
David Strubler, Champlain University

ABSTRACT

Many manufacturing firms still suffer from poor quality of discrete output in spite of the control chart techniques being in vogue for several decades. These firms continue to face significant problems in correctly applying control chart techniques, especially, when dealing with outputs from multiple fill-heads. These problems can either take the form of overreacting to false alarms, or delaying detection of genuine errors due to assignable causes, or failing to detect process errors, or even misapplying the statistical process control methods. Even if such a process is in a state of statistical control, there is a high likelihood that one or more fill-head could drift away from the target due to an assignable cause. It is important that such process changes be detected as soon as possible and corrective actions taken. We suggest that the selection of an appropriate sampling size in a sampling plan can have strong bearing on the sensitivity of detecting these process changes.

The objective of this paper is to investigate the effect of sample size on the sensitivity of detecting undesirable process variance due to assignable cause in a multiple fill-head production process under alternate sampling methods. We also examine the impact of process capability on the sensitivity of control charts. Finally, the paper provides a discussion on the economics and limitations of taking a large sample.

Keywords: Stratified, Random, Power Curve, Control Charts, Monte Carlo Simulation

INTRODUCTION

Continuous quality improvement (kaizen) demands that manufacturing processes be constantly monitored through statistical process control (SPC) methods. Most manufacturing companies today employ at least basic SPC methods. These methods are extremely useful in early detection of quality problems. However, many firms still face multiple challenges in implementing SPC methods, especially, when dealing with output from multiple fill-head production processes. These firms either apply the SPC methods incorrectly, or waste valuable time to detect non-existing errors and assignable causes (commit Type 1 Error), or even fail to

detect serious errors due to assignable causes (commit Type II Errors). Therefore, firms incorrectly using or those lacking sophistication in SPC methods are often plagued with intermittent or even continuous poor quality resulting in unnecessary rework and scrap that can be prohibitively expensive.

Traditionally, most practitioners of control charts have used "three sigma" control limits to minimize false alarms of undesirable change, known as Type-1 error (α–value). However, setting control limits at "three sigma" limits results in the control charts becoming insensitive to detecting small shifts in the production process by increasing Type-II error (β–value). For "kaizen" or continuous quality improvement initiatives to succeed, it is imperative for companies to be able to detect with greater statistical power even those assignable causes that lead to small changes in the process. Wheeler (1983) defines statistical power as the likelihood of a sample statistic falling outside the control limits on the first sample taken just after the occurrence of a shift in the process. It is to be noted that greater the statistical power to detect a process shift, lower is the probability for Type-II error since statistical power is given by $1-\beta$.

Osborn (1990) suggests several approaches for reducing Type-II error, thus increasing the sensitivity of the control charts in detecting process shifts. The first approach is to set control limits at less than 3-sigma limits, say at 2-sigma level. However, this approach runs the risk of increasing Type-I error, thus the likelihood of more false alarms. The second approach for increasing the sensitivity of control charts is to use cumulative information contained in quality run charts based on patterns formed by taking several sequential samples. This approach would incur delays due to multiple sampling. It is not of much help in situations when samples are either taken infrequently, or if process shift occurs for just a brief period of time, or if it becomes critical to detect process shift as soon as it occurs due to quality cost considerations. The third approach to enhance the sensitivity of control charts relates to the sample size used in developing them. As the sample size is increased, the control limits tend to become tighter, thus increasing the statistical power of the control charts to detect process shifts without the detrimental effect of increasing false alarms. To quantify the statistical power of the control charts for detecting process shifts, power curves or operating characteristic (OC) curves may be used. Power curves provide a graphical representation of the probability of rejection or out of control indication ($1-\beta$) corresponding to a given shift in the process for a particular sample size whereas OC curves provide probability of Type-II error (β) as a function of process shift.

Much of the literature on statistical process control has focused on single fill-head production processes. Scheffe (1949) and King (1952) provided operating characteristic (OC) curves for $\bar{x}$ and R charts with rational samples for "standard-given" cases. For "no-standard" given cases, Olds (1961) offers insights regarding characteristics of control charts when rational sampling is used. While still focused on rational samples, beginning in the 1980s and into the 1990s, we see a change in the literature. There is a clear recognition of the need for methods which can detect slight to moderate shifts in the process. For example, employing multiple detection rules for the average process shift, Wheeler (1983) and Palm (1990) demonstrate tables

of the power function for the $\bar{x}$ chart. Following suit, Davis et al. (1993) also utilize the "statistical power" of an $\bar{x}$ chart.

According to Osborn (1990) there is a serious lack of understanding about statistical power and its importance in determining sample size to use for control charts among SPC practitioners. He attributes two primary reasons for such a situation. First, the classic quality control articles by Shewart (1931), Grant & Leavenworth (1987), and Ott (1975) have overemphasized the use of small sample sizes for a variety of reasons. Such a mindset tends to imply that small process shifts or quality problems are not worth detecting and runs counter to the basic quality principles of continuous improvement (kaizen). Second, concepts related to statistical power tend to be abstract and difficult to grasp. Most often, such concepts appear mostly in very technical academic journals with limited audience, but largely ignored in popular QC literature.

However, with manufacturing increasing in complexity production processes often consist of multiple fill-heads. With increased complexity, late detection of even slight shifts in the process can exacerbate costly specification error and scrap. According to Caulcutt (1995) and Evans (1993), in such cases, choosing an appropriate sampling plan both in terms of sample size and method becomes a primary issue for the control charts to function effectively. Wheeler (1983) and Osborn (1990) further highlight this view point by asserting that the control chart will be sensitive enough to detect even minor process shifts without setting off false alarms if QC staff use the correct method and correct sample size. There is limited literature on statistical methods to effectively monitor production processes with multiple fill-heads. More recently, Lanning et al. (2002) employed an adaptive fractional approach to processes with a relatively large number of populations. Other researchers recommend multivariate techniques (Runger et al., 1996), paired Shewart and CUSUM charts (Mortell and Runger, 1995), group control charts (Montgomery, 1982), and plots of raw data, residuals methods, and analysis of variance to examine multiple fill-head machines (Otto & Snee, 1973).

When applying quality control techniques to monitor multiple fill-head processes, random and stratified sampling represent the two commonly used alternate sampling methods. A "random" sample is one in which all of the items in the population of interest have the same probability of being selected in the sample. A "stratified" sample is one which consists of samples taken from each strata within the population of interest (Burr, 1979; Grant, 1988). Too often in industry, sampling size and methods that are presumably logical, convenient, or simple win the day on the manufacturing floor (Caulcutt, 1995; Mayer, 1983; Osborn, 1990; Squires, 1982). Of course, if the process consists of just one fill-head, a rational sample selected from successive time periods of production would be the basis for the control charts ($\bar{x}$ and R) (Grant, 1988; Wadsworth, 1986). However, for multiple fill-head production processes, the sampling size and sampling method choices are not so simple.

Therefore, the objective of this paper is to investigate the effect of sample size on the sensitivity of control charts in detecting undesirable process variance due to assignable cause in a

multiple fill-head production process under alternate sampling methods. Note that the changes in process variance, as compared to those in process mean, are more detrimental for quality output and difficult to correct. These changes are typically monitored using R (range) control chart. We also examine the impact of process capability on the sensitivity of control charts. Since sampling distribution for range statistic is difficult to define, this paper uses Monte Carlo simulation to develop probabilities for detecting changes in process variance under alternate sampling methods. A discussion on sampling economics and limitations is also provided.

R-CHART FOR MODEL CASE SCENARIO

Today, many production processes involve multiple fill-heads for production systems. For example, a production process involving multiple fill-heads would require filling an equal number of vials at a time to a specified weight. When a batch of vials is filled, an equal number of new batches will replace it. Measurements are derived from an equal number of different strata. Now the goal is to design an effective control system for this production process using appropriate sampling size and sampling method (random or stratified). That is, in designing a process control system utilizing R-chart, a quality control professional must then determine whether it would be better to sample N vials at random from the process or sample one from each fill-head for a total of N.

Obviously, if the process is in control, the quality control practitioner can take either a stratified or random sample without any deleterious effects. When vials from different populations are mixed, they form a "pool of mixed product." Over a period of time, random samples of n vials are taken and measured. The R chart would then be determined after N samples are taken, informing the practitioner whether the process is in a condition of statistical control. If affirmed, the same chart could be used to monitor the process.

To be considered a "stratified" sample, each succeeding sample would have to be made up of N corresponding measurements that are from each fill-head at specified intervals of time. In the case where all fill-heads are the same, conventional control charts would present the data in such a way that the stratified sample would appear to be a random sample from a common population. But if the various strata have different averages, the stratification problem would become obvious through a conventional control chart which would focus data points in the direction of the center line.

There are three methods by which stratification difficulties can be remedied. First, Burr and Weaver (1949) and Westman and Lloyd (1949) recommend making corrections in the data for each strata (the difference between the mean and grand mean of corresponding strata in order to scale down the R). This approach offers, with a common mean, probability distributions that overlap for all strata. The second method requires creating distinct control charts for each strata. The third recommended method is to adjust the fill heads or the process physically toward a common average value, i.e., fix the strata to a common mean.

For the remainder of this paper, we assume that the different strata are the same in either of two situations: 1) the strata are corrected upon identifying a stratification problem or 2) the strata were identical from the start. A shift in a single stratum is probable even when the process appears to be in control. It is incumbent upon the quality control professional to detect the shift as quickly as possible. Therefore, this research concentrates on determining the role of sample size and sampling method (random or stratified) on the ability of R-chart to detect a single stratum shift in a multiple fill-head process with high sensitivity.

POWER CURVES FOR R CHART UNDER STRATIFIED SAMPLING

If a multiple fill-head production process in-control is suddenly affected by an assignable cause, it may cause a single fill-head to shift above or below the target value resulting in undesirable process variance. When this happens, for quality control purposes, it becomes important to detect the single fill-head quickly and take corrective actions. The following notations have been used to develop the power curves for R chart using Monte Carlo simulation for three different sample sizes under stratified and random sampling methods.

n = sample size
N = size of each strata
$\bar{x}$ = sample mean
R = Sample range
s = number of strata in the process
δ = a real no. indicating the magnitude of mean shift as a multiple of standard deviation
σ_x = standard deviation of each strata (Assumed to be equal)
T = initial mean of each fill-head (assuming they are equal or have been fixed in accordance with cases 1 or 3)
P_L = P(R < LCL)
P_U = P(R > UCL)

Assume that the process is in a state of statistical control and the mean of each probability distribution of four fill-head populations in the process is T with standard deviation σ_x. The LCL and UCL for the R-chart can be shown to be $D_3\overline{R}$ and $D_4\overline{R}$ (where D_3 and D_4 are constants depending on the sample size) respectively for n=ks where k represents the number of replicated sets of one observation taken from each of s strata. Without loss of generality, the standard normal distribution with T=0 and σ_x=1 will be used for the purpose of simulation.

(a) ***LCL and UCL values for the R-chart when sample size (n) = 4***

As stated earlier, the control limits for the R-chart are given by:

$$LCL = D_3 \overline{R} \quad \text{and} \quad UCL = D_4 \overline{R} \tag{1}$$

On substituting

$\overline{R} = \sigma_x.d_2$, in (1) we get:

$$LCL = D_3.\ \sigma_x.d_2 \ \text{ and } UCL = D_4.\ \sigma_x.d_2 \tag{2}$$

where σ_x=1 for standardized normal random variate.

Since $D_3 = 0$, $D_4 = 2.282$, and $d_2 = 2.059$, for n=4, from the quality parameter table in Grant et al. (1988), we get the control limits as:

$$LCL = 0 \text{ and } UCL = 4.698 \tag{3}$$

(b) ***LCL and UCL values for the R-chart when sample size (n) = 8***

For n=8, since the quality parameter values for D_3=0.136, D_4=1.864, and d_2=2.847, from equation (2) above the control limits for the R-chart can be determined to be:

$$LCL = 0.3872 \quad \text{and} \quad UCL = 5.3068 \tag{4}$$

(c) ***LCL and UCL values for the R-chart when sample size (n) = 12***

For n=12, the quality parameter values for D_3=0.283, D_4=1.717, and d_2=3.258. From equation (2) above we get the R-chart control limits as:

$$LCL = 0.922 \quad \text{and} \quad UCL = 5.594 \tag{5}$$

The control limits for the R chart given by equations (3), (4), and (5) will be used to determine the probability of rejection for developing power curves as a function of sample size.

Now, if the mean of a single fill-head shifts by ($\delta\sigma_x$) above or below the mean (T), the analytical expressions for a new common mean and standard deviation for the R-chart are difficult to develop since the statistical sampling distribution for range (R) is unknown. Hence, we propose to use Monte Carlo simulation to calculate the probability of detecting this shift in one stratum by R chart.

To perform Monte Carlo simulation for the above scenarios, a software package, Insight.xla (Business Analysis software for Microsoft Excel), is used. The three step approach that the software requires to perform the Monte Carlo simulation for stratified sampling plan is as follows:

Step 1. Build Model for S=4 Different Strata

The fill values for each of the three fill-heads in a state of statistical control were generated using a random number generating function, gen_Normal (0,1). This function randomly generated standardized normal values (z) with a mean of 0 and a standard deviation of 1 for each of the three fill-heads. For the fourth fill-head whose mean shifted by $\delta\sigma_x$, the corresponding function used was, gen_Normal ((δ, 1), where $0.5 \leq \delta \leq 10$. Expression for calculating the *Range* values for the four fill-heads were incorporated in the worksheet.

Step 2. Specify Simulation Setting

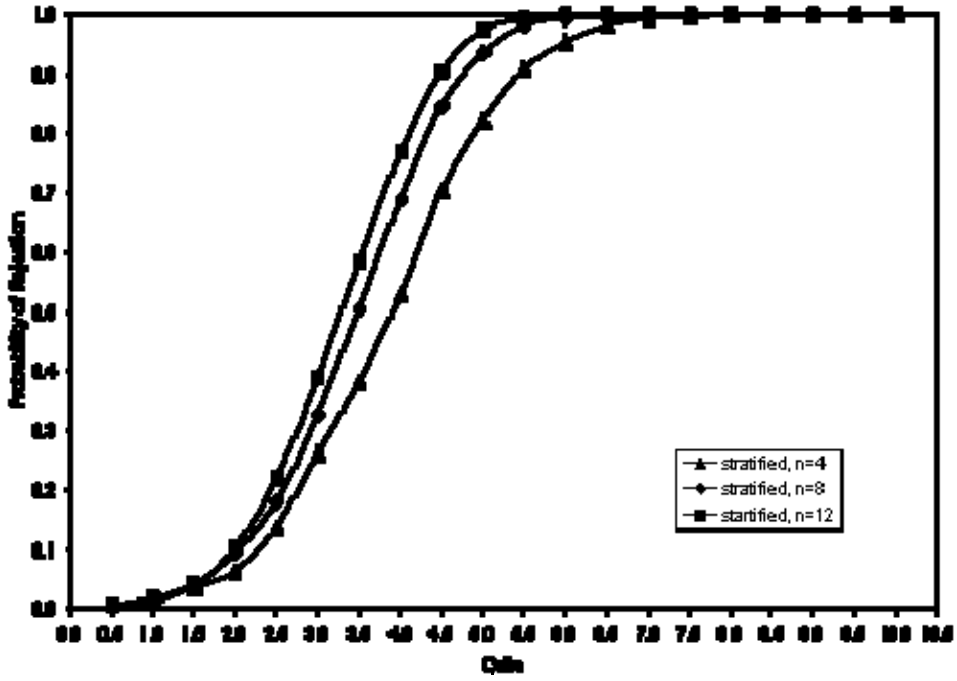
The primary simulation setting that the software requires is the "number of trials". The number of trials was determined by asking the question - what sample size would enable the determination of the probability (proportion) of rejection (P) to within 0.005 with 95% confidence for the overlapping probability distributions of all the four fill-heads. Knowing that the probability of an out of control indication is approximately 0.003 for the R control chart for a process in-control, a sample size of 2000 would allow an estimate for the probability of out of control indication within ± 0.0024 with 95% confidence.

Step 3. Run Simulation and Examine Results

It is assumed that the mean of the fourth fill-head shifts by $\delta\sigma_x$ such that $0.5 \leq \delta \leq 10$. The shift values (δ) are considered in increments of 0.5. A simulation run for 2000 trials was performed for each shift (δ) value and the resulting values for R recorded.

The R values so obtained in Step 3 were then compared with the control limits as given in equations (3), (4), and (5) and the proportion of values falling below LCL (P_L) and above UCL (P_U) determined. Figure 1 shows the power curves for R chart under stratified sampling as a function of sample size.

Figure 1: Power curves for R-chart under stratified sampling method as a function of sample size



POWER CURVES FOR R CHART UNDER RANDOM SAMPLING

Assume the same initial conditions of the process as given in section (4) for stratified sampling. Monte Carlo simulation is used to determine the sensitivity of R chart in detecting the shift in a single fill-head mean. The three step approach to run Monte Carlo simulation for random sampling plan is as follows:

Step 1. Build Model for S=4 Different Strata

Recall that the random sample involves selection of items in such a way that all the items in the population of interest have the same probability of being selected. As a result, the probability of a measurement being selected from each of the 4 fill-heads is 0.25. Since the distribution of 3-fill-heads constitutes a common population, it can be shown that for a random sample, the probability of a measurement being selected from the common population is 0.75 and that from the distribution of the shifted fill-head is 0.25. Hence, a random sample of size 4

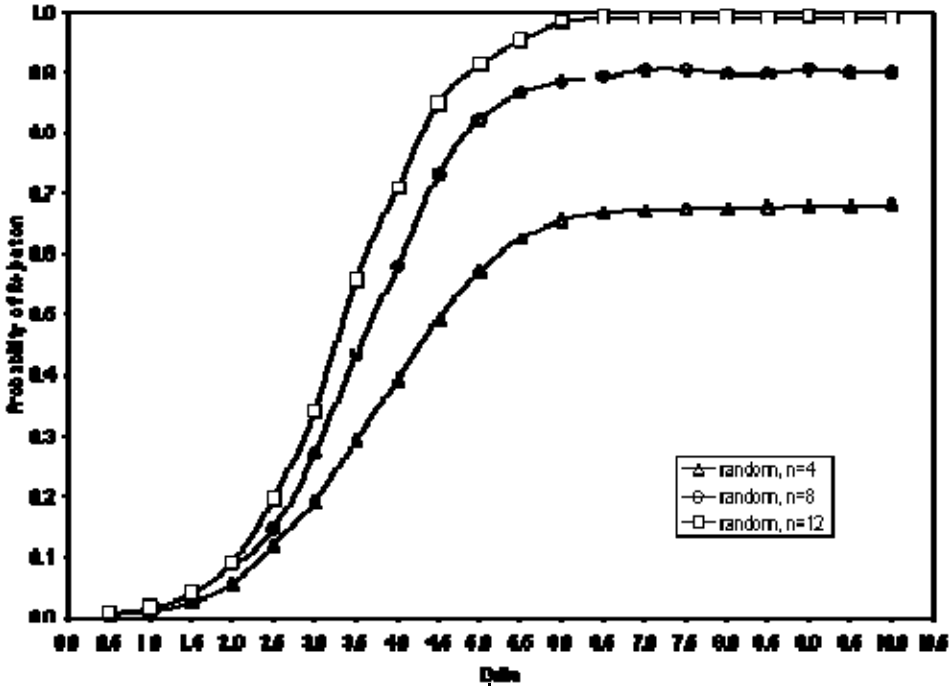
was generated by incorporating a logical random number generating function, *IF Rand () > 0.25, gen_Normal (0,1), gen_Normal (δ,1),* in 4 different cells of the worksheet.

This would ensure that all the 4 measurements in the random sample come from one or any combination of the 4-strata. The range formula for the random samples was also incorporated in the worksheet.

Steps 2 and 3. These steps were the same as under stratified sampling.

The proportion of R values which fall outside the LCL (P_L) and UCL (P_U) for different sample sizes were then determined. Figure 2 shows the power curve for R chart under random sampling plan as a function of sample size.

Figure 2: Power curves for R-chart under random sampling method as a function of sample size



DISCUSSION

Figure 1 shows the statistical power of detecting a fill-head shift in a multiple fill-head production process under stratified sampling method at different sample sizes. At lower levels of

process shifts where $0.5\sigma_x \leq \delta \leq 2\sigma_x$, the likelihood of detecting the shift on the first sample after the shift has occurred is insensitive to the sample size. For example, the probability of detecting $1\sigma_x$ shift is 1.15% and 1.85% at sample sizes of 4 and 12 respectively. Thus, even by tripling the sample size, the probability of detecting process shift increased only marginally. On the other hand, the sensitivity of detecting process shift such that $2.5\sigma_x \leq \delta \leq 5\sigma_x$ increases significantly with the sample size. For example, the probability of detecting a $3.5\sigma_x$ shift in stratum mean is only 32.4% at n=4 which increases to 58.45% at n=12. One may prefer to deal with the detection insensitivity at n=4 by taking multiple samples and identifying out of control signals by examining the resulting data patterns from those sequential samples. Even though this pattern based multiple sampling can increase the likelihood of detecting process changes, but at the cost of significant time delays involved and production of defective output during this period. Thus, if a $3.5\sigma_x$ shift is determined to be critical from quality perspective, one must increase the sample size for early detection of process shift. For a fill-head shift $(\delta) > 6\sigma_x$, the likelihood of detection begins to approach 100% and becomes insensitive to the sample size.

Figure 2 shows similar results for detecting a stratum shift under random sampling method as under stratified sampling. One noticeable difference under random sampling method is that sample size can play a significant role in detecting higher shift levels where $\delta > 6\sigma_x$. As an example, the detection probability for shifts $\geq 6\sigma_x$ under random sampling stabilizes around 67% when n=4 compared to 99% when n=12.

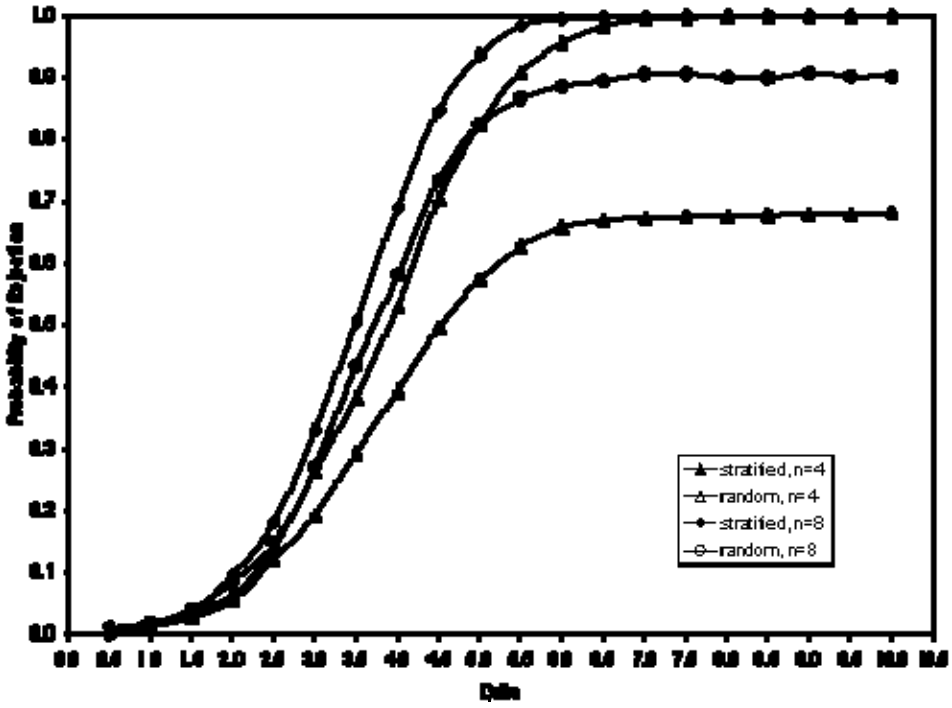
On comparing Figures 1 and 2, it is interesting to note that for a given sample size the probability of detection is equally sensitive under stratified and random sampling methods for smaller shift levels i.e. $0.5\sigma_x \leq \delta \leq 2\sigma_x$. However, for $\delta > 2\sigma_x$ stratified sampling, as compared to random sampling, tends to be more sensitive in detecting process shift. For example, at n=8, the probability of detecting a $3\sigma_x$ shift is 32.8% and 27% under stratified and random sampling methods respectively. Thus, for processes with smaller capability index where a minor process shift can cause serious quality problems, either sampling method based on cost and/or convenience may be used for detecting the presence of assignable causes. On the other hand, for processes with higher capability index where detection of higher shift levels becomes critical, a QC practitioner must use stratified sampling method.

Figure 3 shows the comparative power curves for R-chart under both the stratified and random sampling methods for sample sizes 4 and 8. It again confirms that the statistical power of detecting a stratum shift increases with increasing sample size. Since sample size is usually controllable, a practitioner must choose a sample size such that it detects a stratum shift with a specified level of probability. It is interesting to note that as sample size increases from 4 to 8, the relative difference in sensitivity for detecting process changes between stratified and random sampling methods diminishes significantly.

Figure 3 also shows the interplay between sampling cost (economics), sampling convenience, and the desired sensitivity of a control chart for shift detection. Note that for

stratum shift levels such that $3\sigma_x \leq \delta \leq 5\sigma_x$, random sampling at n=8 provides higher likelihood of shift detection as compared to stratified sampling method at n=4. Thus, if the marginal cost of sampling is low and random sampling is more convenient, then it may represent a better trade-off from a practical standpoint. Alternately, if the marginal cost of sampling is high due to either higher cost of each sampled unit or the destructive testing involved, then a stratified sampling at lower sample size may represent a better trade-off.

Figure 3: Comparative power curves for R-chart under random and sampling method for n=4, 8



CONCLUSION

This paper developed power curves for R chart under stratified and random sampling as a function of sample size to investigate the relative sensitivity in detecting a stratum shift from the target. There exists a lot of ignorance among QC practitioners about sample size and the

differentiating role it can play in detecting a process shift with a high level of sensitivity. The paper finds that if R-chart is the only tool used by a practitioner to monitor a multi-strata production process, then the stratified sampling method should be preferred. The paper also examined the moderating role of sample size in detecting process shifts as a function of process capability index. Lastly, the paper demonstrated the trade-off that exists between sampling economics, sampling convenience, and the desired level of probability for detecting a fill-head shift and a related need to strike a balance between the above trade-offs.

REFERENCES

Burr, I.W. (1979). *Elementary statistical quality control*. New York: Marcel Dekker, Inc.

Burr, I.W. & R.W. Weaver (1949). Stratification control charts. *Industrial Quality Control*, March, 11-15.

Caulcutt, R. (1995). The rights and wrongs of control charts. *Applied Statistics*, 44(3), 279-288.

Costa, A.F.B. (1997). $\bar{X}$ chart with variable sample size and sampling intervals. *Journal of Quality Technology*, 29(2), 197-204.

Davis, R.D., F.C. Kaminsky & R.J. Burke (1993). Detecting process shifts with x-bar charts. *Production and Inventory Management Journal*, 1st Quarter, 25-31.

Evans, J.R. & W.M. Lindsay (1993). *The management and control of quality*. West Publishing Company, 384-386.

Grant, E.L. & R.S. Leavenworth (1988). Statistical quality control. New York: McGraw-Hill, 6th Edition.

King, E.P. (1952). The operating characteristic of the average chart. *Industrial Quality Control*, November, 30-32.

Lanning, J.W., D.C. Montgomery & G.C. Runger (2002). Monitoring a multiple stream filling operation using fractional samples. *Quality Engineering*, 15(2), 183-195.

Mayer, R.R. (1983). Selecting control chart limits. *Quality Progress*, 16(9), 24-26.

Montgomery, D.C. *Introduction to statistical quality control*. John Wiley and Sons, New York, 3rd Edition.

Mortell, R.R. & G.C. Runger (1995). Statistical process control of multiple stream processes. *Journal of Quality Technology*, 27(1), 1-12.

Olds, E.G. (1961). Power characteristics of the control charts. *Industrial Quality Control*, July, 4-10.

Osborn, D.P. (1990). Statistical power and sample size for control charts - survey results and implications. *Production and Inventory Management Journal*, 4th Quarter, 49-54.

Ott, E.R. & R.D. Snee (1973). Identifying useful differences in a multiple-head machine. *Journal of Quality Technology*, 5(2), 47-57.

Palm, A.C. (1990). Tables of run length percentiles for detecting the sensitivity of Shewart control charts for averages with supplementary runs rules. *Journal of Quality Technology*, 22(4), 289-298.

Runger, G.C., F.B. Alt & D.C. Montgomery (1996). Controlling multiple stream processes with principal components. *International Journal of Production Research*, 34(11), 2991-2999.

Runger, G.C., F.B. Alt & D.C. Montgomery (1996). Contributors to a multivariate statistical process control signal. *Communications in Statistics - Theory and Methods*, 25(10), 2203-2213.

Scheffe, H. (1949). Operating characteristics of average and range charts. *Industrial Quality Control*, May, 13-18.

Squires, F.H. (1982). What do quality control charts control? *Quality*, November, 63.

Wadsworth, H.M., K.S. Stephens, & B.A. Godfrey (1986). *Modern methods for quality control and improvement.* New York: John Wiley.

Westman, A.E.R. & B.H. Lloyd (1949). Quality control charts for x and r adjusted for within-subgroup pattern. *Industrial Quality Control*, March, 5-11.

Wheeler, D.J. (1983). Detecting a shift in process average: tables of the power function for $\bar{X}$ charts. *Journal of Quality Technology*, 15(4), 155-170.

STUDY OF A FAST FACET MINING ALGORITHM USING DIFFERENT TYPES OF TEXT

Jose Perez-Carballo, California State University, Los Angeles

ABSTRACT

A fast clustering algorithm designed to mine facets from text was tested using several corpora containing different types of text including technical textbooks, a cookbook, web blogs, Wikipedia articles, and electronic mails. A "facet" is an aspect of a topic. For example the following sets would be reasonable facets in a cooking domain: ingredients (e.g. apples, cayenne pepper, chocolate), utensils (e.g. egg slicer, funnel, grater, potato masher), processes (e.g. basting, poaching, pressure cooking), etc. Several studies, have shown that interfaces that present results organized into categories or faceted hierarchies meaningful to users may help them make sense of their information problem as well as the information system itself. The algorithm studied here is based on the hypothesis that multi-word terms that appear in a similar grammatical context are likely to belong to the same facet. The results show a difference in performance of the algorithm depending on the kind of text. Text that tends to be more structured, such as a Java textbook, or Wikipedia articles, results in a larger number of the facets generated by the algorithm being judged useful by experts. Text that tends to be less structured and informal, such as blogs and email, results in less facets judged useful by experts.

INTRODUCTION

This paper describes tests of an algorithm called Fast Facet Identifier (FFID) described first in (Perez-Carballo, 2009). The tests presented here are intended to determine how well this algorithm performs with different kinds of documents.

The Fast Facet Identifier (FFID) Algorithm

FFID (Fast Facet Identifier) is an algorithm designed to find facets in large collections of documents. In that paper it was reported that: a) FFID discovered sets of facets from large corpora in a short time, and b) the facets discovered by FFID could be useful when building ontologies, as well as user interfaces designed to help users browse large collections of information.

Definition of a Facet

A "facet" is an aspect of a topic (Anderson and Perez-Carballo, 2005). Consider the following example taken from (Perez-Carballo, 2009). The following would be reasonable sets of facets in a collection of documents containing cooking recipes. Each set has a name or label (e.g. "ingredients") and members (e.g. apples, cayenne pepper, chocolate). Other facet sets in this domain would be: utensils (e.g. egg slicer, funnel, grater, potato masher), processes (e.g. basting, poaching, pressure cooking), dishes (e.g. a jiaco, bengal potatoes, bhuna khichuri, black-eyed peas, kale), herbs (e.g. basil, chicory, dill), etc.

The Usefulness of Facets

Traditional library classifications have always been based on facets. Ranganathan (1892-1972) described a facet system in the 1930s (Svenonius, 1992; Anderson and Perez-Carballo, 2005). A faceted hierarchical classification uses a set of category hierarchies (instead of only one). Each hierarchy corresponds to a different facet (dimension or property). Any topic can be described specifying each of the relevant facets. For example: look for all cooking recipes that involve facet ingredient "chicken", and facet process "grilled", and facet cuisine "spanish". Several researchers (English et al., 2002; Yee et al., 2003; Stoica et al., 2007, Hearst, 2006; Venkatsubramanyan & Perez-Carballo, 2007) have tested interfaces that use facets in order to support information exploration and browsing. Such tests have shown that user interfaces based on facets allow users to build much more effective queries, as well as support more effective browsing tools.

FFID: A Good tool to Find Facets

FFID finds good facets and it does it fast. A good facet discovering system should be able to identify multi-word terms such as "bengal potatoes" and "bhuna khichuri", decide that they belong in the same facet set, and determine a useful label for the set. In the case of the two previous multi-word terms a useful label could be "dishes". FFID uses a fast simplified clustering algorithm that allows the identification of hundreds of facet sets from a corpus of hundreds of thousands of sentences in a very short time (seconds). This makes FFID practical as a tool to be used by human experts when building interfaces but also as an algorithm that can be used in real time to provide browsing options to users.

Automatic Discovery of Facets

Knowledge engineers are often hired at great expense to combine their knowledge architecture skills with the domain knowledge extracted from subject matter experts in order to

create ontologies or faceted hierarchical classifications that can be used to enhance document classification and browsing tools. FFID is a tool that can make the job of knowledge engineers easier, faster and cheaper, as was demostrated in (Perez-Carballo, 2009). In a production environment FFID is given as input a large collection of documents (which we call a "corpus" in this paper). The corresponding output of FFID will be a collection of sets of facets, many of them with useful labels. A knowledge architect can use these facets in order to build the faceted hierarchical classification that is often the basis of browsing tools or interfaces. Also, FFID is fast enough to generate term suggestions in real time that can used to help the user to build better queries. The usefulness of FFID was validated by knowldge engineers and indexing experts in tests reported in (Perez-Carballo, 2009).

Tests of FFID

All the experiments described in (Perez-Carballo, 2009) were performed on corpora derived from Wikipedia. The question remained whether the same algorithm would prove equally useful on other kinds of corpora. Wikipedia articles may follow consistent formats that may provide grammatical patterns that affect the results obtained using FFID. The possibility also existed that the algorithm could have been optimized to work well with Wikipedia text but might not work as well with other kinds of corpora without significant modification. In this paper we report the results of a study of FFID that compares its performance on corpora other than those derived from Wikipedia articles.

PREVIOUS WORK

Several researchers (Hearst, 2006, Perez-Carballo, 2009) have described the advantages of the kinds of facets discovered by clustering terms vs labels obtained using techniques that cluster documents such as Vivisimo (Rivadeneira and Bederson, 2003). The facets obtained by clustering terms tend to be more useful to users (Hearst, op. cit.). The sets of related terms obtained by clustering documents may be misleading to users among other reasons, because different aspects of the topic may appear at the same level. Strictly speaking we could not even call such sets “facets”. For example composers, performers, and opera characters, may be grouped together because they may appear in the same document or fragment. While items, such as “Metropolitan Opera” and “State Opera Prague”, that might belong to the same aspect, or facet, of the topic may not be grouped together. A knowledge engineer would have created sets of conductors, singers, orchestras, venues, formats, etc. in order to facilitate browsing of the information by users. Such sets are really facets in the traditional sense of the word as used by information scientist and librarians (Anderson, J. D. & Perez-Carballo, J., 2005). The algorithm proposed in (Perez-Carballo, 2009) produces sets that are more like facets than the sets generated by clustering documents.

Usability studies that compare different kinds of groupings show that systems that use facet categorizations are more useful to people (Hearst, 2006). Users seem to find facets easier to understand and use. This is the same kind of categorization that librarians prescribe to help users browse through unfamiliar topics (Anderson & Perez-Carballo, 2005).

It is difficult and very expensive to build facet hierarchies purely by human intellectual labor without the help of automation. Using a pre-built classification (like the WordNet approach described in Stoica and Hearst, 2004) it is possible to miss topics that are important in the corpus, or emerging topics (such as current events and people) that may not appear in WordNet. On the other hand, Hearst's system (CASTANET) is very successful at extracting terms from corpora that do not provide a lot of context, such as a collection of scientific paper titles. FFID uses the richness and redundancy of the Web (e.g. a term can be found in many different sentences) in order to discover multi word terms and classify them into useful facets. It does this well even when it finds multi word terms that it has not encountered before.

Description of the FFID Algorithm

FFID uses a fast simplified clustering algorithm that allows the identification of hundreds of facet clusters from a corpus of hundreds of thousands of sentences in a very short time. In our experiments involving corpora of sizes of the order of hundreds of thousands of words, the processing time of each corpus was of the order of minutes.

The central hypothesis used by FFID is that multi-word terms (MWTs) that appear in a similar grammatical context are likely to belong to the same facet. For example, "violin" and "flute" (musical instrument) are likely to appear in similar grammatical contexts which are likely to be different from the contexts in which "funnel" and "grater" (cooking utensils) appear.

The following examples show sentences where MWTs (shown in bold face) appear in similar contexts:

*1- dip in **beer batter** and shake gently*
*2- dip in **seasoned flour** and shake gently to remove excess*
*3- garnish with **parsley**, if desired*
*4- garnish with **pecan halves**, if desired*
*5- Free scores by **Anton Bruckner** in the Werner Icking Music Archive*
*6- Free scores by **Luca Marenzio** in the Choral Public Domain Library - ChoralWiki –*

The previous sentences provide evidence that "beer batter" and "seasoned flour" belong to a facet, "parsley" and "pecan halves" to another, and "Anton Bruckner" and "Luca Marenzio" to yet another. FFID identifies MWTs and clusters them according to the degree of similarity of the contexts in which they appear. In some cases, FFID may be able to identify a useful label for the facet itself (e.g. Composer) as well as clustering the corresponding values (Vivaldi, Brucker, etc.) under that label.

The methods followed by the system described here are not very different from the methods that librarians and indexers are taught to use when they are charged to find facets from texts (Anderson & Perez-Carballo, 2005). Facets have a very strong correspondence with grammatical categories. Librarians are taught to create sentences that describe the topic of an item they want to classify. Each grammatical category they find is made into a facet. See (Perez-Carballo, 2009) for more examples.

A crucial aspect of FFID is the way in which it clusters the grammatical contexts. If each pair of contexts were to be compared using some similarity measure, the complexity of the algorithm will be O(n2). This means that if clustering a few thousand sentences takes on the order of minutes, clustering tens of thousands may take days. This is the reason why clustering algorithms are seldom practical. The algorithm used by FFID to cluster contexts can cluster hundreds of thousands of terms in seconds. In order to avoid O(n2) complexity, in FFID the contexts are processed first in order to transfom them into strings that, when sorted, will result in similar contexts being next to each other in the sorted list. The sorted list of contexts is then examined from beginning to end picking up the clusters of similar contexts. The resulting complexity is, thus, the complexity of sorting (O(nlog(n))).

The full algorithm used by FFID was described precisely in (Perez-Carballo, 2009). In the next paragraphs the basic structure of the algorithm is described for the purposes of this paper.

FFID consists of the following phases:

I. Term identification phase.

Compute multi-word terms (MWTs) from the corpus.
(1) using the Brill tagger (Brill, 1995), tag the corpus with part of speech information

(2) using predetermined regular patterns of part of speech (POS) tags, identify sequences of terms that are good candidates for multi-word terms (MWTs). For example: a sequence of singular nouns is a candidate for MWT. An example of the regular patterns used in the experiments described in this paper are specified here using the syntax of the Java language. In this example "JJ," "FW," "NN," etc. are POS tags.

(1)([^]+/JJ){0,1}([^]+/((FW)|(NN)|(NNS)|(NNP)|(NNPS))+
*(2)([^]+/(NNP |NNPS))((of/IN){0,1}[^]+((FW)|(NN)|(NNS)|(NNP)|(NNPS)))**

Some examples of MWTs identified by this method are the following:

Le/NNP Nozze/FW Di/NNP Figaro/NNP
Royal/NNP Concertgebouw/NNP Orchestra/NNP
Caucasoid/NNP North/NNP Africans/NNPS

(3) each MWT is saved with the sentence where it was found in a list called term-sentence list. The sentence will be used later to cluster the MWTs according to the context in which they appear in the corpus. The result of this phase is a list of all the terms found where each terms is preceded by the words that are the context in which the term was found in the corpus. Here is an example of a fragment from that sorted list. Each line in this list was derived from a phrase found in the corpus such as "Travel guide to Afghanistan from wikitravel."

-to/TO guide/NN Travel/NN;from/in wikitravel/nnp ./.;Afghanistan/NNP
-to/TO guide/NN Travel/NN;from/in wikitravel/nnp ./.;Africa/NNP
-to/TO guide/NN Travel/NN;from/in wikitravel/nnp ./.;Bali/NNP
-to/TO guide/NN Travel/NN;from/in wikitravel/nnp ./.;Belgium/NNP
-to/TO guide/NN Travel/NN;from/in wikitravel/nnp ./.;Chicago/NNP

The elements of the sorted list are of the form L;R;term, where L are the words (in inverse order) found in a sentence to the left of the term (e.g. "Travel guide to") and R are the words found to the right (e.g. "from wikitravel"). From this sorted list the system will be able to decide that all the terms in the example are likely to belong to the same facet. Notice that in this case cities, countries and continents are confused into the same facet because there are travel guides to them by wikitravel.

II. Normalization phase.

(1) apply a normalization process to the sentences. Regular patterns are used in order to reduce some sentences into a normal form. The details of all the regular patterns used is decribed in detail in (Perez-Carballo 2009). The patterns used in these experiments are listed in Table 1, belo. For example, the following two sentences:

- most deaf people within england speak british sign language, bsl, a sign language native to britain.

- list of countries where english is an official language

After tagging with POS tags and applying the regular expressions used to identify certain patterns, the following two string are obtained:

- ;PATTERN_ISA -language/nn ;BSL/NNP

- ;PATTERN_ISA -language/nn ;English/NNP

The normalized forms above are added to the term-sentence list. The normalization pattern "PATTERN_ISA" corresponds to label P5 in Table 1.

III. Facet identification phase

(2) Sort the term-sentence list. Some of the sentences may have been normalized by the previous step. In this example, after normalization, the terms BSL and English are near each other when the contexts (;PATTERN_ISA -language/nn) are sorted.
(3) Parse the sorted term-sentence list in order to identify MWTs with similar contexts. MWTs with similar contexts will be next to each other in the sorted list. In our running example, this step

identifies a facet that includes BSL and English, as well as another facet that includes Afghanistan, Africa, Bali, Belgium, and Chicago.

DISCUSSION OF THE METHOD USED BY FFID

For our method to cluster two MWTs together we only need to find some sentences in the corpus in which the MWTs appear in the same syntactic context. Thus the two sentences "Travel guide to France from Wikitravel" and "Travel guide to England from Wikitravel" provide evidence that "France" and "England" appear in similar contexts so they may belong to the same cluster. All other MWTs that appear in the same context are added to the same facet. Our hypothesis is that if two terms belong to the same facet then in a big enough corpus it should be possible to find several contexts that are similar enough for our method to cluster the two terms together. The more clusters we find that share the same two terms, the more likely it is that those terms belong to the same facet.

One of the advantages of the method described here is that it is possible to identify facets (e.g. "countries") and facet values (e.g. "France", "England") not known in advance. For example, places are recognized as places because they appear in the same kind of grammatical contexts. the sentence "Travel guide to Freedonia from Wikitravel" allows the system to decide that "Freedonia" is likely to belong to the same facet as any term X that appears in a sentence "Travel guide to X from Wikitravel".

The algorithm described above produces several different types of clusters depending on the pattern recognized. We identify the different types using labels P1, P2, etc. Table 1 shows the kinds of clusters used in the experiments with an example for each one. The precise regular patterns used to identify each of the cluster types is described in (Perez-Carballo, 2009) using the syntax of the Java programming language, here we give only a very simple example for each pattern. Notice that the system is able to generate a label (third column on Table 1) for each one of the clusters.

Table 1: Normalization patterns used in the experiments (with an example of each pattern)

id	Pattern Cluster Label	Example	Elements example (X)
P1	X of the Y	X of the ars nova	beginning, theory, time
P2	Y such as X	Effects such as X	chroma-key, filters, reverberation, whistles
P5	X is a Y	X is a language	BSL, English, Latin, Parsi
P6	Y's X	Machaut 's X	lyric output, lyric poems, masterpiece, narrative output, poems, poetry
P7	Y including X	Forms including X	canon, guaracha, zouk
P8	Y X	List of compositions -by X	Alban Berg, Alexander Scriabin, Antonín Dvořák, Arnold Bax, François Couperin

TEST DESIGN

What was measured: in order to test the effectiveness of the system with respect to different corpora (see next section), we measured the proportion of facet sets that were judged "useful" by a three expert panel. This is a measure of precision. It was decided, because of its difficulty, not to measure recall, i.e., of all possible facet sets mentioned in each corpus what percentage was identified, or for each set, what percentage of the possible facet values was correctly included in the appropriate set. (Perez-Carballo, 2009) describes experiments involving a single corpus about music taken from Wikipedia that was used to measure the proportion of facet sets judged useful by expert judges. The experiments described in that paper showed that the algorithm was fast and effective in order to help human experts identify good facets. The experiments in this paper test the performance of the algorithm with different kinds of text.

How it was measured: for each corpus tested, a file was created that contained at most 20 clusters for each type of cluster generated by the system. The set of 20 clusters per label was selected at random from all the clusters generated by the system. This file was shown to a group of 3 experts that included 2 knowledge engineers and a professional indexer. They were asked to look at the sets and judge them "not useful at all" (score 0), "somewhat useful" (score 1), "very useful" (score 2). They were asked to interpret "useful" as meaning "this cluster would have helped them shorten the time and effort required to create facet taxonomies by hand". The experts worked together to score the clusters and reach a judgment by consensus. This process is similar to the team work of knowledge engineers in the real world.

When the algorithm was described above, the patterns that are used to conflate several different contexts were listed in Table 1. Some of these patterns may generate useful clusters more often than others. In the experiments described in this paper we measured the difference in quality of these different patterns. The purpose of the tests is to find out whether our methods to extract facets works better for some kinds of text than with others.

After all clusters were generated from each one of the corpora used in the experiments, a label was given to each cluster depending on its normalization pattern. The labels are P1 to P7 and are shown on Table 1 with their corresponding patterns. P8 corresponds to clusters generated without using a normalization pattern.

Here are the instructions given to the judges:

In this study you are going to help us find which kind of word sets you think are more useful to people when they are trying to organize information. This file contains sets of words created using different techniques. Each set looks like the following example:

_____*24: P5_15 -- score:*

Label- ... is a city
London
Los Angeles
Paris

The first line after the line that has the word "score" is the label of the set. Sometimes the label has three periods. In the example above the periods indicate that you could use the terms in sentences such as: "London is a city". Other labels may look like:

25: P6_1 -- score:
Label- set of cities
London
Los Angeles
Paris

Other labels may not seem very meaningful, but remember you are going to be scoring the quality of the sets not the quality of the labels. So you may find the set very useful although its label may not be very clear. Your job is to give to each set a score from 0 to 2 where:

0: not useful at all
1: somewhat useful
2: very useful

DESCRIPTION OF THE CORPORA USED

In order to compare the performance of the system with respect to several different kinds of text eight different collections, or corpora, were used. The following is a list of the text collections we used in order to perform the tests reported in this paper:

name: CASTANET_ES
description: cooking recipes corpus provided by Marti Hearst
files: 3538
total size: 2.9 MB
number of words: 617,267
source: Marti Hearst's Castanet project (Stoica et al., 2007; Stoica and Hearst, 2004).

name: COOK
description: cooking book from Wikibooks
files: 2251
total size: 5.7 MB
number of words: 1,142,157
source: http://en.wikibooks.org/wiki/Cookbook (March 2011).

name: DISH
description: spidered from the home page of "The Daily Dish" - a blog by Andrew Sullivan
files: 659
total size: 2.7 MB
number of words: 486,784
source: http://andrewsullivan.theatlantic.com/ (March 2011).

name: HUFF
description: files spidered from the home page of "The Huffington Post"
files: 105
total size: 1.7 MB
number of words: 339,623
source: http://www.huffingtonpost.com/ (March 2011).

name: JAVABOOK
description: Java book from Wikibooks
files: 114
total size: 1.7 MB
number of words: 327,117
source: http://en.wikibooks.org/wiki/Java_Programming (March 2011).

name: JDA
description: Book with James D. Anderson (Anderson, J. D. & Perez-Carballo, J. 2005)
files: 22
total size: 1.4 MB
number of words: 247,660
source: http://comminfo.rutgers.edu/~carballo/ird2005.html (March 2011)

name: MUSIC
description: articles from Wikipedia spidered from article about "classical music"
files: 1158
total size: 15.5 MB
number of words: 2,969,377
source: http://en.wikipedia.org/wiki/Classical_music (March 2011).

name: CSMAIL
description: email messages
files: 6
total size: 2.9 MB
number of words: 643,312
source: 619 college email messages from my mailbox

ANALYSIS OF THE TESTS

The examples given to the judges in their preliminary instructions (see previous section) are very clear but some of the clusters the experts actually found in the output of the system required some domain knowledge. For example, one cluster found by the judges contained the terms: "Kilikia" and "Kotayk" that none of the experts had ever seen before. The label generated by the system suggested that these were beers. The experts googled the terms and decided that indeed the terms were different kinds of beers so they decided this was a good facet. This is one of the advantages of this system. It will discover associations that the experts were not aware of, even between terms that were unfamiliar to the experts.

In cases such as the one described in the previous paragraph the cluster label generated by the program aided the judges by helping them realize the connection between the terms. In other cases the label did not help. In those cases the terms may have seemed related to the judges but they would give the set a low mark because the label did not seem good or not related of the terms.

It often was difficult for the judges to decide whether a set would be useful or not, even if the terms seemed to belong to the same category. This is often the case in the work of knowledge engineers and indexers. How useful a facet is may depend ultimately on the situation and the community of users that are expected to use the facet. The first step in creating any classification is to determine what are the categories of topics and features that will be more useful to the targeted user community (Anderson and Perez-Carballo, 2005). Consequently, an automatically generated faceted classification should be filtered and refined in order to include all and only the facets that are expected to make it the most useful and usable for its users. Even if the output of FFID has to be refined by an expert, it would still save a lot of time, effort, and money to start from the facets identified by a system such as this. It is also likely that the expert

would not have to have a lot of domain expertise in order to be able to refine the facets that are identified automatically.

Table 2 shows the data collected during the tests. Since the corpora are of different sizes we used the sum of all scores divided by the number of clusters as a measure of how good the clusters extracted from that corpus are. That number can be used to compare across corpora even if the number of clusters extracted is different. The maximum value for that parameter would be 2 if all clusters were judged "very useful". The minimum value would be 0 if all clusters were judged "not useful at all". This is because the judges were asked to give each cluster a score from 0 to 2 where: 0 means "not useful at all," 1 means "somewhat useful," and 2 means "very useful."

Table 2: Test results

Corpus name	size (MB)	total nr clusters	clusters per MB of input	nr of clusters in sample	best cluster type (score)	score for clusters in sample
CASTANET	2.9	360	124.14	60	p8 (1.85)	1.6
JAVABOOK	1.6	174	108.75	60	p5 (1.83)	1.52
JDA	1.4	259	185.00	49	p6 (1.8)	1.45
COOK	5.7	782	137.19	82	p8 (1.75)	1.37
MUSIC	15.5	1555	100.32	120	p6 (1.9)	1.32
CSMAIL	2.88	541	187.85	54	p8 (1.8)	1.2
HUFF	1.7	122	71.76	51	p6 (1.3)	1.08
DISH	2.7	60	22.22	44	p6 (1.33)	0.8

Table 2 presents the results of the judgments on all the corpora tried in the experiments. The number of clusters generated for the DISH and HUFF corpora is much smaller (22.22 clusters per MB of input and 71.76, respectively) than for other corpora (e.g. 124 for CASTANET). DISH and HUFF contain blog style text. In both cases, the readers of the blog write comments. The kind of text readers write is in general short, very diverse, and not always grammatical (in some cases not even relevant to the original article). The CSMAIL corpus contains about 600 email messages sent to my school account. The FFID algorithm was able to generate more clusters per megabyte (187.85) from CSMAIL but the number of useful clusters seems small compared to other corpora, such as the Java book. The three corpora that resulted in the least number of useful clusters (DISH, HUFF, and CSMAIL) contain text that tends to be more diverse and less formal. Of them, CSMAIL is a little more focused (on university issues), and it produced more useful clusters than the other two (HUFF and DISH), which are much more diverse, since they cover the current world and political news of the day, and contain very informal, blog-style, user comments. COOK and MUSIC are collections of documents about specific topics: a cooking book and a collection of articles about the history of western classical music, respectively. These two, more focused corpora, generated more clusters of higher quality. Finally the JDA, JAVABOOK, and CASTANET corpora represent collections about specific topics. In all three cases the topics contain some areas in which syntax is important. The JDA

book contains many examples about the syntax of indexing, the JAVABOOK collection contains examples of java code, and the CASTANET corpus is a collection of highly structured recipes. Each one of the CASTANET recipes follows the same form: name, ingredients, preparation. In the case of these three corpora two things happen: the topic is focused, and there are many passages that follow a very formal syntax. Consequently, our algorithm finds many clusters that were judged useful. For example, one of the clusters found by the judges contained all the keywords of the Java language. Other clusters contained sets of ingredients, kinds of dishes, etc.

CONCLUSIONS

Corpora with a lot of structure (i.e. patterns that appear often in the text) and about the same topic will produce more clusters that will be more likely to be useful. A recommendation for the use of the FFID algorithm in production situations is to segment the documents that will be fed to it into collections that are topically consistent. For example do not mix articles about world politics with articles about the Java language. The resulting facets will be more likely to be useful sets if the documents are more consistent. The kind of language used in the documents will also affect the number of useful facets produced. Topics that are more likely to use structured text, such as examples of programming language code, or a collection of documents that all follow the same format, such as recipes, will be more likely to generate useful facets. Finally, text that does not have any topical unity, written by a diverse group of writers, and not following any kind of format (such as emails or blog comments), will result in the less amount of useful facets. But even for that kind of text, FFID was shown to produce enough useful facets by our group of indexers and knowledge engineers to make it a useful enough tool in the process of creating classification taxonomies.

ACKNOWLEDGEMENTS

The author would like to thank Marti Hearst for sharing the CASTANET data that we used to run comparisons between FFID and CASTANET. James Anderson read earlier versions and provided, as usual, valuable advice.

REFERENCES

Anderson, J. D. & Perez-Carballo, J. (2005). *Information retrieval design*. St. Petersburg, Fl: University Publishing Solutions.

Brill, E. (1995). Unsupervised learning of disambiguation rules for part of speech tagging. *Proceedings of the 3rd Workshop on Very Large Corpora*, 1-13.

English, J., Hearst, M. A., Sinha, R., Swearingen, K., & Yee, K.-P. (2002). Hierarchical faceted metadata in site search interfaces. *Conference on Human Factors in Computing Systems,* 628-639.

Hearst, M.A. (2006) Clustering versus faceted categories for information exploration. *Communications of the ACM,* 49(4), 59-61

Perez-Carballo, Jose. (2009) Semi-automated identification of faceted categories from large corpora. *Academy of Information and Management Sciences Journal (AIMSJ).* 12(2), 91-111.

Rivadeneira, W. & Bederson, B. (2003). A study of search result clustering interfaces: Comparing textual and zoom able user interfaces. *HCIL technical report*, (Technical Report HCIL-2003-36).

Stoica, E. & Hearst, M. A. (2004). Nearly-automated metadata hierarchy creation. *The Companion Proceedings of HLT-NAACL'04.* 117-120.

Stoica, E., Hearst, M. A., & Richardson, M. (2007). Automating creation of hierarchical faceted metadata structures. *Proceedings of NAACL HLT*. 244–251.

Svenonius, E. (1992). Ranganathan and classification science. *Libri*, 17(42):176–183.

Taylor, R. (1968). Question-negotiation and information seeking. *College and Research Libraries*, 29(3):178–194.

Venkatsubramanyan, S. and Perez-Carballo, J. (2007). Techniques for organizing and presenting search results: A survey. *Journal of Information Science and Technology (JIST)*, 4(2). p45

Yee, K.P., Swearingen, K., Li, K., and Hearst, M. A. (2003). Faceted metadata for image search and browsing. *Proceedings of the SIGCHI conference on Human factors in Computing.* 401-408.

COORDINATING PRICING AND INVENTORY PURCHASING DECISIONS OF A SUPPLY CHAIN FOR AN E-TAILER IN FACE OF QUANTITY DISCOUNTS

Patrick Lee, Fairfield University

ABSTRACT

Traditional models have assumed that when a product's demand curve is known and stationary, a retailer of a product would find it optimal to replenish the inventory with a fixed quantity, and to sell the product at a fixed price. Current studies have shown that an E-tailer is better off if he/she would increase the price during an inventory replenishment cycle when price and demand elasticity is present. This research shows how this strategy can be adopted for E-tailers in face of quantity discounts offered by the manufacturer of a supply chain to encourage sales, and how retailers can be misled to adopt a suboptimal policy by using the cost-based model alone. We further illustrate that a continuously increasing price policy can lead to significant increase in profits over a fixed pricing policy.

Key words: inventory model, supply chain, E-tailers, EOQ, continuously increasing price, profit maximization

INTRODUCTION

Ford Harris's (1913) study on batch sizing started the studies on inventory policy a century ago. The Harris Model has then been forgotten until the idea was later published by Wilson (1934) in Harvard Business Review. Following the Review, the Harris's batching rule is now known as the Wilson Economic Order Quantity (EOQ) as it applies to inventory control.

This classic model assumes that a retailer buys a product at a constant unit cost, incurs a fixed cost per order, stores the product at a constant inventory carrying cost per unit per year, and faces a deterministic and constant demand rate over an infinite horizon, the retailer's optimal strategy is to buy a fixed quantity every time he or she replenishes the inventory. Ignoring inventory related costs, classical price theory tells us that when a product's demand is price sensitive but the demand curve is known and stationary, the retailer's optimal strategy is to charge a single price throughout the year. Although Whitin (1955) was the first one to integrate the concepts of inventory theory with the concepts of price theory, however, he did not formally investigate the simultaneous determination of price and order quantity decisions of a retailer.

Kunreuther and Richard (1971), perhaps the predecessors of supply chain management, then showed that when demand is price elastic, centralized/coordinated decision-making (using

simultaneous determination of optimal price and order quantity) was superior to the common practice of decentralized decision-making whereby the pricing decisions were made by the marketing department while the order quantity decisions were made by the purchasing department independently. Although Kunreuther and Richard (1971) were perhaps unaware of Whitin's (1955) paper, their model was very similar to Whitin's (1955) model. Assuming a known and stationary demand curve along with the conditions of the EOQ model, Arcelus and Srinivasan (1987, page 173) asserted: "given constant marginal costs of holding and purchasing the goods, the firm will want to maintain the same price throughout the year". Again, they assumed a fixed single selling price throughout each inventory cycle. What they did not realize is that, even though marginal holding costs are constant per unit, a firm's holding costs at any particular time within an inventory cycle are a function of inventory on hand, which itself is a function of the time from the beginning of the inventory cycle.

Since Whitin's (1955) work, numerous authors (Tersine and Price, 1981; Arcelus and Srinivasan, 1987; Ardalan, 1997; Hall, 1992; Martin, 1994; Arcelus and Srinivasan, 1998; Abad, 2003) have used Whitin's (1955) and Kunreuther and Richard's (1971) models as foundations to their own models. But none of these authors have ever questioned Whitin's (1955) and Kunreuther and Richard's (1971) assumption that the retailer's optimal strategy would be to sell the product at a fixed price throughout the inventory cycle. The fact that Whitin's (1955) and Kunreuther and Richard's (1971) assumption of a single price throughout an inventory cycle leads to suboptimal profits for the retailer is due to declining carrying costs as a function of time. It seems that any optimization model allowing a retailer with a price-insensitive demand to set the selling price arbitrarily would push the price to infinity. In other words, in that situation, price is not seen as a decision variable for any mathematical model. Given an arbitrary price (and corresponding demand), the retailer's only strategy is to minimize his inventory ordering and holding costs by using the EOQ model.

Considering a situation of price sensitive demand, Abad (1997; 2003) found that, in the case of a temporary sale with a forward buying opportunity, a retailer's optimal strategy is to charge two different prices during the last inventory cycle of the quantity bought on sale—a low price at the beginning of the inventory cycle and a higher price starting somewhere in the middle of the cycle. Yet, Abad (1997; 2003) did not consider a similar strategy in every regular inventory cycle of a product with price sensitive demand. Inspired by Abad's (1997; 2003), Joglekar, Lee and Farahani (2008), showed that a continuously increasing price strategy: charging a relatively low selling price at the beginning of an inventory cycle when the on-hand inventory is large, would lead to higher profit, abide the gain is relatively small. Subsequent study by Lee and Joglekar (2012) for gradual inventory production/replenishment cycle showed that significant gains on profit can be acheived by taking advantage of the demand elasticity.

With the widespread use of revenue management or yield management techniques (Feng and Xiao, 2000; McGill and Ryzin,1999; Smith and Leimkuhler, 1992; Talluri and Ryzin, 2000; Weatherford and Bodily, 1992) in the airline, car rental, and hotel industries today, a time-

dependent (or dynamic) pricing strategy has become commonly adopted. Revenue management techniques are typically applied in situations of fixed, perishable capacity and a possibility of market segmentation (Talluri and Ryzin, 2000). In recent years, retail and other industries have begun to use dynamic pricing policies in view of their inventory considerations. The recent recession has brought forth dynamic pricing to a new light.

As retail sales dropped, retailers were facing unusual built-ups of inventory that would lead to order cancellations affecting all parties of the supply chain. To counter such demand issues, wholesalers started to offer quantity discounts to retailers to boost sales. Hence, the E-tailers are presented with an opportunity to review their inventory policy, and pricing.

In addition, supply chain literature has shown that coordination of supply chain activities such as pricing (marketing) and inventory policies (production/distribution) would yield greater rent than acting as individuals. The case of bullwhip effect is well-known when supply chain partners that are hedging against risks and variations of demands, independently adjust the ordering quantities wildly to prevent stock-outs leading to magnification of inventory orders up the supply chain. The results, of course, would lead to huge inventory costs of retailers, and wholesalers, and production scheduling problems for manufacturers, and suppliers.

In the situation when manufacturers use a quantity discount to encourage larger sales, literature shows that the optimal ordering quantity to minimize inventory costs is derived from balancing setup cost, holding cost, and costs of inventory at various discount levels. However, with the supply chain approach of coordinating sales demands, and inventory, an E-tailer would benefit more if he/she would focus on profit optimization.

Joglekar, Lee and Farahani (2008) proved that by using profit maximization, one would enjoy higher profits by simply switching to price optimization in the EOQ (Economic Order Quantity) context. Furthermore, they showed that E-tailers who can easily change price via computing systems can obtain higher profits by changing price continuously. Although, the profit increase in an original EOQ inventory models is relatively small, they do provide a tool on price increasing schemes that can lead to higher profits.

This research would continue the investigation of profit maximization approach on quantity discounts when manufacturers would offer their supply chain retailers to achieve larger sales. In the next section, we shall review the cost model with quantity discounts (or bumpy breaks). Following that, we shall recapitulate the fixed price model, and extend the model for quantity breaks. The fourth section will show how the continuously pricing model can be implemented in face of quantity discounts. Finally, we would illustrate with numerical examples how all three models would perform under the same set of data, and offer some concluding remarks on cost models versus profit based models, and possible future research directions.

COST BASED QUANTITY DISCOUNT MODEL

The traditional EOQ models with quantity discount can be easily found in any operation or inventory management text. The general formulation is to find the optimal ordering quantity within each break, and when that quantity does not happen within the (price) breaks, we would bring the corresponding optimal ordering quantity up or down to the break point. The total inventory cost per period would then equal to the sum of inventory cost plus the cost of the goods sold per period. The following notations are based on fixed quantity review policy:

C_i = retailer's known and constant unit cost of buying the product within the quantity discount break, B_{i-1} and B_i,

Q_i= ordering quantity, when unit price is C_i, per cycle,

S = retailer's known and constant ordering cost per order,

I = retailer's carrying costs per dollar of inventory per year,

D = retailer's annual demand per period, and

B_i = quantity discount break, i, when ordering quantity $B_i > Q_i \geq B_{i-1}$, with corresponding unit cost of C_i.

To obtain the optimal ordering quantity with discount, one would find the optimal, Q_i, by minimizing, TC, the total cost per period, i.e., we minimize

$$TC = (D/Q_i)S + ½(Q_iIC) + DC_i \text{ for all } i. \quad (1)$$

Notice that with constant demand/consumption, the fixed quantity review policy is the same as fixed periodic review policy with B_i being given as ordering cycle time, T_i where $T_i = B_i/D$.

Thus, the periodic review model would be to minimize, TC with respect to T, such that:

$$TC = S/T + ½(DTIC_i) + DC_i, \text{ where } T = (2S/DIC_i)^{1/2} \text{ for all i.} \quad (2)$$

Note that if we simply switch the cost minimization model to maximization of profits via optimal price, such as maximizing ($P^*D - TC$) where P is the price for selling each product, the solution would set P to be infinite, rendering the solution impractical.

FIXED PRICE QUANTITY DISCOUNT MODEL

Both papers of Whitin "(1955), Kunreuther and Richard (1971) consider a situation where all the other assumptions of the EOQ model are valid but demand is price sensitive, with a known and stationary demand curve. Whitin's (1955) notation is different from Kunreuther and Richard's (1971) notation. Although the model is applicable to any form of the demand function, for simplicity, we use a linear demand function.

Let the notations of the previous section hold. In addition, we assume,

$D_1 = a - b\,P_1$, where a and $b > 0$, and a is the maximum demand per period,
P_1 = retailer's selling price per unit in this model,
t = time elapsed from the beginning of an inventory cycle,
T_1 = duration of an inventory cycle,
Q_{1i} = retailer's order quantity per order in this model when unit cost is C_i, and
Z_1 = retailer's profit per period.

It is assumed that,

$P_1 > C_i$, (for all i), and D_1 = retailer's annual demand as a function of the selling price, P_1, hence, $D_1 = a - b\,P_1$, where a and b are nonnegative constants, a representing the theoretical maximum annual demand (at the hypothetical price of \$0 per unit) and b representing the demand elasticity (i.e., the reduction in annual demand per dollar increase in price). Although P_1 would remain constant throughout a cycle, we choose to express it in an affine function form to be consistent with the price increasing model in the next section.

Note that since D_1 must be positive for the conceivable range of values of P_1, $a > bP_1$ for that range of values of P_1, and since $P_1 > C_i$, it follows that $a > bC_i$.

The fixed pricing model of Joglekar, et al (2008) for the constant cost, C, inventory model; with Q being the ordering quantity per order, then $Q = D_1/T_1$, the profit per period is given by,

$$Z_1 = (P_1 - C - ICT_1/2)(a - b\,P_1) - S/T_1 \tag{3}$$

Differentiating Z_1 with respect to P_1 and T_1, the first order conditions for the maximization of this function are:

$$P_1 = (1/2)(a/b + C) + (ICT_1/4), \tag{4}$$

where T_1 is a positive root of the following equation:

$$T^3 - 2[(a - bC)/(bIC)]T^2 + 8S/(I^2C^2b) = 0. \quad (5)$$

Substituting T_1 into (3), we obtain the profit per period as

$$Z_1 = (b/4)[(a/b) - C - (ICT_1/2)]^2 - S/T_1 \quad (6)$$

By using Excel's Solver®, one would be able to find the optimal T_1, and hence, the optimal price from equation (4), and the corresponding demand $D_1 = a - bP_1$, per period, and the optimal ordering quantity Q.

To adopt this model for ordering quantity with discount, we substitute C by C_i, for equation (6) above. From the resulting optimal T_1, we can compute the corresponding D_1, and the optimal ordering quantity, Q_{1i}. When C_i is not eligible for discount under the above conditions due to the breaks, we then, will proceed like the cost based model to utilize the optimal ordering quantity at the breaks, either B_{i-1}, or B_i, accordingly.

We derive the optimal pricing, P_1, at price breaks in the following for B_i, in general. With ordering quantity of B_i, the inventory cycle time, T_i, is given by

$$B_i = D_1T_1$$

or, $T_1 = B_i/(a - bP_1)$, after substitution of $D_1 = a - b\,P_1$. (7)

By substituting (7) into (3), the profit per period,

$$Z_1 = (aP_1 - bP_1^2 - C_ia + bC_iP_1) - IC_iB_i/2 - Sa/B_i + SbP_1/B_i$$

Differentiating Z_1 and setting the first order condition to zero for optimal P_1, we obtain,

$$P_1^* = (a + bC_i + Sb/B_i)/(2b) \quad (8)$$

CONTINUOUSLY INCREASING PRICE MODEL WITH QUANTITY DISCOUNTS

Joglekar, et al (2008) developed the model of continuously increasing price within an inventory cycle. We recapitulate their model below for E-tailers. We retain all the notations of the previous section for the fixed pricing model. We also add the following notation:

t = time elapsed from the beginning of an inventory cycle,

$P_2(t)$ = the retailer's selling price at time t,
= $f + gt$, where f and g are nonnegative decision variables,

T_2 = duration of an inventory cycle,

Q_{2i} = retailer's order quantity per order in this model when unit cost is C_i, an

Z_2 = retailer's profit per period.

Thus, the retailer's annual profit under this model is given by maximizing Z_2, with respect to f, g, and T_2.

$$Z_2 = (a - bf)(f - C_i - IC_iT_2/2) + (a - 2bf + bC_i)gT_2/2 + bg(IC_i - g)T_2^2/3 - S/T_2 \quad (9)$$

Notice that the model is a periodic review model similar to the fixed price model while the cost based model is a quantity review model. However, since the cost based model is shown to be easily transformed to a periodic based model, this would make our comparisons easier later.

As suggested by Joglekar, et al (2008), we can utilize Excel Solver® to find f, g, and T_2 simultaneously. The optimal ordering quantity per order Q_{2i}* is given by,

$$Q_{2i}^* = (a - bf)T_2 - (bg/2)T_2^2 \quad (10)$$

For using this model with quantity discount breaks, we, again, have to consider the scenario when Q_{2i}* is not in the range of corresponding price breaks. Under such circumstances, we would have to adjust the ordering quantity to the break point, B_i, as in the previous two models. That is, we would maximize (9) with respect to f, g, and T_2, subject to the constraint,

$$B_i = (a - bf)T_2 - (bg/2)T_2^2 \quad (11)$$

NUMERICAL EXAMPLES

Consider the following situation with price breaks, when the retailer's cost of a product is C_1 = \$7 per unit when ordering quantity is below B_1 = 2000 units/order, C_2 = \$6.50 per unit when ordering quantity is between B_1 = 2000 and B_2 = 6000, and C_3 = \$6.00 per unit when ordering quantity is greater than or equal to B_2 = 6000 units. We further assume that the ordering cost, S, is \$400/order, the inventory carrying cost rate, I, is 0.4 per unit per year, and for price and demand elasticity parameters, a and b, are 50,000 and 5,000, respectively. Notice that for the cost based model, demand D is given and constant over time, while for the price-based models, D_1 and D_2 are dynamic depending on C_i. In order to be able to make comparisons, we arbitrarily pick D_1* when unit price is \$7 as annual demand D of the cost-based model.

Tble 1. Comparisons of Optimal Policies of Cost-Based Model, Fixed Price Model, and Continuously Increasing Price Model															
Assumptions: S = \$400/order, I = 0.4/per unit per year, a = 50000, b = 5000. Discount breaks are: \$7 per unit for Q < 2000, \$6.5 per unit for $Q \leq 2000 < 6000$, and \$6.00 per unit for $6000 \leq Q$.															
	Cost-Based Model			Fixed Pricing Model					Continuously Increasing Price Model						
C_i	D	Q	TC	D_1	Q_1	T_1	P_1	Z_1	D_2	Q_2	T_2	f	g	$P_2 = f + g*T_2$	Z_2
\$7.00	6781.597	1391.977	51368.71	6781.597	1391.977	0.205258	8.643681	7249.245	6035.031	1416.315	0.209281	8.5	1.4	8.792994	7284.319
\$6.50	6781.597	2000	48036.7	8250	2000	0.242424	8.35	11012.5	7434.434	2000	0.243169	8.196997	1.299986	8.513113	11054.01
\$6.00	6781.597	6000	48341.69	9833.35	6000	0.610168	8.0333	12138.89	9833.333	6000	0.617566	7.686344	1.200011	8.427429	15364.95

Table 1 shows the optimal solution for each of the models presented in this paper. Note that for the optimal ordering quantities Q, Q_1 and Q_2, the actual solution when C_i is \$6.50 or \$6.00 per unit are in fact smaller than the required discount breaks. Hence, the optimal solutions for those ordering quantities are set at break points, B_i. The shaded cells represent the best option for each model, respectively.

For the cost-based model, we minimize TC, the total inventory cost plus the cost of products per year. The optimal policy is to purchase 2000 units per order for a total annual cost of \$48,036.70. For both the fixed price model and the continuously increasing price model, we elect to order 6,000 units per order instead, contributing to \$12,138.89, and \$15,364.95 in profits/year, respectively. Thus, the continuously increasing price model garners an increase of \$3226.06 in annual profit, or 26.58 % over the fixed price model.

Although EOQ models based on costs are generally insensitive to changes of parameters, such as ordering cost, holding cost, and so on. Table 1's cost-based model shows that the difference in total cost between ordering 2000 and 6000 units per order is a mere 0.635 % difference. Unfortunately, the same logic does not apply to the profit maximization models. Should we have adopted the 2000 units per order as our inventory policy, based on fixed pricing model, the retailer would lose 9.272 % profit, while compared with using the continuously increasing profit model, a retailer would lose a drastic 28.057 % of profit instead.

The Joglekar, et al (2008) study shows that the continuously increasing price models are superior to the fixed price models. In the basic EOQ models, the profit increase of the continuously increasing price models deliver higher profit/per period over the fixed price models by about 1 to 2 %. However, Lee and Joglekar's (2012) study discovers that when some of the EOQ restrictions are changed, including the current models of quantity discounts, the profit increase could be drastic: over 25 % increase in profit, while total demand per year, or the number of units sold per year actually decreases.

Although we are not going to repeat the sensitivity analysis here performed by Joglekar, et al (2008), and Lee, et al (2012), we did perform some informal tests on the profit maximization models, and found that the results are consistent with the previous reports. In particular, the continuously increasing price models are consistently much more profitable than the fixed price models by selling less per period.

On a separate analysis, the cost-based model arrives at the same optimal policies as the profit maximization models when the annual demand D increases. That is, when D increases from 6781, the optimal policy of ordering 2000 units/per order also shifts to 6000, in agreement with the other models. Obviously, the results are very much dependent on the parameters such as

the price and demand elasticity, *a* and *b* being used, or the cost factors employed in the models. Based on these findings, a retailer should proceed with caution when relying on the traditional cost-based models alone for determining the optimal inventory policies with discount breaks since they might be misled as in our numerical example.

CONCLUSIONS

Since Harris's (1913) pioneer work on lot sizing and inventory, it is customary to set inventory policies according to the cost structure of the company. Since Abad (1997, 2003), and Joglekar, et al (2003, 2008), and Lee, et al (2012), it has been shown that price driven inventory policy (or with supply chains: a coordinated pricing and stocking policy) actually would increase the profit drastically. Although the continuously increasing price model does not offer much profit increase for the traditional EOQ situation, it, however, turns out to be most effective in dealing with other members of the EOQ family. In fact, Lee, et al (2012) and the previous section provide solid evidence that coordinating price and inventory in a supply chain would help both E-tailers and traditional retailers with gradual production, and price discounts.

Admittedly, the continuously increasing price models can be seen as more complex and more difficult to implement, but given today's technology, especially for E-tailers, the continuously increasing price scheme can be easily adopted via web servers. In fact, E-tailers can coordinate their inventory, replenishment, and pricing with their suppliers and customers with today's internet technology. Given the ease of implementation and in the event of quantity discounts with suppliers, E-tailers should seriously consider adopting the profit maximization policies instead of the traditional cost-based replenishment policies for achieving better profits.

This study together with the research by Joglekar, et al (2008) and Lee, et al (2012) have shown that continuous increasing price models with deterministic demands have consistently reigned in significant profits over the fixed pricing models. We would extend our study and report our findings for probabilistic demands of a supply chain and how coordinating price and inventory would impact the overall profit versus the traditional inventory models in the future.

PERTINENT LITERATURE REFERENCES

Abad,,P. L., "Optimal policy for a reseller when the supplier offers a temporary reduction in price," *Decision Sciences*, vol. 28, no. 3, pp. 637–649, 1997.

Abad, P. L., "Optimal price and lot size when the supplier offers a temporary price reduction over an interval," *Computers and Operations Research*, vol. 30, no. 1, pp. 63–74, 2003.

Arcelus, F. J., and G. Srinivasan, "Inventory policies under various optimizing criteria and variable markup rates," *Management Science*, vol. 33, no. 6, pp. 756–762, 1987.

Ardalan, A., "Combined optimal price and optimal inventory replenishment policies when a sale results in increase in demand," *Computers and Operations Research*, vol. 18, no. 8, pp. 721–730, 1991.

Arcelus, F. J., and G. Srinivasan, "Ordering policies under one time only discount and price sensitive demand," *IIE Transactions*, vol. 30, no. 11, pp. 1057–1064, 1998.

Boyaci, T., and G. Gallego, "Coordinating pricing and inventory replenishment policies for one wholesaler and one or more geographically dispersed retailers," *International Journal of Production Economics*, vol. 77, no. 2, pp. 95–111, 2002.

Chase, R. B., F. R. Jacobs and N. J. Aquilano, "Operations management with competitive advantage," McGraw-Hill/Irwin, 13th Edition, 2010.

Christophe, http://www.logisitik.com/learning-center/inventory-management/item/466-economic-order-quantity-eoq-model.html, 20 October, 2009.

Feng, Y., and B. Xiao, "Continuous-time yield management model with multiple prices and reversible price changes," *Management Science*, vol. 46, no. 5, pp. 644–657, 2000.

Hall, R., "Price changes and order quantities: impacts of discount rate and storage costs," *IIE Transactions*, vol. 24, no. 2, pp. 104–110, 1992.

Harris, F., "How many parts to make at once," Factory, The Magazine of Management, vol. 10, no. 2, 1913.

Joglekar, P., P. Lee, and A. M. Farahani, "Continuously increasing price in an inventory cycle: an optimal strategy for e-tailers," Journal of Applied Mathematics and Decision Sciences, Vol. 2008, Article no. 483267, 2008.

Kunreuther, H., and J. F. Richard, "Optimal pricing and inventory decisions for non-seasonal items," *Econometrica*, vol. 39, no. 1, pp. 173–175, 1971.

Lee, P., and P. Joglekar, "Continuously Increasing Price in a Gradual Usage Inventory Cycle: An Optimal Strategy for Coordinating Production with Pricing for a Supply Chain, to appear in Academy of Information and Management Sciences Journal, 2012.

Martin, G. E., "Note on an EOQ model with a temporary sale price," *International Journal of Production Economics*, vol. 37, no. 2-3, pp. 241–243, 1994.

McGill, J. I., and G. J. van Ryzin, "Revenue management: research overview and prospects," *Transportation Science*, vol. 33, no. 2, pp. 233–256, 1999.

Smith, B. C., J. F. Leimkuhler, and R. M. Darrow, "Yield management at American Airlines," *Interfaces*, vol. 27, no. 1, pp. 8–31, 1992.

Talluri, K. T., and J. G. van Ryzin, *The Theory and Practice of Revenue Management*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

Tersine, R. J., and R. L. Price, "Temporary price discounts and EOQ," *Journal of Purchasing and Materials Management*, vol. 17, no. 4, pp. 23–27, 1981.

Weng, Z. K., "Channel coordination and quantity discounts," *Management Science*, vol. 41, no. 9, pp. 1509–1522, 1995.

Weng, Z. K., "Modeling quantity discounts under general price-sensitive demand functions: optimal policies and relationships," *European Journal of Operational Research*, vol. 86, no. 2, pp. 300–314, 1995.

Whitin,T. M., "Inventory control and price theory," *Management Science*, vol. 2, no. 1, pp. 61–68, 1955.

Wilson, R., "Scientific routine for stock control", Harvard Business Review, vol. 13, no. 1, 1934.

A MODEL FOR MANAGING RENTAL FLEETS IN THE NEW COMPETITIVE LANDSCAPE: MAINTENANCE, PRODUCTIVITY, CORPORATE BRANDING AND LEGAL IMPLICATIONS

Anthony Narsing, Macon State College
Terry Sanders, Macon State College
Mimi Ford, Macon State College

ABSTRACT

Rental companies across the United States are paying close attention to profitability. The current economic landscape has changed, and unfortunately it forces those companies to move away from the outdated model of utilizing rental cars between four and six months, after which they are disposed of on the used car market for high profits. An improved model is needed to address the dynamic changes in the new competitive landscape, which may significantly impact profitability and other business operational factors. The authors of this paper propose a model, examining maintenance costs, productivity, branding and legal implications for more effective management of rental companies.

INTRODUCTION

Those who lease or rent vehicles for personal or business use often do so with the assumption that maintenance, defects or recall issues have all been addressed. From a consumer advocacy point of view rental car companies must practice due diligence, which simply means that they are responsible for a car's overall maintenance while it is in their fleet. Sadly enough, this may not be the case. A jury awarded $15 million in the wrongful death of two sisters, ages 20 and 24, who had rented a vehicle from Enterprise Rental Car Company. The company failed to repair a leaking power steering fluid, which was cited under a recall. It was determined that leaking power steering fluid was ultimately the cause for the fatal accident. This type of occurrence has damaging effects on corporate branding, which significantly impacts revenue, customer satisfaction, reputation, trust, corporate image and social responsibility.

A designed structured approach is required to closely monitor and take corrective action to resolve problems like recalls. Units in the fleet can be fixed systematically without the risk of grounding the operation. A fleet manager can assess the severity of its recalls by working closely with the manufacturer to resolve problems in a timely fashion. In addition to recall issues, it is imperative for fleet managers to proactively respond to basic wear and tear issues such as those related to tires, brakes, hoses, oil changes and belts, as well as any repairs outside of those covered by the manufacturer's warrantee.

As the number of new car acquisitions continues to decline, fleet operators are faced with tremendous challenges in managing a diverse mix of new and old units. The Big Three (GM, Ford and Chrysler) have significantly reduced both program cars and deep discounts (Sawyers, 2011). This essentially means that older units can potentially accumulate as much as 50,000 miles (Sawyers, 2010), which leads to significant increases in maintenance cost. With a reduction of program cars, fleet managers must absorb added expenses to prepare a rental unit for its disposal on the used car auction market place.

The authors of this paper propose the Fleet Maintenance Planning (FMP) model, which will yield significant benefits to rental car fleet operators in both small and large markets. The output from this model will assist operators to better manage and improve maintenance costs, increase productivity, enhance corporate branding and limit business liability.

LITERATURE REVIEW

Several authors discuss fleet maintenance as a secondary topic. No one has directly identified a model to address maintenance needs. Erns, Horn, Kilby and Krishnamoorthy (2010) discuss rental fleet scheduling problems which arises when a rental unit is rented from one location and returned to another. They link this problem to revenue management issues but do not discuss fleet maintenance directly. Lieberman and warren (2007) are concerned more with pricing and revenue management capabilities rather than exploiting customer needs for safer and well maintained rental units. These two criteria alone would significantly influence revenue. If a customer's perception is that rental units are deemed unsafe, then certainly pricing policy becomes a moot point as no one wants to rent an unsafe car. Cho (2005) addresses determinants for used rental car resale values such as branding; however, he does not mention that overall maintenance, which is even a bigger factor. Antich (2011) specifically focuses on fleet car maintenance and identifies key factors such as oil changes, tires, increases to repair shop labor rates, which are likely to adversely impact maintenance and productivity.

Unfortunately after an exhaustive search, the literature search did not produce any articles, which specifically identified rental operators and any specific fleet maintenance programs. The authors targeted trade many popular trade journals such as "Automotive Fleet," which openly discusses the need for maintenance programs as a fleet begins to age. However, they do not identify maintenance programs with rental car operators. Several on-line media outlets such as Auto Rental News repeatedly (Brown, 2012) discuss fluctuating earnings, RPD (revenue per day) pricing schemes, while keeping expenses low (i.e. fleet costs). One could argue that keeping expenses low could have widespread implications on other cost issues such as those related to timely repairs and or preventative maintenance.

MANAGING FLEET OPERATIONS

Depressed economic conditions, contracting fleet sizes, fewer program cars, elimination of deep discounts from the Big Three (Ford, Chrysler and GM) manufacturers, aging units, and reduced fleet allocation by manufacturers (i.e. aimed at protecting residuals) are an indication

that changes in the rental car industry are here to stay (Thompson, 2009 and Brown, 2011). Fleet operators in both small and large markets are adversely affected by these changes. Unfortunately, those operating in small markets are finding it increasingly difficult to manage their operations, and yet maintain an acceptable level of profitability. Data collected from a local small market fleet operator (who wishes to remain anonymous) between 2009-2011 reveals much of the same problems listed above. Figure 1 clearly shows an overall pattern of declining monthly revenues, while Figure 2 confirms the positive correlation between monthly rentals and revenues.

Figure 1. Monthly Revenues 2009-2011

Source: local rental car fleet operator

Figure 2. Monthly Rentals 2009-2011

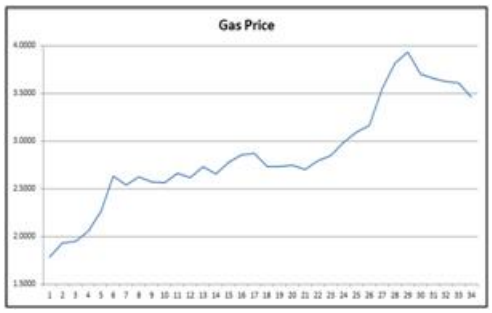
Source: local rental car fleet operator

Both revenues and the number of rentals declined steadily in 2009 as the recession had a firm grip on the economy; however, there were signs of recovery during the first half of 2010 as the number of rentals increased giving a sharp boost to revenue. Unfortunately, rentals and revenues declined sharply in the latter half of 2010 and showed only marginal gains in 2011. An overall pattern of declining revenue together with the number of rentals can be attributed to several factors: higher fuel prices resulting in fewer planned trips, businesses cutting back on travel, and increased carpooling.

Figure 3 illustrates sharp increases in the price of regular fuel during the same period from 2009-2011, which essentially confirms the patterns in revenue and the number of rentals shown in Figures 1 and 2, respectively. The modest gains in revenue displayed during the first half of 2010 is explained largely by the steady price of regular fuel, which hovered between $2.63 to $2.75 per gallon as shown in Figure 3. There are many factors which influence the price of fuel, some of which include: political unrest in the Middle East (i.e. Egypt, Iran, Syria, Iraq influences world supply and demand), world priced of crude oil, a weakened U.S. dollar, China's surging demand for crude and oil futures trading by commodity brokers. Though rental car operators are not directly involved in influencing changes to fuel prices, unfortunately they are adversely affected as are many other businesses that rely on its use such as the airline industry.

Figure 4 simply bolsters and confirms the adverse impact felt by rental car operators during the same period. An upward CPI trend is indicative of a weakened economy as consumers are faced with challenges of managing increases to household expenditures. Certainly such a trend cannot be welcoming news to rental car operators, or any other businesses for that matter.

Figure 3. Price of Fuel from 2009-2011



Source: http://www.eia.gov/

Figure 4. CPI from 2009-2011

Source: http://www.bls.gov/

MAINTENANCE AND PRODUCTIVITY

Current Maintenance Practice

In this section an overview of current practice and discussion of key attributes associated with fleet maintenance are presented. An important consideration is the current practice of managing fleet maintenance particularly with a mix involving both current and aging units. The authors contacted several local managers at rental car agencies to inquire about their fleet maintenance practices. Unfortunately, many were very apprehensive and reluctant to share information. For the purpose of our research, the manager of a large chain decided to share insights about current maintenance practices on the condition that company information remained anonymous. When asked about a specific fleet maintenance program, the manager stated that there was no formal program in place nor was any suggested by corporate office. Maintenance was handled on an as-needed basis and hence data collection is reported for cost associated with any untimely repairs. Emphasis was placed on basic preventative maintenance such as oil changes. That is, whenever the car reached a certain mileage, its internal computer would alert the operator that it was time to change the engine's oil. The engines found in most new cars allow for oil changes every 7,000 to 10,000 miles resulting from a switch to higher motor oil (Antich, 2011) standards. Unfortunately, new oil standard grade denoted by GF-5 already in use in new Chrysler and Ford models can expect an increase in oil service cost about 15% or an added $2 to $3 per quart. On the subject of tires, which can add about $85.45 per vehicle (Antich, 2009); the manager indicated that tires are only serviced if there was an obvious problem such as a puncture, which would normally be the responsibility of the driver and not covered under the rental agreement. Tire rotation is not even a planned service. At the next oil change interval depending on the service provider, a multi-point inspection may be conducted. I pointed out that a typical car loses about 2 psi in each tire per month. Hence waiting for the next oil change, a car could be potentially losing anywhere from 2-8 psi per tire depending on service date. Drop in tire pressure can lead to tremendous under-steering and ultimately loss of control, which is further compounded if the vehicle is carrying a significant load.

In summary, it was revealed that no formal preventative maintenance existed at any of their rental locations. It was common practice to perform a quick visual inspection, while preparing the car for the next rental customer. The inspection is not mechanical but rather cosmetic such as detecting unusual scratches and body abrasions. Their maintenance practice corroborates the findings in the literature search, which suggested that maintaining profits is the goal at the expense of keeping costs low. When pressed about recalls, the manager indicated that they are certainly important; however, there was no published protocol at the local level to handle such issues. Repairs of any kind are only triggered by customer complaints. Otherwise the rental units are returned to service as quickly as possible. The manager suggested that their methods of dealing with maintenance are not unique to them, but rather it is similarly practiced by their competitors for whom he had previously worked.

The shortcomings of using an adhoc process for addressing maintenance issues present many problems, particularly as rental units age. Failure to inspect for basic wear in components such as belts, brakes, tires, hoses or checking fluid levels can all lead to higher repair costs, when operators do not subscribe to formal preventative maintenance program. Failure to address recall issues in a timely manner can lead to unsafe operation thereby placing a customer at risk, which was the case with the Enterprise who failed to fix the leak in a power steering hose as described in a recall bulletin by Chrysler. The result was a fiery crash and the loss of two innocent lives.

As an example, most cars today use a single belt called a serpentine belt to drive major engine components such as the alternator, water pump, power steering and air conditioning. These belts are subjected to extreme working conditions and can fail without notice due to hairline cracks, which can be overlooked by an untrained individual. A skilled technician will often remove the belt for careful inspection. A failed serpentine renders a vehicle in operable leaving a rental customer stranded until the vehicle is towed to a repair facility.

Again we can site brake inspection as another example, which can lead to increased maintenance cost. Braking systems are more advanced and designed to last longer. However, rental units are never driven by the same individual hence the rate at which wear takes place will vary thereby affecting stopping distance. For instance, this can create problems for travelers who have to make emergency stops with a loaded vehicle. Trained technicians would remove each tire and measure the thickness of the rotor and drum followed by an inspection of both shoes and pads. These measurements would be recorded and monitored by the operators for each rental unit. The short list of examples presented offer insights of what can go wrong when a preventative maintenance program is not present. Not only are repair costs likely to increase, but also impending liability risks of operating a poorly maintained vehicle. The FMP model proposed in this paper forces fleet operator to maintain a preventative maintenance program, document all occurrences and to act on needed repairs before component failure, thereby limiting both operating cost and reducing personal risk. In this way, a customer will be much safer when driving a rental vehicle.

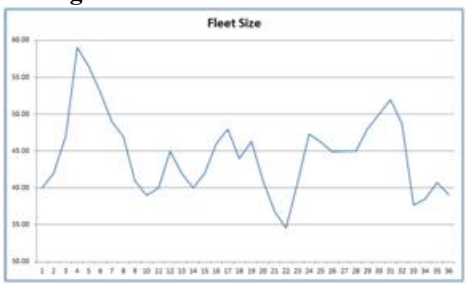
Attributes of Fleet Maintenance

An effective fleet maintenance plan is critical to the survival of operators because they are forced to adopt a new model. This new model involves smaller fleet sizes with rental units

held for longer periods perhaps accumulating as much as 50,000 miles before disposal at the used car auction. Figure 5 shows a steady decline in fleet size for the local rental fleet operator from 2008-2009. Ironically, in 2008 GM asked the federal government for bailout assistance to avoid filing chapter 11 bankruptcy. A strong resale market in 2010 allowed the small operator to replace his fleet and hence reduce maintenance costs. Most rental car companies like Enterprise Rent-a-Car and Budget strive to keep their cars from one to two years, where mileage can accumulate between 18,000 to 30,000 miles before moving those units onto their used car lot. Of course, this is their goal which may or may not be applicable to all of their rental units.

The Big Three have since cut their production, which means fewer fleet allocations to rental car operators with even smaller discounts. Fleet operators found it to be very difficult to dispose of their used rental units because the used car market was already saturated, which translates into much smaller profits. In some instances disposal may result in significant losses at which point the decision to sell becomes moot. As a result, fleet operators are challenged to maintain their existing rental units with an understanding that the composition of their fleet will include both new and aging (greater than six months) units. Maintenance by definition is moving beyond simple oil changes. Instead, operators must address and anticipate critical repairs particularly for those units that begin to accumulate significant mileage.

Figure 5: Fleet Size from 2009-2011



Source: Local fleet rental operator 1

GE Capital Solutions Fleet Services is a fleet management company located in Eden Prairie, Minnesota. In the last 17 years, they have been regularly conducting surveys of fleet operators to track and compile maintenance data of their passenger cars. Their results reveal significant findings, which are very relevant and applicable to all fleet car rental operators as it is invaluable in helping them to better understand and plan their maintenance program. In their most recent study of 13,318 passenger cars from January 1 to December 31, 2011, they found a slight decrease in overall fleet maintenance costs as compared to the calendar year 2010. Fleet operators were able to capitalize on a strong 2010 resale market, which allowed them to keep their overall maintenance costs "flat". The concept of keeping costs flat is achieved as fleet operators replace older units in a timely manner to avoid preventative and costly repairs commonly associated with aging units. Table 1 displays maintenance cost associated with fleet maintenance from 2009 to 2011.

Table 1: Fleet Maintenance Cost from 2009-2011

	Total Maintenance		**Oil Change Services Cost**		**Replacement Tire Cost**		**Average Repair Cost per Unit**	
Calendar Year	Cost per Unit per Month	Cost per Mile	Cost per Unit	Cost per Mile	Cost per Unit per Month	Cost per Mile	Cost per Unit per Month	Cost per Mile
2009	$56.11	$0.0292	$8.72	$0.0045	$18.75	$0.0098	$26.92	$0.0140
2010	$55.89	$0.0292	$9.01	$0.0047	$16.05	$0.0084	$28.51	$0.0150
2011	$53.49	$0.0276	$9.01	$0.0046	$15.78	$0.0081	$26.01	$0.0134

Source: GE Capital Fleet Services (2011)

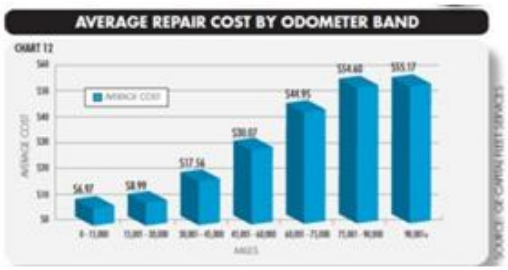
Total maintenance costs include tires, maintenance repairs (which include unscheduled services such as brakes, suspension, engine, transmission, electrical, and other), along with other preventative maintenance for passenger cars. As shown in Table 1, a strong resale market, increased interval between oil changes according to OEM specification, and an overall improvement in the quality of cars helped to keep maintenance costs flat. Price of oil change services increased by six percent from last year (Strom, 2011), which is reflected in part by increases in crude oil. Manufacturers are switching to new motor oil for "improved efficiency and engine protection" as noted by Strom. GM's new cars and trucks are designed to use Dexos, while Ford is moving toward the new GF-5 oil standard for its vehicles. Some fleet operators are beginning to rely on the unit's on-board oil service indicator for determining when to change oil. Of course, this strategy is acceptable as long as driving conditions and environment are kept constant. Unfortunately, this is an unrealistic expectation as fleet managers must keep a detailed log from each unit to provide effective maintenance.

Fleet managers were able to negotiate fleet pricing for tire replacement as reflected by the lower rate shown in Table 1. However, tire pricing levels are expected to rise this year as increases are expected in raw materials, crude oil, and an overall high global demand for rubber. Figure 6 indicates an upward trend in average repair costs as rental units remain longer in service. Average repair costs appear to almost triple after a rental unit's first year in service increasing from $8.76 in year 1 to $24.18 in year 2. Figure 7 shows a similar upward trend in repair cost as mileage increases. Most new vehicle warrantee is three years or 36 months, whichever comes first. Average repair costs unrelated to warrantee increased from $6.97 to $17.56, when a rental unit reaches up to 45,000 miles. Beyond 45,000 miles, rental units average repair costs increase significantly as major repair components begin to fail.

Figure 6. Average Repair Cost by Months in Service

Source: GE Capital Fleet Services

Figure 7. Average Repair Cost per Odometer Band



Source: GE Capital Fleet Services

Productivity is expected to decline as maintenance repair costs continue to climb with aging rental units. To ensure the highest level of productivity, fleet managers must adopt a progressive preventative maintenance program. It is best that they have a dedicated licensed and insured reputable repair shop with an agreed pricing contract for basic operations such as oil changes and a routine 21 point inspection. Any deviation from a dedicated maintenance program will result in unsafe vehicles and expose the company to financial losses, litigation and may adversely impact branding.

CORPORATE BRANDING

The issue of branding for car rental companies must include image. Branding is considered a name assigned to individual products, services or a group of complementary products. Some rental car companies use their corporate image to deliver many complimentary products to the marketplace.

Corporate image reflects the feelings consumers and businesses have about the overall organization and its individual brands (Low and Baak, 2012). The image may be divided into two distinct categories which are the consumer perspective and the corporate perspective.

Rental car companies spend millions of dollars to create, change or rejuvenate their existing image. For example Avis selected the marketing and design firm of Siegel and Gale to create a new identity in readiness for its flotation on the US market and to help it compete in the increasingly competitive car rental market. Its task will be to create a new corporate image for the car rental giant in readiness for its initial public offering which was filed for this month (Design Week, 2012).

In the case of Avis the consumer perspective will be of utmost importance due to several factors including confidence in purchasing decisions, product assurance and social acceptance of the product. From the corporate perspective a good image allows rental car companies to charge more for goods and services, attract better employees and create greater channel power not to mention the more favorable image on Wall Street.

Avis employs the family brand for many related products used to enhance their rental car services. For example, Avis markets Avis where2® GPS Navigation, Avis Blast - Portable XM Radio and Avis Access (Disability Products)[1] The addition of disability products elevates Avis by offering the following services:

- Scooter Rentals: Avis provides mobility scooter rentals
- **Transfer Board**: Enables easier transfer from a wheelchair into and out of the vehicle.
- **Panoramic Mirror**: Provides a much larger field of view for any driver.
- **Swivel Seat**: Allows the driver or passenger to turn his or her body in the car seat with limited effort.
- **Hand Controls**: Enables drivers to accelerate or brake using a hand-controlled device.
- **Spinner Knob**: Enables the driver to turn the steering wheel with a full turning radius using only one hand.
- **Easy Access Bus**: Offers an electrically-operated ramp, two ADA compliant wheelchair positions, wider doors, spacious aisles and low luggage racks.

Enterprise Rental Car Company brands their rental service as local establishments that deliver rental cars directly to their customers either at home or at work. The consumer perspective of Enterprise is greatly enhanced through this image primarily as social acceptance.

Enterprise Rent-A-Car is the latest organization to enroll as an NABC(National Auto Body Council) diamond level member, the highest level of NABC corporate membership. Like

several other corporate members, Enterprise's support of the NABC will include participation in Recycled Rides, the NABC-sponsored nationwide awareness project whereby NABC members repair and donate recycled vehicles to families and service organizations in need in their communities. This move by Enterprise should motivate consumers by assurance of product reliability.

Hertz Rental Car utilizes their corporate logo and image to further market their many brands. Hertz corporate vision states the following: "Hertz will be the corporate leader in mobility and equipment solutions"[2] The term mobility solutions may be a substitute for rental cars and if we look carefully at Hertz complimentary and flanker brands we will see why.

Hertz utilizes the corporate image (logo) to market several products and services as extensions to the flagship brand of rental cars. For example:

- Hertz Rent a Car: Providing Quality Car Service for over 90 years.
- Hertz Local Edition®: In Your Neighborhood, providing local and insurance replacement rentals.
- Hertz Equipment Rental: A leading provider of heavy construction equipment and tool rental and sales
- Hertz Car Sales®: Selling one-year old vehicles from the Hertz fleet at low prices.
- Hertz Truck & Van Rental: Whether you are moving across town or making large deliveries, we are here to help.
- Hertz Entertainment Services offers everything you need for your next big film or television production or live event.

Hertz employs family brands(different logo) for several other products such as, Hertz On Demand™ a pay-as-you go car membership club, Donlen Corporation which is a global provider of innovative fleet leasing and management solutions for corporate fleets across the U.S. and Advantage Rental Car which offers a value rental rate in key vacation destinations.

In summary, we are seeing the core business of the rental car industries diversifying through complimentary and family brands of products in order to increase revenue and grow the business.

LEGAL ISSUES

Previous Law and Background

Under prior law, some states held car rental companies "vicariously liable." Under this doctrine, an accident victim injured due to the negligent operation of a rental car could hold the car rental company responsible for injuries regardless of whether or not the car rental company was negligent, or in any way at fault. Liability was imputed to the car rental company through mere ownership of the vehicle. In 2005, Congress passed the Transportation Equity Act, a provision of which eliminated vicarious liability, as it applies to rental companies. Referred to as the Graves Amendment, the law provides that a rental car company will not be held liable for the

tortuous actions of the driver of a car that the company owns, rents, or leases to an individual so long as the rental car company is not negligent or guilty of criminal wrongdoing. 49 U.S.C. Sec. 30106(a)(2005)

Current Law

Although the Graves Amendment has served to remove car rental companies from vicarious liability, they can still be liable under other theories of tort law in negligence such as negligent maintenance and negligent entrustment (Coverage Counsel, 2009).

Negligent Maintenance and Strict Product Liability

As car rental companies are now keeping the cars in their fleets for a longer period of time, negligence in failing to maintain the vehicles in a safe condition becomes of increasing concern. In addition, the car rental companies can be a party-defendant to a strict product liability lawsuit. Although the manufacturers will likely ultimately bear the liability in the lawsuit, the rental car company will be called upon to defend itself to the claim and incur steep legal costs in such a lawsuit (Scott, C. and Kissane, P.A., 2011)

In some instances, the vehicle may be subject to a safety recall mandated by the National Highway Traffic Safety Administration. Under current law, car dealerships cannot sell recalled vehicles to consumers; however, no law bans car rental companies from renting such vehicles to consumers. Although car rental companies are not subject to supervision of the NHTSA, this does not absolve them of liability. Under tort law, they have a duty to fix these vehicles, as evidenced by a California case. In 2011, a jury awarded $15 million in damages to the family of Raechel and Jacqueline Houck. The twin sisters were killed in a fiery crash involving a recalled PT Cruiser, rented to them from Enterprise Rent-a-Car. In the lawsuit, it was brought out that the vehicle had a safety recall issue with a power steering hose that could leak and cause a fire. Enterprise had not made the necessary repairs on the vehicle. Although Enterprise was not subject to the NHTSA recall notice, the company was held to be negligent by not repairing the vehicle. Of note, a bill has been introduced to the Senate (named for the Houck twins) that will prohibit the rental of defective cars and trucks, and bring the rental car companies under the control of the NHTSA.

Negligent Entrustment

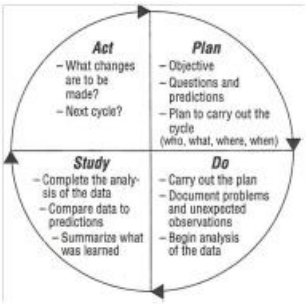
Under the theory of negligent entrustment, a rental car company can be liable in tort for negligent "entrusting" the operation of a vehicle to another. If a car rental company should know that a potential rental operator of a vehicle will pose an unreasonable risk to others and does entrust that vehicle to that person, the rental company is liable if this renter injures another

through negligent operation of the vehicle. However, the threat and magnitude of such lawsuits is not affected by the aging fleets of the car rental companies.

MODEL DEVELOPMENT

The Fleet Maintenance Planning (FMP) model is built on the principal of continuous improvement. It uses the Plan-Do-Act-Check (PDCA) management method whose application was highlighted extensively by W. Edward Deming (Kanji, 1996; Smart, Henderson, et.al., 1993). Deming used this and other quality control tools and highlighted their application in assisting companies to better manage and improve their business processes. The concept of PDCA, which has four stages, is built on a scientific basis with its roots derived from the work of Francis Bacon, a pioneer, advocate and practitioner of using scientific methods. This methodical approach permits management to create a structure around which fleet operators can effectively manage their operations and use a feedback mechanism to check and make corrections when bottlenecks or constrictions are identified within their processes (Marquis, 2009).

Figure 8: PDCA Model



(Source: Moen and Norman, 2006)

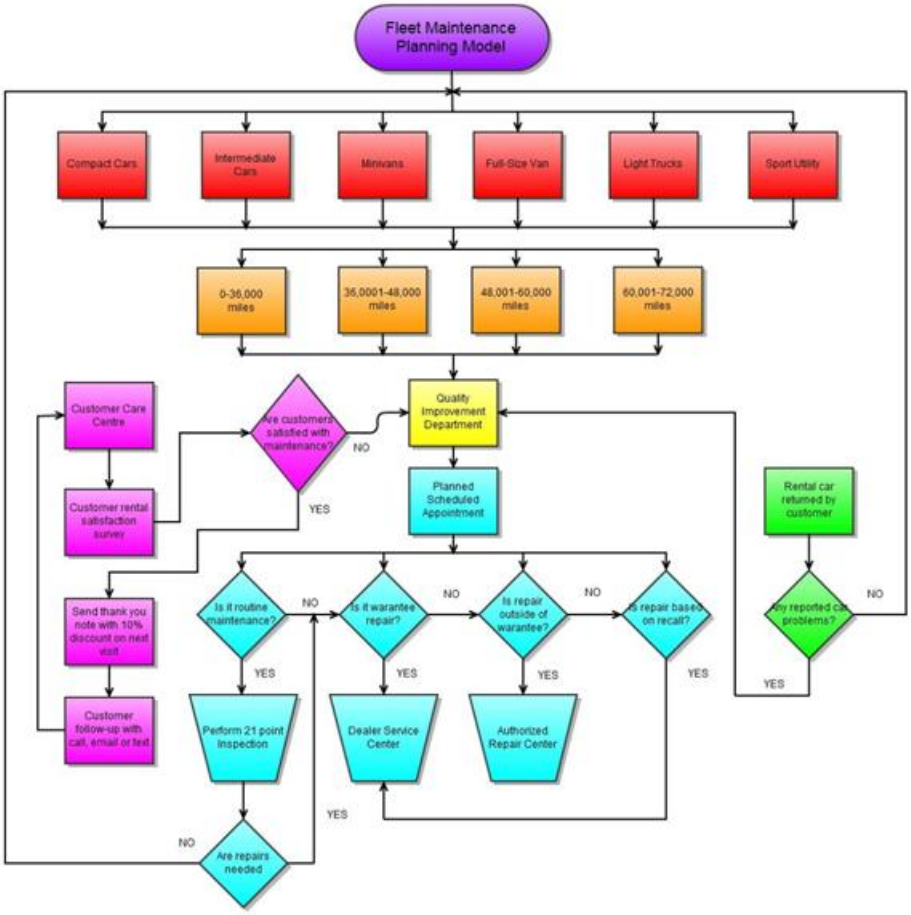
The PDCA method showed in figure 8 forms the basis in the development of the FMP. The output from the model presented in the this paper will greatly assist fleet operators in collecting data, planning and controlling operating cost, minimize downtime, improve productivity, act on recalls, and listen to customer maintenance concerns. The four typical stages to the PDCA method are given below:

PLAN: companies set goals and identify methods needed to reach those goals, which can be accomplished by designing a new or revising an existent business process

- **DO**: This is a prototype stage where companies can implement the plan and establish metrics to measure its performance
- **CHECK**: Review results from prototype test in DO stage and compare against goals established in PLAN stage. Provide feedback to management.
- **ACT**: Implement refined and improved solution by deciding on changes needed to meet company goals.

The FMM as shown in Figure 9 illustrates all four stages of the PDCA method (Stevenson, 2009). As noted, the PLAN stage is perhaps the most important step because it specifically outlines company goals and objectives in its long-term strategic plan.

Figure 9. Fleet Maintenance Planning Model



In the "**PLAN**" stage, a fleet is presented using six categories as noted by the "red" symbols in Figure 9: compact, intermediate, mini vans, full size vans, light trucks and sport utility vehicles, which is consistent among operators (i.e. Hertz, Avis, Economy etc…). Each class of vehicle is further classified into four mileage groupings shown in by the "orange symbols": 0-36,000 miles, 36,001-48,000, 48,001-60,000 and 60,001-72,000. This mileage grouping is based on the criteria presented in Automotive Fleet, which was used to capture maintenance data by tracking repairs over a 12-month period across several categories as shown in Table 2.

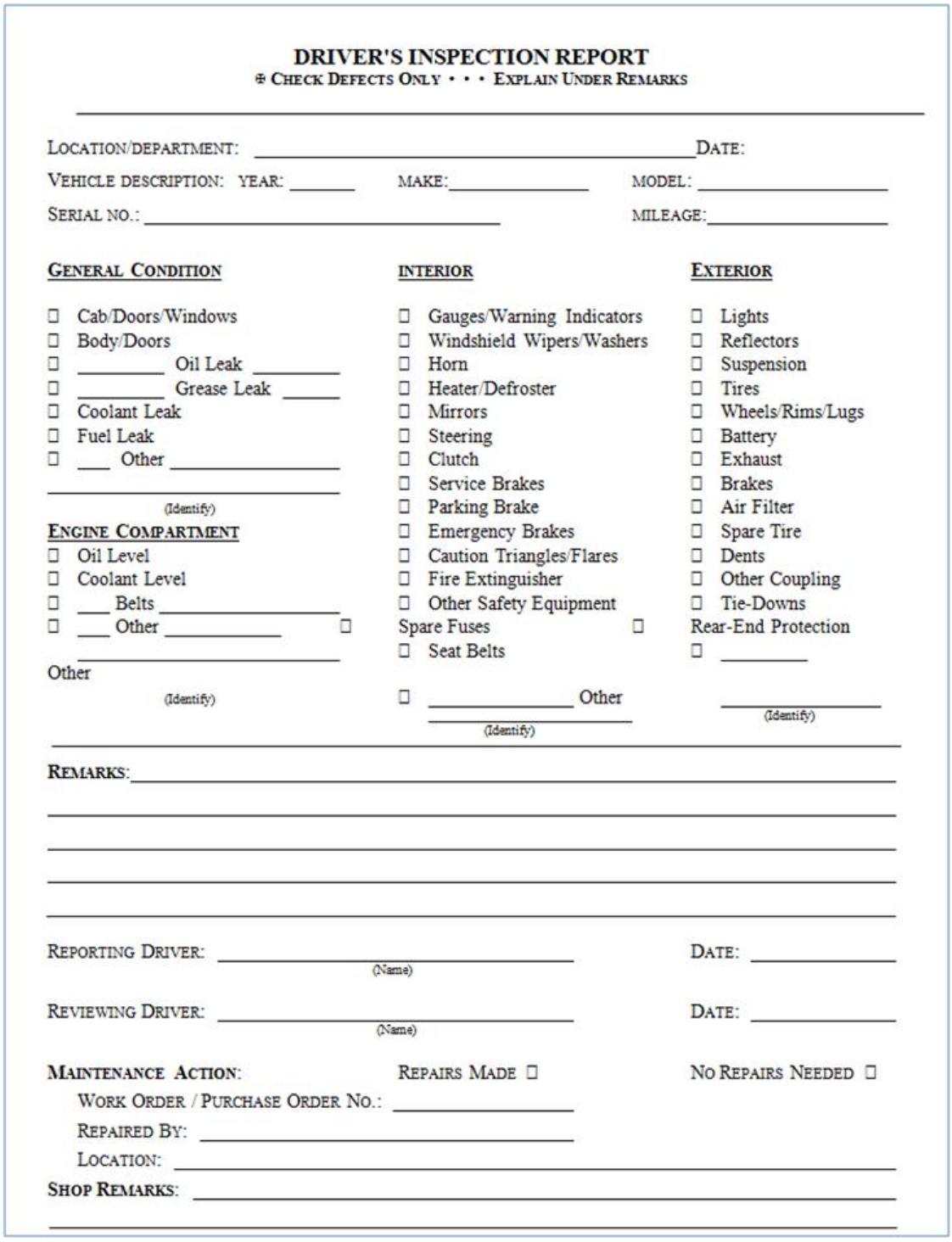
Table 2. Repair Categories Associated with Annual Fleet Maintenance			
Item No.	**Category**	**Item No.**	**Category**
1	Tires	8	HVAC
2	Brakes	9	Suspension
3	Engine	10	Exhaust
4	Steering	11	Starting System
5	Fuel System	12	Ignition
6	Cooling	13	Electrical
7	Charging System	14	Transmission

A "Quality Improvement (QI)" department oversees all aspects associated with organizing and planning maintenance needs within a fleet (Stevenson, 2009; Jacobs and Chase, 2011). Fleet managers would work closely with their QI coordinator to ensure that both throughput and workload are not adversely impacting fleet utilization levels, which is particularly important for operators during high-peak periods such as Thanksgiving. In others words, the QI coordinator along with the fleet manager would systematically devise a plan where maintenance for high volume rentals are prioritized at the highest level, while maintenance for less frequently rented units are worked on later.

The blue symbols highlight the "**DO**" stage where operators pilot their process before full scale implementation. In this way, mistakes can be identified and potentially resolved without significant financial and operational disruption. The QI coordinator works at executing the maintenance schedule once it is reviewed and approved by the fleet manager. Appointments are scheduled based on the type of maintenance needs, which can be simply routine as in a basic or detailed multi-point inspection (i.e. similar to the ones indicated below in Figures 10 and 11), warrantee repair, non-warrantee repair or recall related.

At the "**CHECK**" stage data is gathered twofold: 1) customer returns rental unit indicated by the "green" symbols and 2) post customer comments from "customer" satisfaction survey shown by the "pink" symbols. The importance of this stage is to compare the goals and objectives, which were established in the "**PLAN**" stage using the gathered data. Any deviation or departure from stated goals must be addressed and corrected before full implementation is pursued in the "**ACT**" stage. Attendants checking in each rental unit can administer a quick survey, while speaking with the customer and logging their responses into their PDA. The "Customer Care Center" shown by the "pink" symbol would send a customer satisfaction survey out to those who might have used the drop box when returning their rental unit. All customers would receive a follow-up visit via telephone call, text, or by e-mail. As a courtesy, fleet managers could offer a 10% discount on their next rental as a token of customer appreciation.

Figure 10: Basic Multi Point Inspection



DRIVER'S INSPECTION REPORT

✠ CHECK DEFECTS ONLY • • • EXPLAIN UNDER REMARKS

LOCATION/DEPARTMENT: _____ DATE:

VEHICLE DESCRIPTION: YEAR: _____ MAKE: _____ MODEL: _____

SERIAL NO.: _____ MILEAGE: _____

GENERAL CONDITION

- ☐ Cab/Doors/Windows
- ☐ Body/Doors
- ☐ _____ Oil Leak _____
- ☐ _____ Grease Leak _____
- ☐ Coolant Leak
- ☐ Fuel Leak
- ☐ ___ Other _____

_____
(Identify)

ENGINE COMPARTMENT

- ☐ Oil Level
- ☐ Coolant Level
- ☐ ___ Belts _____
- ☐ ___ Other _____

_____
(Identify)

INTERIOR

- ☐ Gauges/Warning Indicators
- ☐ Windshield Wipers/Washers
- ☐ Horn
- ☐ Heater/Defroster
- ☐ Mirrors
- ☐ Steering
- ☐ Clutch
- ☐ Service Brakes
- ☐ Parking Brake
- ☐ Emergency Brakes
- ☐ Caution Triangles/Flares
- ☐ Fire Extinguisher
- ☐ Other Safety Equipment

Spare Fuses ☐

- ☐ Seat Belts
- ☐ _____ Other _____
(Identify)

EXTERIOR

- ☐ Lights
- ☐ Reflectors
- ☐ Suspension
- ☐ Tires
- ☐ Wheels/Rims/Lugs
- ☐ Battery
- ☐ Exhaust
- ☐ Brakes
- ☐ Air Filter
- ☐ Spare Tire
- ☐ Dents
- ☐ Other Coupling
- ☐ Tie-Downs

Rear-End Protection

- ☐ _____

_____
(Identify)

REMARKS: _____

REPORTING DRIVER: _____ DATE: _____
(Name)

REVIEWING DRIVER: _____ DATE: _____
(Name)

MAINTENANCE ACTION: REPAIRS MADE ☐ NO REPAIRS NEEDED ☐

WORK ORDER / PURCHASE ORDER NO.: _____

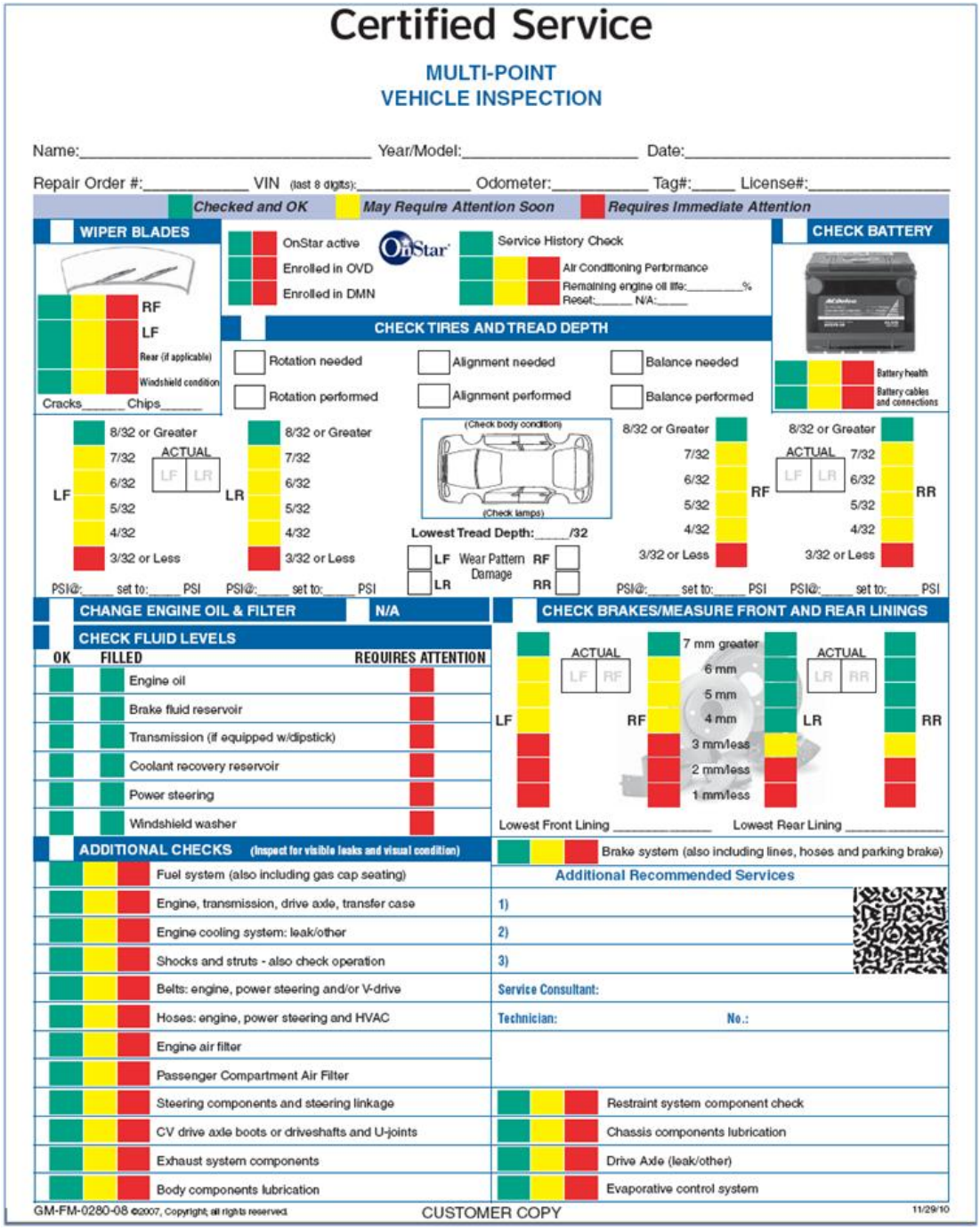
REPAIRED BY: _____

LOCATION: _____

SHOP REMARKS: _____

Source: www.harrp.com/.../Vehiclemaintenanceformanddriversinspectionrep.

Figure 11: Detailed GM Multi Point Inspection



Source: http://www.mycertifiedservice.com

The final stage of the PDCA method is to "**ACT**". At this stage management are able to perform a final review, verify and validate inner workings of the FMP model by closely examining the collected data. Fleet operators can make any change(s) deemed necessary before moving towards full implementation of the FMP model. All participants, both internal and external to the process, can share any input before standardization is adopted. As fleet managers move forward towards adoption, they must understand that the PDCA method seeks continuous improvement. Hence, fleet operators can rest assured that the FMP model will continue to evolve as more opportunities for improvement are presented along the process.

MODEL IMPLICATIONS AND FUTURE WORK

The FMP model developed in this paper successfully incorporates the PDCA method in a rental car fleet maintenance program. The incorporated PDCA method promotes continuous improvement and makes use of a dynamic feedback process to monitor performance against stated company goals and objectives.

The FMP model specifically addresses the cores issues discussed in this paper, fleet maintenance, productivity, branding and legal implications. Fleet maintenance is central to the development of the FMP model. The QI department plays a significant role in this process as it's responsible for overseeing the process of continuous improvement by keeping channels of communication open between the fleet managers, fleet owners and customers. Its responsibility is to ensure that every unit is properly maintained particularly for those units which are aging relative to others in the fleet. QI will ensure that every vehicle is maintained for safe and reliable operations by working closely with its authorized repair centers. It is important for fleet operators to forge long-term relationships with their repair centers, much like Wal-Mart does with its suppliers. It builds trust, mutual respect and ensures quality of service.

There are two feedback loops in this model as both are designed to collect data, which can then be used to make changes deemed necessary to ensure 100% customer satisfaction. Customers can report any maintenance problems as they drop their rental off. The advisor checking in their rental will log their concern, which will be addressed by the QI department. Advisors can offer a 10% discretionary discount as a "goodwill" gesture.

Sometimes customers simply drop off their rentals without ever seeing an advisor. In this case, a survey would be sent to those individuals. If they are happy, a 10% discount is offered on their next visit. Otherwise, the QI department is given the task to deal with any negative maintenance experiences. In both cases, there will be a follow-up by the either the customer care center or the QI department with every event logged for quality control. The advantage to this strategy is to ensure that customers do not seek alternate on-line blogs to vent their concerns as this serves only to adversely affect the company branding. Too often companies ignore customer complaints until it begins to affect business. In a business climate affected by an economic

downturn, it is prudent for rental car fleet operators to act quickly on customer maintenance related complaints.

Recalls are no longer ignored but rather they are systematically addressed on a prioritized basis by the QI department. It is not necessary for fleet operators to ground their fleet at the sight of a recall. Instead, the QI department would work closely with the dealer service center to identify the severity of the recall problem and schedule it for repair. In this way, fleet operators will avert any legal issues and hopefully not rent a car if the recall is deemed detrimental to its safe operation.

This research has established a solid framework and foundation, which enhances the way rental car companies approach fleet maintenance. The structure of the FMP model can be adapted in whole or parts to any fleet car maintenance program. The next step is find a company, who is willing to use this model as a pilot study from which data can be collected to monitor its effectiveness. This could form the basis of a follow-up research paper.

CONCLUSION

Rental car fleet operators have adopted a new model, one that includes fewer fleet allocation, no more program cars, and declining discounts from manufacturers. Fleets must manage a mix of both new and aging rental units. The FMP model, which is driven by the PDCA method establishes maintenance accountability and uses feedback for achieving continuous improvement. It allows rental car fleet operators to address and manage issues such as productivity, branding and legality as they are all linked to fleet maintenance. In an environment where change is constant, fleet owners must embrace this new competitive landscape and work relentlessly to exceed maintenance needs in their fleet. In the end, they will have a safer fleet and much happier customers.

ENDNOTES

1 http://www.avis.com/car-rental/content/display.ac?contentId=avis-access-US_en-001311

2 https://images.hertz.com/pdfs/VMVWeb.pdf

REFERENCES

Antich, M., “Maintenance Costs”, Automotive Fleet, March 2009

Antich, M., “Maintenance Costs”, Automotive Fleet, March 2011

Brown, C., “Car Rental: A tough Year for Pricing”, Auto Rental News (February, 24, 2012), www.autorentalnews.com

Brown, C., “6 Predictions for the Car Rental Industry in 2012”, Auto Rental News (December, 19, 2011), www.autorentalnews.com

Cho, J.S., “Determinants of Used Rental Car Prices”, Journal of Economics Research 10 (2005) 277-304.

Coverage Counsel, “Graves Amendment Found Constitutional—again”, Friday April 3, 2009.

Coverage Counsel, "Second Departmant Reaffirms its Finding that Graves Amendment is Constitutional—Personal Injury Action against trustee of Vehicle Lessor is Dismssed", Friday April 30, 2010.

Design Week .com/UK April 10, 2012

Jacobs, F. and Chase, R., "Operations an Supply Management", McGraw-Hill Publisher, 2011.

Kanji, G., "Implementation and pitfalls of total quality management", Total Quality Management, Taylor Francis, Volume 7, Issue 3**,** 1996, p.331-343

Low, K.E. and Baack, D., "Integrated Advertising, Promotion and Marketing Communications" Prentice Hall (2012)

Marquis, H., "How To Roll the Deming Wheel", source: http://www.itsmsolutions.com/newsletters/DITYvol5iss28.pdf, Vol. 5.28 • July 15, 2009.

Moen, R and Norman, C., "Evolution of the PDCA Cycle", http://pkpinc.com/files/NA01_Moen_Norman_fullpaper.pdf, 2006.

Sawyers, A., "Rental Fleets Fret About Supply, Hold Vehicles Longer", Automotive News, April 25, 2011: 38

Scott, C. and Kissane, P.A., "The Not So "Grave" results of the Graves Amendment for Rental Car Companies (Spring 2011 Litigation Quarterly), May 18, 2011.

Smart, Henderson, C., Winning, T., et al, "benchmarking learning and teaching: developing a method", quality assurance in education, emerald publishing (1993).

Stevenson, W., "Operations Management", McGraw-Hill Publisher, 2009, p.405-450.

Thompson, C., "Dollar thrifty Slashes its Chrysler Purchases", Automotive News, October 19, 2009.

Young, J.J. and Tavel, C.A, "Vicarious Liability of Rental Firms Preempted by Federal Law", Vol. 234-no.36, Monday, August 22, 2005.

DATA MINING TECHNIQUES FOR PREDICTION OF DIFFERENT CATEGORIES OF DERMATOLOGY DISEASES

Dinesh K. Sharma, University of Maryland Eastern Shore
H. S. Hota, Guru Ghasidas University, India

ABSTRACT

Diagnosis of health conditions of any person requires lots of expertise and accuracy because it is directly related to the life of a human being. Data mining techniques may extract relevant patterns from the health care data and can be used in the decision making process. This paper presents two data mining techniques: Support Vector Machine and Artificial Neural Network, along with an ensemble of these two techniques for classification of different categories of Erythemato-Squamous diseases. In the proposed ensemble model, we combine the above two models by using a confidential weighted voting scheme. Performance of all models and its ensemble model are calculated by using some error measures like accuracy, specificity and sensitivity as well as gain chart. Proposed models are tested on dermatology data set downloaded from the University of California at Irvine (UCI) machine learning repository. Empirical results show that the ensemble model has achieved a remarkable performance with the highest accuracy of 99.25% and 98.99% at training and testing stages respectively.

Keywords: *Erythemato-Squamous, Dermatology Data, Support Vector Machine (SVM), Artificial Neural Network (ANN), Ensemble Model.*

INTRODUCTION

In medical science, diagnosis of health conditions is a challenging task. Medical history data comprises of a number of tests essential to diagnose a particular disease and the diagnosis are based on the experience of the physician; a less experience physician can diagnose a problem incorrectly. Hence, it is possible for the health care industry to increase the advantages through the use of data mining techniques to develop a decision support system (DSS) which will diagnose the problem uniformly and intelligently. Therefore, an effective and intelligent health care DSS for diagnosis of different types of diseases is an essential requirement for health care.

Dermatology is a study of skin disease that is extremely complex and difficult to diagnose, and ultimately may be a leading cause of skin cancer. The six different categories of these diseases share the similar clinical features of erythema (Güvenir & Emeksiz, 2000; Elsayad, 2010b). Classification is a robust technique in medical mining. Even though most

studies are conducted in the field of classification to diagnose erythemato-squamous diseases, researchers still are working to find the best classifier for this kind of dataset (Übeyli, 2008 & 2009; Elsayad, 2010b).

Several authors (Güvenir, 1998; Güvenir & Emeksiz, 2000; Nanni, 2006; Elsayad, 2010b) have used data mining techniques for the diagnosis of erythemato-squamous disease. Güvenir et al. (1998, 2000) were the pioneers in this area and have done lots of works to develop a classifier. In their work (Güvenir & Emeksiz, 2000), they have developed a graphical user interface (GUI) with all visible information based on nearest neighbor, naïve Bayesian and voting features intervals-5 techniques to assist physician involved in this domain. A domain expert physician can use this DSS to diagnose the disease while intern-doctors can use it to verify their knowledge; model has achieved remarkably high classification accuracy of 99.2% based on the data set collected on their own. Other authors (Bojarczuk, 2001; Übeyli & Guler, 2005; Nani, 2006; Polat & Gunes, 2009; Übeyli & Dogdu, 2010; Barati et al., 2011) have also used various data mining techniques such as a decision tree, neuro-fuzzy, k-means clustering, and SVM techniques for the same purpose and achieved accuracy between 94.22% to 98.3%. Elsayad (2010b) has investigated this problem and developed a data mining based ensemble model using multilayer neural network, decision tree and linear discriminant analysis (LDA) techniques and got 98.23% classification accuracy. Recently, Xie and Wang (2011) applied support vector machine with novel hybrid feature selection methods and achieved 98.61% classification accuracy. Among all the above authors, Güvenir and Emeksiz (2000) achieved the highest classification accuracy of 99.2%.

In this study, we have presented classifications techniques with ensemble of Support Vector Machine (SVM) and Artificial Neural Network (ANN), for the classification of different categories of erythemato-squamous diseases. The dermatology dataset is taken from the University of California at Irvine (UCI) machine learning dataset (web source http://archive.ics.uci.edu/ml/datasets.html, last accessed on Jan 2012) to demonstrate the techniques. The classification accuracy obtained in this piece of research work using ensemble model is remarkably close to that of Güvenir and Emeksiz (2000) and is highest among all other models suggested in the literature. The proposed model is the only ensemble model using SVM and artificial neural network techniques with highest classification accuracy of 98.99%. Hence, our model is a competitive model as compared to the model developed by the different authors using data mining techniques.

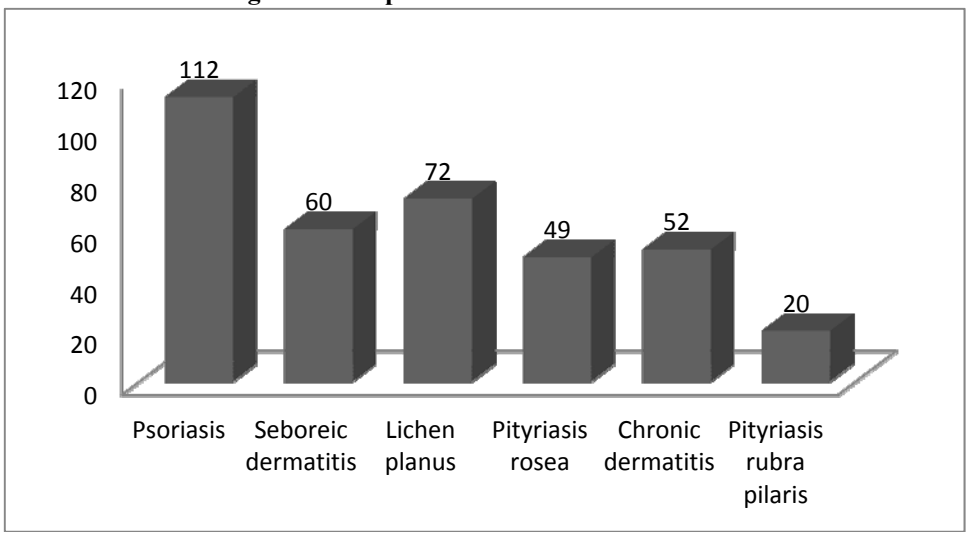
DATA SET DESCRIPTION

Each sample of dermatology dataset is classified into six categories: psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis and pityriasis rubra pilaris. The dataset contains 35 features out of which the first 34 features are considered as input and the 35th feature is considered as target (Class). Table 1 shows the features of dermatology dataset. There

are 365 instances in the data set containing all the six classes. Out of 365 instances, 112 instances belong to psoriasis class, 60 instances belongs to seboreic dermatitis class, 72 instances belong to lichen planus class, 49 instances belongs to pityriasis rosea class, 52 instances belong to chronic dermatitis class and 20 instances belong to pityriasis rubra pilaris class; sample distribution of various classes of the data set is shown in Figure 1.

Table 1: Dermatology dataset with its features	
1. Erythema	19. Parakeratosis
2. Scaling	20. Clubbing of the rete ridges
3. Definite Boarders	21. Elongation of the rete ridges
4. Itching	22. Thining of the suprapapillary epidermis
5. Koebner Phenomenon	23. Spongiform pustule
6. Polygonal Papules	24. Munro microabcess
7. Follicular Papules	25. Focal hypergranulosis
8 Oral mucosal involvement	26. Disappearance of the granular layer
9 Knee and elbow involvement	27. Vacuolisation and damage of basal layer
10. Scalp involvement	28. Spongiosis
11. Family history	29. Saw-tooth appearance of retes
12. Melanin incontinence	30. Follicular horn plug
13. Eosinophils in the infiltrate	31. Perifollicular parakeratosis
14. PNL infiltrate	32. Inflammatory mononuclear infiltrate
15. Fibrosis of the papillary dermis	33. Band-Like infiltrate
16. Exocytosis	34. Age
17. Acanthosis	35. Condition
18. Hyperkeratosis	

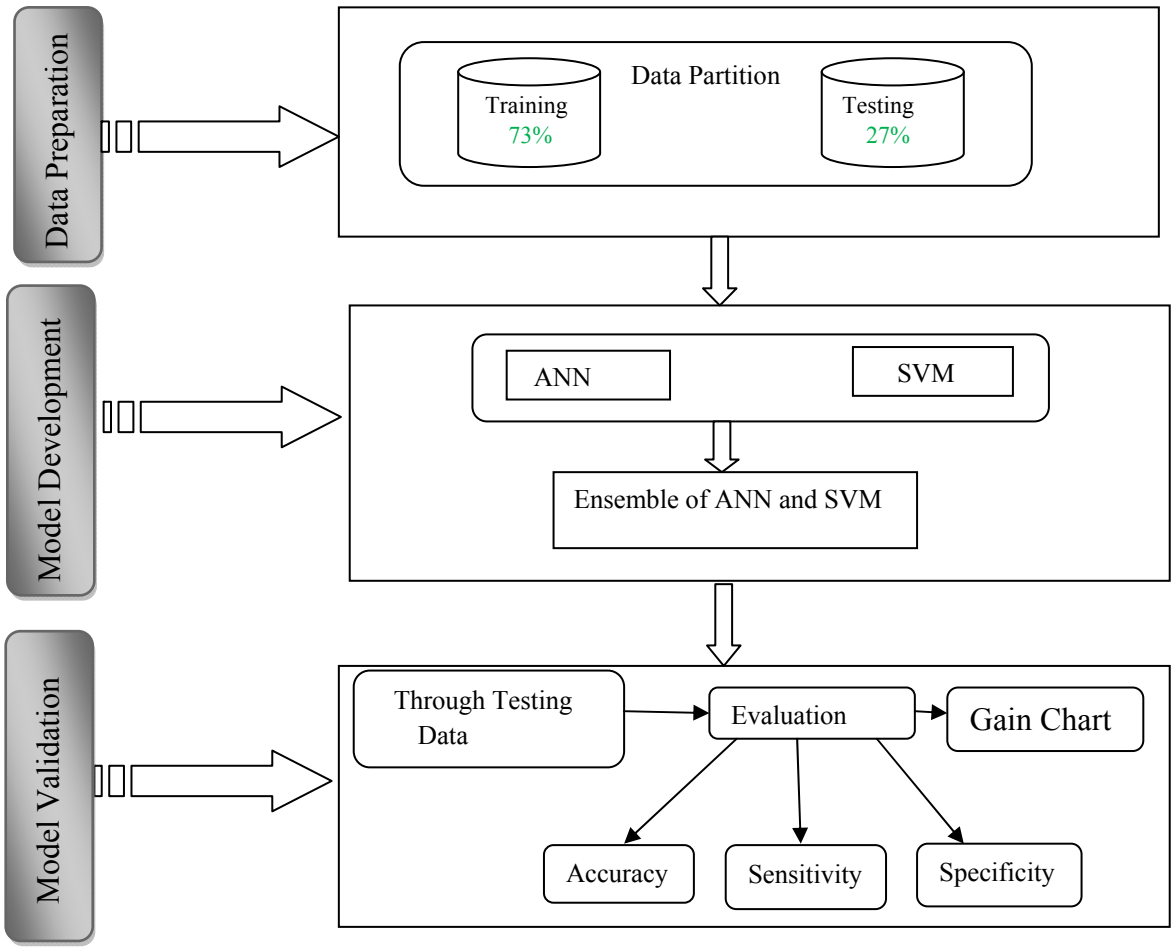
Figure 1: Sample class distribution of data set



MODEL IMPLEMENTATION

As shown in Figure 2 model implementation processes is divided into three different components namely: (i) Data preparation (ii) Model development and (iii) Model validation. In data preparation step, we prepare the data to feed it to the models. Two mutually exclusive datasets, a training dataset comprising 73% of the total dermatology dataset, and a testing dataset of 27% are created. Classification techniques are applied on this data. Out of 365 numbers of instances in dermatology dataset, 266 (73% Approx.) instances are taken as training set and 99 (27% Approx.) instances are taken as testing set. In model development and model validation steps, a classification model with desired accuracy is first developed then by feeding testing data, accuracy of the various developed models are measured using error measures like accuracy, sensitivity, and specificity. Classification model is an analysis technique used to describe data classes. In this procedure, a model or classifier is constructed to predict categorical labels.

Figure 2: Model implementation process



Data classification is a two-step process. In the first step, a classifier is built describing a predetermined set of data classes or concepts. This is a learning step (or training phase), where a classification algorithm builds the classifier by analyzing or "learning form" a training set made up of database tuples and their associated class labels. In the second step, the model is used for classification. A test set is used to test tuples and their associated class labels. The accuracy of a classifier on a given test set is the percentage of the test set tuples that are correctly classified by the classifier (Han et al., 2011).

Models are developed in two phases: training and testing, training refers to a building or development of a new model by using historical data, and testing refers to trying out the model on new, previously unseen data to determine its accuracy and physical performance characteristics. Training is typically done on a large proportion of the total data available, whereas testing is done on some small percentage of the data. The training dataset is used to train or build a model. Once a model is built on training data, the accuracy of the model on unseen data (testing) can be found. Different data mining techniques used to develop models in the model development stage is explored herewith these are mainly ANN and SVM based classification algorithms.

Artificial Neural Network (ANN)

ANNs are classic models of the way the nervous system operates. ANN is used for classification, prediction and pattern recognition. The network learns by examining individual records, generating a prediction or class for each record, and making adjustments to the weights whenever it makes an incorrect prediction/classification. This process is repeated many times, and the network continues to improve its predictions/classification accuracy until one or more of the stopping criteria have been met. In this research work, ANN is trained with Error back propagation algorithm (EBPA). The error back-propagation algorithm is different from another in respect to the process by which the weights are calculated during the learning period of the network. The general difficulty with the multilayer perceptions is calculating the weights of the hidden layers in an efficient way that would result in extremely small or zero output error. When the hidden layers are increased, the network training becomes more complex. To update weights, the error must be calculated. The error, which is the difference between the actual (calculated) and the desired (target) output is easily measured at the output layer. It should be noted that at the hidden layers, there is no direct information of the error. Therefore, other techniques should be used to calculate an error at the hidden layer, which will cause minimization of the output error, and this is the ultimate goal.

The training of the ANN is done in three stages the feed-forward of the input training pattern, the calculation and the back-propagation of the error, and updating of weights. The testing of the BPN involves the computation of feed-forward phase only. There can be more than one hidden layer (more beneficial), but one hidden layer is sufficient. Even though the training is

unusually slow, once the network is trained it can produce its outputs extremely rapidly. In this study, ANN is constructed of three independent layers: input, hidden and output layers with default control parameters.

Support Vector Machine (SVM)

SVM is a robust classification and regression technique (Mitra & Acharya, 2004) that maximizes the predictive accuracy of a model without over-fitting the training data. SVM is particularly suited to analyzing data with extremely large numbers (for example, thousands) of predictor fields. SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data is not otherwise linearly separable. A separator between the categories is found, and then the data is transformed in such a way that the separator could be drawn as a hyper plane. Following this, characteristics of new data can be used to predict the group to which a new record should belong. The Support Vector Machines (SVM) are a general class of learning architectures, inspired by the statistical learning theory that performs structural risk minimization on a nested set structure of separating hyper planes. Given a training data, the SVM learning algorithm generates the optimal separating hyper plane in terms of generalization error. The support vector machine is hugely popular as a high-performance classifier in several domains in classification. It obtains a set of support vector which characterizes a given classification task. The basic idea is to construct a hyper plane as the decision surface such that the margin of separation between positive and negative examples is maximized. Here, the error rate of a learning machine is considered to be bounded by the sum of the training error rate and a term depending on the Vapnik Chervonenkis (VC) 1 dimension. Given a labeled set of N training samples (X_i, Y_i), where $X_i \in R_n$ and $Y_i \in \{-1, 1\}$, the discriminate hyper plane is defined as:

$$f(X_q)= \Sigma Y_i \propto i \; K(X_q, X_i)+b$$

Here, K(.) is a kernel function and the sign of $f(X_q)$ determines the membership of query sample X_q. Constructing an optimal hyper plane is equivalent to determining all nonzero $\propto$ is which corresponds to the support vectors, and the bias b. The expected loss of making a decision is the minimum.

Ensemble Model

Two trained models, ANN and SVM, are ensemble together to form a new model known as ensemble model. An ensemble model is defined as a set of individually-trained classifier whose predictions are combined when classifying a new data. The ensemble model combines the output of several classifiers produced by weak learner into a single composite classification (Pal, 2007). It can be used to reduce the error of any weak learning algorithm. The purpose of combining all these classifiers together is to build an ensemble model which will improve classification accuracy as compared to each individual classifier. The two models are

combined by using a confidential weighted voting scheme (Elsayad, 2010) where weights are based on the confidence value of each prediction. Then the weights are summed, and value with the highest total is again selected. The confidence for the final selection is the sum of the weights for the winning values divided by the number of models included in the ensemble model.

PERFORMANCE MEASUREMENT

Performance of each individual classifier and ensemble classifier can be evaluated by using some well-known error measures: accuracy, sensitivity and specificity. These measures (Han et al., 2011) are defined by true positive (TP), true negative (TN), false positive (FP) and false negative (FN) under positive (P) and negative (N) cases. Table 2 represents a matrix showing cases of TP, TN, FP, and FN.

Table 2: Structure of confusion matrix for two class problems

Actual Vs Predicted	Positive (P)	Negative (N)
Positive (P)	True Positive (TP)	False Negative (FN)
Negative (N)	False Positive (FP)	True Negative (TN)

The following performance measures (Elsayad, 2010a,b) can be evaluated using the above table.

Accuracy measures the proportion of correct predictions considering the positive and negative inputs. It is highly dependent of the data set distribution which can easily lead to wrong conclusions about the system performance. It is calculated as follows:

$$\text{Accuracy} = \text{Total hits} / \text{Number of entries in the set} = (TP + TN) / (P + N) \quad (1)$$

Sensitivity measures the proportion of the true positives, that is, the ability of the system on predicting the correct values in the cases presented. It is calculated as follows:

$$\text{Sensitivity} = \text{Positive Heats} / \text{Total positive Heats} = TP / (TP + FN) \quad (2)$$

Specificity measures the proportion of the true negatives, that is, the ability of the system on predicting the correct values for the cases that are the opposite of the desired one. It is calculated as follows:

$$\text{Specificity} = \text{Negative hits} / \text{Total Negatives} = TN / (TN + FP) \quad (3)$$

EXPERIMENTAL RESULTS AND DISCUSSION

After presenting training dataset and testing dataset to each classifier along with ensemble classifier, a confusion matrix (Zurada & Lonial, 2005) is obtained to identify true

positive, true negative, false positive, and false negative values as shown in Tables 3 and 4 for training and testing data sets respectively.

Table 3: Confusion matrix of different classifiers for training dataset

Classifier Model	Predicted Class / Actual Class	Psoriasis	Seboreic Dermatitis	Lichen Planus	Pityriasis Rosea	Chronic Dermatitis	Pityriasis Rubra Pilaris
ANN	Psoriasis	75	0	0	0	1	0
	Seboreic Dermatitis	0	39	0	0	0	0
	Lichen Planus	0	0	57	0	0	1
	Pityriasis Rosea	0	3	0	34	0	0
	Chronic Dermatitis	1	0	0	0	38	0
	Pityriasis Rubra Pilaris	1	0	0	0	0	16
SVM	Psoriasis	76	0	0	0	0	0
	Seboreic Dermatitis	0	37	0	2	0	0
	Lichen Planus	0	0	58	0	0	0
	Pityriasis Rosea	0	0	0	37	0	0
	Chronic Dermatitis	0	0	0	0	39	0
	Pityriasis Rubra Pilaris	0	0	0	0	0	17
Ensemble of ANN and SVM	Psoriasis	76	0	0	0	0	0
	Seboreic Dermatitis	0	37	0	2	0	0
	Lichen Planus	0	0	58	0	0	0
	Pityriasis Rosea	0	0	0	37	0	0
	Chronic Dermatitis	0	0	0	0	39	0
	Pityriasis Rubra Pilaris	0	0	0	0	0	17

Each cell of the table contains the number of instances in a particular class predicted by the corresponding model. The predictions are compared with original classes to identify true positive, true negative, false positive and false negative. Say, for example, in the case of ANN classifier, correctly classified instance for psoriasis class is 75 and misclassified instance for the same class is 1 for training dataset; similarly correct classification for seboreic dermatitis is 39 and misclassification is 0 for training dataset. At the training stage SVM and ensemble models are producing exactly same output while at testing stage ensemble model is more accurate than SVM model, hence there is an improvement in case of ensemble model . If we will examine Table 4, the numbers of correctly classified instances from each of the classifiers are in an acceptable range. Based on these tables, accuracy, sensitivity and specificity (Anyanwu & Shiva, 2008) are calculated using equations 1, 2 and 3 respectively, and results are tabulated in Tables 5 and 6 for training and testing data sets respectively. From Tables 5 and 6, it is clear that all of the three statistical measures are almost 100%, except for some of the classes. Say, for example, in the case of ensemble model, sensitivity for seboreic dermatitis is 94.87% while specificity for pityriasis rosea is 97.80%. The statistical measures for all the classes of ensemble model are

either high or equal as compared to an individual model. This shows the highest performance of ensemble of SVM and ANN model.

Table 4: Confusion matrix of different classifiers for testing dataset

Classifier Model	Predicted Class / Actual Class	Psoriasis	Seboreic Dermatitis	Lichen Planus	Pityriasis Rosea	Chronic Dermatitis	Pityriasis Rubra Pilaris
ANN	Psoriasis	36	0	0	0	0	0
	Seboreic Dermatitis	0	20	0	1	0	0
	Lichen Planus	0	0	14	0	0	0
	Pityriasis Rosea	0	1	0	11	0	0
	Chronic Dermatitis	0	0	0	0	13	0
	Pityriasis Rubra	0	0	0	0	0	3
SVM	Psoriasis	36	0	0	0	0	0
	Seboreic Dermatitis	0	21	0	0	0	0
	Lichen Planus	0	0	14	0	0	0
	Pityriasis Rosea	0	3	0	9	0	0
	Chronic Dermatitis	0	0	0	0	13	0
	Pityriasis Rubra	0	0	0	0	0	3
Ensemble of ANN and SVM	Psoriasis	36	0	0	0	0	0
	Seboreic Dermatitis	0	21	0	0	0	0
	Lichen Planus	0	0	14	0	0	0
	Pityriasis Rosea	0	1	0	11	0	0
	Chronic Dermatitis	0	0	0	0	13	0
	Pityriasis Rubra	0	0	0	0	0	3

Another way to compare the performance of different classifiers is the gain chart (Zou et al., 2007). The gains chart plots the values in the gains percentage using the following equation:

(Hits in increment / total number of hits) x 100% (4)

The cumulative gain chart always starts at 0% and ends at 100% as we go from left to right. For a good model, the gains chart will rise steeply towards 100% and then levels off. Figure 3 show the cumulative gain chart of ensemble model for testing dataset. If we will examine this figure it is clear that it is following the characteristics of gain chart . The overall response rate of ensemble model is found to be higher than the individual model for all the classes.

Table 5: Comparative statistical measures for different classifiers for training dataset				
Model	Class of Disease	Sensitivity (%)	Specificity (%)	Accuracy (%)
ANN	Psoriasis	98.68	98.96	98.87
	Seboreic Dermatitis	100.00	98.67	98.87
	Lichen Planus	98.27	100.00	99.62
	Pityriasis Rosea	91.89	100.00	99.87
	Chronic Dermatitis	97.43	99.56	99.24
	Pityriasis Rubra Pilaris	94.11	99.60	99.24
SVM	Psoriasis	100.00	100.00	100.00
	Seboreic Dermatitis	94.87	100.00	99.24
	Lichen Planus	100.00	100.00	100.00
	Pityriasis Rosea	100.00	99.12	99.24
	Chronic Dermatitis	100.00	100.00	100.00
	Pityriasis Rubra Pilaris	100.00	100.00	100.00
Ensemble of ANN and SVM	Psoriasis	100.00	100.00	100.00
	Seboreic Dermatitis	94.87	100.00	99.40
	Lichen Planus	100.00	100.00	100.00
	Pityriasis Rosea	100.00	99.12	99.24
	Chronic Dermatitis	100.00	100.00	100.00
	Pityriasis Rubra Pilaris	100.00	100.00	100.00

Figure 3: Gain chart of ensemble model for testing data set

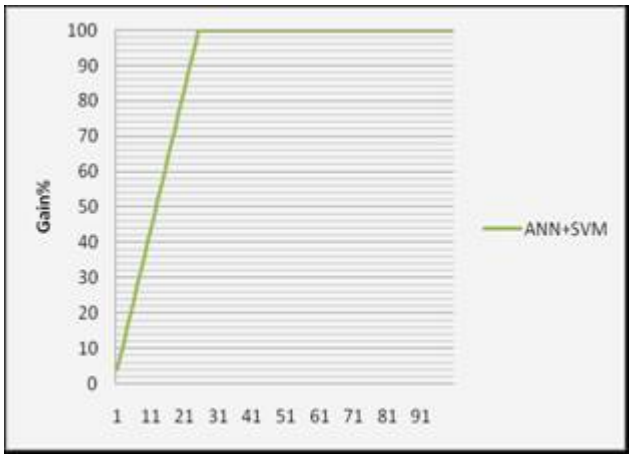
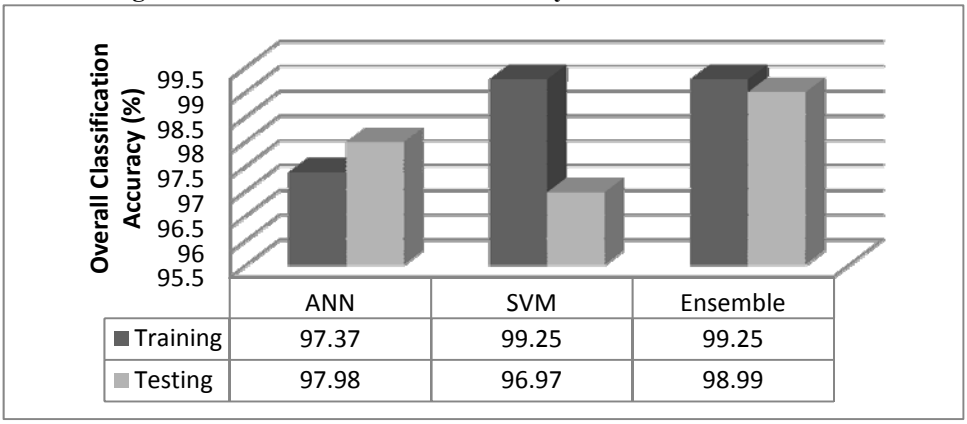


Table 6: Comparative statistical measures for different classifiers for testing dataset				
Model	Class of Disease	Sensitivity (%)	Specificity (%)	Accuracy (%)
ANN	Psoriasis	100.00	100.00	100.00
	Seboreic dermatitis	95.23	98.71	97.97
	Lichen planus	100.00	100.00	100.00
	Pityriasis rosea	91.66	98.85	97.97
	Chronic dermatitis	100.00	100.00	100.00
	Pityriasis rubra pilaris	100.00	100.00	100.00
SVM	Psoriasis	100.00	100.00	100.00
	Seboreic dermatitis	100.00	96.15	96.96
	Lichen planus	100.00	100.00	100.00
	Pityriasis rosea	75.00	100.00	96.96
	Chronic dermatitis	100.00	100.00	100.00
	Pityriasis rubra pilaris	100.00	100.00	100.00
Ensemble of ANN and SVM	Psoriasis	100.00	100.00	100.00
	Seboreic dermatitis	100.00	98.79	98.98
	Lichen planus	100.00	100.00	100.00
	Pityriasis rosea	91.66	100.00	98.98
	Chronic dermatitis	100.00	100.00	100.00
	Pityriasis rubra pilaris	100.00	100.00	100.00

Table 7: Total classification accuracy of training and testing dataset					
Model	Cases	Training dataset		Testing dataset	
		Number of instances	Accuracy (%)	Number of instances	Accuracy (%)
ANN	Correct	259	97.37	97	97. 98
	Wrong	7	2.63	2	2.02
SVM	Correct	264	99.25	96	96.97
	Wrong	2	0.75	3	3.03
Ensemble of ANN and SVM	Correct	264	99.25	98	98.99
	Wrong	2	0.75	1	1.01

Figure 5: Overall classification accuracy of various classifiers



CONCLUSIONS

In this study, we applied two different data mining techniques: SVM and ANN on dermatology dataset. By combining scores of these two models, a more precise ensemble model is obtained. They are all combined by using confidential weighted voting scheme. Table 7 shows the overall classification accuracy for training and testing data set respectively for the three models. It is found that accuracy (Kaur & Wasan, 2006) of ensemble model is 99.25% for training dataset while it is 98.99% for testing dataset which is higher than all other individual models. Accuracy of the ANN model is 97.37% and 97.98% respectively for training and testing datasets respectively while accuracy of the SVM model is 99.25% and 96.97% respectively for training and testing datasets. Performance of each model has also been investigated with the help of a gain chart. In all respects, the ensemble model is performing well. Hence, this model can be recommended for the classification of dermatology data set as decision support system in health care. To further improve the classification accuracy of model soft computing tools like genetic algorithm, fuzzy logic and particle swarm optimization and its hybrid techniques can be used.

REFERENCES

Anyanwu, M.N., & Shiva, S.G. (2008). Comparative analysis of serial decision trees classification algorithms. *International Journal of Computer Science and Security,* 3(3), 230-240.

Barati, E., Saraee, M., Mohammadi, A., Adibi, N, & Ahamadzadeh, M.R. (2011). A survey on utilization of data mining approaches for dermatology (Skin) disease prediction, Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Health Informatics (JSHI), March Edition, 1-11.

Elsayad, A. M. (2010a). Predicting the severity of breast masses with ensemble of Bayesian classifiers. *Journal of Computer Science*, 6(5), 576-584.

Elsayad, A. M. (2010b). Diagnosis of erythemato-squamous diseases using ensemble of data mining methods. *ICGST-BIME Journal*, 10(1), 13-23.

Kaur, H. & Wasan, S.K. (2006). Empirical studies on applications of data mining techniques in health care. *Journal of Computer Science*, 2(2), 194-200.

Güvenir, H., Demiröz, G., & Ilter, N. (1998). Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. *Artificial Intelligence in Medicine*, 13(3) 147-165.

Guvenir, H., & Emeksiz, N. (2000). An expert system for the differential diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, 18**(1)** 43–49.

Han, J., Kamber, M., & Pei, J. (2011). Data mining: *Concepts and Techniques,* 3rd edition, Morgan Kaufmann, San Francisco, CA. USA.

Mitra, S., & Acharya, T. (2004). Data mining multimedia: Soft computing and bioinformatics. New York: John Wiley.

Nanni, L. (2006). An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. *Neurocomputing*, 69, 842–845.

Pal, M. (2007). Ensemble learning with decision tree for remote sensing classification. *World Academy of Science, Engineering and Technology* 36, 258-260.

Polat, K., & Güneş, S. (2009). A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. *Expert Systems with Applications*, 36(2) 1587-1592

Übeyli, E., & Guler, I. (2005) Automatic detection of erythemato-squamous diseases using adaptive neuro fuzzy inference system. *Computer in biology and medicine*, 35, 421-433.

Übeyli, E. (2008). Multiclass support vector machines for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, 35(4) 1733–1740

Übeyli, E. (2009). Combined neural networks for diagnosis of erythemato-squamous diseases. *Expert Systems with Applications*, 36(3) 5107–5112.

Übeyli, E., & Dogdu, E. (2010). Automatic detection of erythemato-squamous diseases using k-means clustering. *Journal of Medical Systems* 34, 179-184.

UCI (2012). Web source: http://archive.ics.uci.edu/ml/datasets.html,last accessed on Jan 2012).

Xie, J., & Wang, C. (2011). Using support vector machine with novel hybrid features selection method for

Diagnosis of erythemato-squamous diseases. *Expert System with Applications,* 38(5), 5809-5815.

Zou, K.H., O'Malley, A. J., & Mauri, L., (2007). ROC analysis for evaluating diagnostic test and predictive models. *Journal of the American Heart Association*, 115:654-657.

Zurada, J., & Lonial, S. (2005). Comparison of the Performance of several data mining methods for bed debt recovery in the health care industry. *Journal of Applied Business Research*, 21(2), 37-53.