

ACCOUNTING | HEALTH CARE | NON-PROFITS | E-COMMERCE

KELLEY SCHOOL OF BUSINESS

Fall 2012

OnAnalytics

diagnose. predict. optimize.

**ANALYTICS
SHARPENS INSIGHTS**

Four applications of converting
data into insights



Welcome

The past 15 years have seen an explosive growth in data creation, collection, and analysis. This expansion, coupled with almost unlimited computer processing capabilities and increasing sophistication in analytics software, has led to a dramatic shift in the relationship between business and analytics. Executives and managers have begun to recognize data not merely as an interest area for statisticians and modelers but more directly as a vital business asset. Business data, which now contains an incredible breadth and depth of information about almost every facet of business activity, holds value that can be unlocked through data analysis and business analytics to provide a competitive advantage.

As a result of this widespread recognition of the value of business data, we are seeing an evolution of business analytics that in many ways mirrors that of information technology. IT was once a specialty area that few businesses incorporated in house. Today, IT is an integral part of business operations that savvy organizations know they must leverage in order to achieve efficiency and outperform their competitors. The very same can be said for business analytics. An organization that ignores or underutilizes business analytics may, quite frankly, be jeopardizing their potential for growth, competitiveness, and greater efficiencies.

Business analytics has always been a key strength for the Kelley School. Although the term “business analytics” is relatively new, the practice of using statistical data analysis and analytical modeling to support insightful business decisions is what the Kelley School has always done extremely well. Here at Kelley, we are accelerating the evolution of business analytics by equipping our students with leading-edge tools and strategies. For our MBA program, we have developed one of the country’s first majors in business analytics. In the first year of the program, in which we also offered a major and a minor in business analytics, we attracted more than 70 MBA students to the field. We have also created groundbreaking business analytics certificate programs with our partners from Deloitte Consulting and Booz Allen Hamilton. Additionally, we are partnering with the Indian Institute of Management at Lucknow to offer a graduate program in business analytics directly to students in India. This partnership builds on the global and online capabilities that we have continued to strengthen through our online program, Kelley Direct.

Kelley’s Institute for Business Analytics and the school as a whole are committed to moving the field of business analytics forward across the globe. Doing so will allow organizations everywhere to leverage the rich data resources that are now available, resulting in greater insight, efficiency, and effectiveness in every sector. And that sort of transformation is – and always has been – what Kelley is all about.

Ash Soni

Associate Dean of Information Technology; Professor of Operations & Decision Technologies; ArcelorMittal Faculty Fellow



From the Editors

Welcome to the second issue of *OnAnalytics*, which focuses on four application areas: accounting, health care, non-profits, and e-commerce. For each area, a Kelley School faculty member presents a summary of a recently completed study that utilizes advanced analytics methods. To complement the faculty research reports, we invited leading practitioners to provide different perspectives and to discuss how analytics is affecting practice. The articles from practitioners precede each of the faculty research reports.

Advanced analytic techniques are utilized to address a broad array of accounting, finance, and investment problems. One important segment of such applications is in forensic accounting, specifically in screening transactions and documents for potential fraud. Joseph Cheriathundam, managing director at the StoneTurn Group, discusses how analytics techniques are used to search through account transactions for potential fraud. In addition, he highlights the need for domain expertise as a complement to automated data-driven procedures. James Wahlen, a Kelley School professor, next reports on the development of a model using six accounting metrics to predict future firm performance. The model produced stock portfolios, which significantly outperformed the portfolios created using consensus ratings by financial analysts.

Firms such as Deloitte Consulting have found numerous applications for analytics in the health care field. Anwer Khan, Aditya Sane, Vik Wadhvani, and Mark Zozulia draw on their experience at Deloitte LLP to describe recent and emerging applications of analytics that promise to enhance the quality of care and help to control expenses. Following that article, Kelley School professor Kurt Bretthauer describes how he applied a queuing model to the problem of managing patient flow in a large urban hospital. The model resulted in specific recommendations regarding the number and allocation of beds among four types of inpatient units. With this application, he was able to optimize resource use and the availability of hospital facilities to patients.

Advanced analytic techniques are also being increasingly used to address problems in government and not-for-profit organizations. The objectives in such organizations are typically different from and sometimes more complex than those associated with businesses, but the potential for improved operations is substantial nevertheless. Tom Preston, an executive with the Booz Allen consultancy, draws on his extensive experience with projects for the federal government to provide an overview of the emerging opportunities for applying advanced analytics in that domain. Next, Kelley school professor Alfonso J. Pedraza-Martinez applied advanced data analytics and optimization methods to a vehicle replacement problem faced by the International Committee of the Red Cross. His approach indicates the potential for substantial cost savings. He also highlights an all-too-frequent problem of implementing the results of analytics models in the face of organizational and policy constraints.

The last set of articles in this issue of *OnAnalytics* is concerned with the application of analytics to marketing and consumer behavior issues in e-commerce. Vaibhav Gardé, an analytics executive at FedEx, discusses the evolution from web analytics to much broader opportunities for interaction using digital media, resulting in a proliferation of communication channels that are enabling more personalized, real-time interaction with consumers. Babur De los Santos and Matthijs Wildenbeest, both faculty members at the Kelley School, summarize their recent research on Internet shopping. Their findings indicate that even on the Internet, where price comparisons are extremely easy, consumers engage in a more nuanced decision-making process and that price alone does not drive buying decisions.

We hope that you enjoy the perspectives from both research and practice in this issue of *OnAnalytics*.

Frank Acito

Professor of Marketing
Max Barney Faculty Fellow
acito@indiana.edu

Vijay Khatri

Associate Professor of Information Systems
Arthur M. Weimer Faculty Fellow
vkhatri@indiana.edu

OnAnalytics

diagnose. predict. optimize.

6



Using Context to Improve Fraud Tests

Joseph Cheriathundam,
StoneTurn Group

Business analytics is increasingly employed in identifying abuses and minimizing losses that result from fraud. One of the challenges in fraud detection, however, is the prevalence of false positives, which can lead to fruitless and disruptive investigations. Joseph Cheriathundam of StoneTurn Group outlines a strategy for creating more meaningful fraud detection tests by developing a thorough understanding of the context in which transactions occur.

8

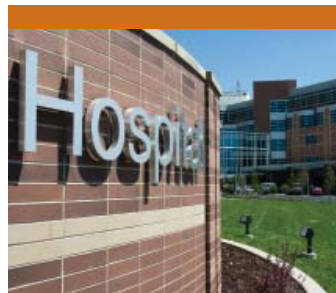


A Superior Strategy for Predicting Earnings and Choosing Stocks

James M. Wahlen

Motivated by skepticism about analysts' consensus recommendations, the researchers set out to craft a superior investment strategy derived from financial statement data. They used a six-signal scoring system to predict future earnings increases and tested their model against financial analysts' consensus recommendations. The researchers' strategy significantly outperformed consensus recommendations, suggesting that analysts are not making full use of the information available when they create consensus recommendations for investors.

10



Applied Analytics in the Health Care System

Anwer Khan, Aditya Sane,
Vik Wadhvani, and
Mark Zozulia,
Deloitte Consulting

The enormous amounts of data related to health care in the United States encompass not only direct patient care information but also the activities of health insurance payers, life science organizations, and public health agencies. Anwer Khan, Aditya Sane, Vik Wadhvani, and Mark Zozulia of Deloitte Consulting describe some of the emerging business analytics strategies and applications for traversing through this maze of data to increase efficiencies and patient outcomes.

12



Unblocking Patient Flow

Kurt M. Bretthauer

The efficient flow of hospital patients between different units of care is not only a crucial element of effective treatment but also an important consideration in conserving resources and managing revenue. With this study, the researchers consider the "blocking" problem that occurs when at-capacity units cannot accommodate patients, causing some patients to remain in a higher unit of care than they require and others to be turned away from the hospital. By creating a simplified and highly accurate model of patient flow, the researchers provide a tool for determining the optimal mix of beds within a hospital's budget constraints and specified management objectives.

14



Leveraging Data in the Public Sector

Tom Preston,
Booz Allen

With annual expenditures of more than \$3 trillion, the federal government stands to benefit enormously from the descriptive, diagnostic, and predictive capabilities of business analytics. Tom Preston leads Booz Allen's Enterprise Resource Planning Advisory Services and works on a number of high-level U.S. government projects. Here, he describes some of ways in which business analytics is being implemented in the public sector, as well as the promising new technologies that offer opportunities to improve government efficiency and effectiveness.

16

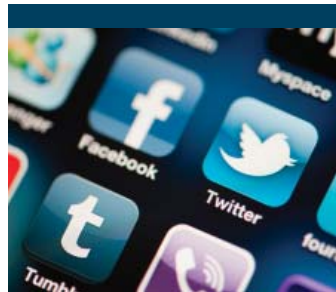


Vehicle Replacement in Humanitarian Operations

Alfonso J. Pedraza-Martinez

Ground transportation is among the most critical factors in delivering international humanitarian aid. In developing their vehicle replacement policies, however, organizations like the International Committee of the Red Cross have relied on manufacturers' recommendations derived from commercial-sector data. With this study, the researchers examine data on vehicle purchase, use, and sale in the context of humanitarian operations in Afghanistan, Ethiopia, Georgia, and Sudan. Their optimization model reveals that what works for commercial fleets does not necessarily apply to vehicles delivering humanitarian aid.

18



Digital Analytics: Driven by Search and Social Media

Vaibhav Gardé,
FedEx

The field of web analytics has grown and evolved so rapidly that it has already taken on a new identity as "digital analytics," emphasizing not only traditional website interactions but also the widespread use of mobile devices and social media. Vaibhav Gardé, a Marketing Principal in the Interactive Marketing practice at FedEx, describes the current trends in digital business analytics and the exciting new career opportunities they present.

20



Estimating Consumer Search Costs in Online Markets

Babur De los Santos and
Matthijs R. Wildenbeest

The massive amount of data generated by e-commerce offers an unprecedented opportunity to test theoretical models of consumer search behavior. Using data on Internet bookstore browsing and transactions, the researchers discover that the sequential search model favored in the literature does not fit with observed online shopping activity. Searches themselves are in fact rare, as customers tend to visit only one store before making a purchase. When they do browse, the fixed sample size model provides a better fit for their search behavior.

Using Context to Improve Fraud Tests

Joseph Cheriathundam

Joseph Cheriathundam is a Managing Director specializing in financial analytics for StoneTurn, an international dispute consulting and financial investigations firm. He began his career as an electrical engineer performing tests and evaluations of military systems, later applying his evaluative skills to financial consulting for PricewaterhouseCoopers, Deloitte, and now StoneTurn. Cheriathundam obtained a business analytics certification from Kelley this year.

jcheriathundam@stoneturn.com



One of the many ways that businesses can use analytics to improve their financial health is in the area of occupational fraud detection. This type of fraud, in which an employee abuses his or her role for personal gain, is estimated to cost organizations an average of 5% of annual revenues, or a worldwide total loss of \$3.5 trillion (Association of Certified Fraud Examiners, 2012). There are many sophisticated techniques now used to detect occupational fraud, including:

- Segmentation and classification of transactions to model normal and abnormal patterns.
- Encoding programs to deliver alerts when abnormalities are detected.
- Testing for missing/anomalous values in sequential and other data.
- Analyzing text data – for example, addresses – to identify suspect entries.

Many analytics-driven fraud detection methods search for outliers – data points that don't fit the normal transaction patterns. The problem

is that dealing with massive amounts of data introduces a number of naturally occurring outliers. This creates one of the biggest challenges when introducing today's analytics into the issue of fraud: false positives. Specifically, false positives lead to unnecessary costs for a company as they review the related employees/ transactions to determine the validity of the alerts. For this reason, it is vital as analysts that we seek ways to limit alerts stemming from such errors.

What we don't want to do, however, is abandon these tests altogether as there is often a value to be gained from them. A better strategy can commonly be derived by complementing your quantitative data with qualitative data to reduce the occurrence of false positives. I'll offer two examples of common fraud tests that can deliver useless results when they are applied without considering the business being analyzed as well as some potential strategies that could prevent these occurrences.



Is the norm abnormal?

One fraud test that is commonly applied to payments is a test for weekend activity. Because corporations are generally not inclined to pay overtime for their accounts payable department, it is reasonable to conclude that payments recorded on a weekend date should be regarded as suspect.

But what if you're dealing with a situation in which days of the week are irrelevant? We were faced with this question recently when we assisted in a review of military spending in Iraq. For those working on the ground in a war zone, the concept of a weekend is non-existent, and a test of weekend activity would have devalued the analysis by flooding our output with false positives.

A conclusion that is often reached in these scenarios is to throw out the test. However, as we looked more closely at the individual source systems, it made sense to include the tests for the disbursement systems that were based in the United States rather than in-theater. By taking that extra step of investigation, we applied the test where it provided value and ultimately delivered a more meaningful output to our client to review for potential fraud.

Are the data appropriate for the test?

Another familiar test involves a mathematical concept known as Benford's Law. Simply put, Benford's Law states that in a set of naturally occurring numbers, the leading digit is distributed in a specific, non-uniform way, with a greater likelihood of occurrence

for the smaller numbers (1 or 2) than larger (8 or 9). Without going into detail as to why this is (contact me if you'd like an explanation), I can say that it bears out in many circumstances relating to expenses.

Recently, however, we had a client for whom there was a spike in their data for expenses with a 6 as the leading digit. In a strict application of the test, this would be a flag for fraud. But we looked at some of the output and found that the company's cell phone packages fell into the range of \$60.00-\$69.99 and were skewing the outcomes. Again, we did not dispense with the test altogether, but we did pull out the cell phone expense records, and the resulting findings provided a more actionable set of results.

Conclusion

There's a lot of enthusiasm at the moment for business analytics, and rightly so. There is increased access to more data and more processing power than ever before. As a result, it is feasible to perform tests that would never have been possible in previous decades. To make effective use of the unprecedented data and computational systems available, analysts need to ensure that communication and contextualization are essential components of their process. They need to understand client operations and design tests to fit not only the data but also the actual business that the data reflects. Doing so will lead to more meaningful and actionable results for our clients, which is critical as we continually seek to optimize the value provided by business analytics.

A Superior Strategy for Predicting Earnings and Choosing Stocks



James M. Wahlen

Professor of Accounting;
James R. Hodge Chair

jwahlen@indiana.edu

Analysts' consensus recommendations are a tool used by investors, but do they really reflect the full information available from financial statement reports? With this study, Jim Wahlen and Matt Wieland developed and tested a rating system that predicts future changes in earnings based on financial statement information. When these future earnings predictions are used to design an investment strategy, the researchers' stock portfolios significantly outperformed portfolios formed using consensus recommendations, suggesting that consensus recommendations fail to incorporate information that could more reliably predict future earnings changes and identify under- or over-priced stocks.

Statement of the problem

Consensus recommendations from financial analysts do not follow a normal distribution but instead skew toward "strong buy," "buy," and "hold" categories, with only a tiny percentage of stocks identified as "underperform" or "sell." Additionally, the "hold" category is amorphous; it could be seen as implying neutrality, ambivalence, uncertainty, or a veiled low opinion of the stock's potential.

These observations and prior research indicate that consensus recommendations are not reliably informative and may even be misleading. A more useful predictive tool would transparently process information from financial statements to offer a reliable indication of future changes in earnings, which in turn provide a fundamental basis on which to identify stocks to buy or sell.

Data sources used

The researchers gathered consensus recommendations for firm-years 1994 through 2005 from the Institutional Brokers Estimate System (I/B/E/S). They also collected financial statement information for the same period from Compustat, and stock prices, returns, and market capitalization data from the Center for Research in Security Prices (CRSP). The sample, containing intersections from the three sources, consisted of 25,168 firm-year observations.

Analytic techniques

Abnormal returns for each firm-year were calculated by compounding monthly raw returns over a one-year holding period and then subtracting the compounded returns on the CRSP size-based decile portfolio. The researchers calculated descriptive statistics on 1-year ahead cumulative and abnormal returns, financial statement analysis variables, market multiples, stock returns, and analysts' earning forecasts across all consensus recommendation levels.

To predict future earnings increases, the researchers used a previously developed six-signal scoring model known as the predicted earnings increase score (PEIS). Firms were ranked into quintiles for each of the six criteria:

1. *Return on net operating assets (RNOA)*. Extreme returns tend to move to mean reversion over time; a firm generating extremely high RNOA will likely encounter competition while extremely low RNOA must implement improvements in profit margins and operating efficiency.
2. *Gross margin signal (GM)*. The growth of gross margins relative to growth in sales reflects the firm's control of production costs relative to product prices.
3. *Selling, general, and administration signal (SG&A)*. This signal reflects changes in operating costs relative to sales. In the presence of sales growth, an increase in the SG&A percentage indicates poor control of overhead whereas a decrease implies cost control and operating leverage. If sales decline, the signals reverse, with an SG&A percentage increase indicating managerial optimism and a decrease suggesting that managers are pessimistic about future earnings.
4. *Asset turnover ratio (ATO)*. Changes in the efficiency of the firm's assets tend to predict the direction of future profitability.
5. *Growth in net operating assets (G^{NOA})*. G^{NOA} interacts negatively with RNOA; low G^{NOA} relative to RNOA indicates increasing efficiency while high G^{NOA} relative to RNOA implies inefficient buildup of operating assets.
6. *Accruals (ACC)*. Measured as the difference between operating income and cash flow while controlling for RNOA, accrual levels are negatively associated with future earnings predictions.



Matt Wieland

Assistant Professor of
Accounting
Kelley School of Business,
Indianapolis

mmwielan@iupui.edu



For each of the six criteria, firms in the lowest and highest quintiles received a score of -1 or +1 according to the direction of the prediction, while firms in the middle three quintiles received scores of 0. The researchers summed the scores of all six signals for each firm-year to compute PEIS. Firms received scores between -6 and +6, with higher scores indicating a likelihood of future earnings increases.

The researchers computed descriptive statistics of 1-year ahead RNOA for the PEIS quintiles for the full sample and each consensus recommendation level. Next, the researchers tested the abnormal returns of portfolios formed based on PEIS. Their primary trading strategy, labeled the “fundamental strategy,” took equally weighted long positions on firms in the highest PEIS quintiles and short positions on firms in the lowest PEIS quintiles. Their benchmark strategy, labeled the “buy/sell strategy,” captured the effects of following consensus recommendations by taking long positions on the strong buy and buy recommendations and short positions on the sell recommendations. Within the stocks labeled “hold” in the consensus recommendations, the researchers tested a “conditional hold strategy” taking long (short) positions on the highest (lowest) PEIS quintile, and compared it with an “unconditional hold strategy” taking long positions for all “hold” stocks.

The researchers also implemented a regression equation to control for the annual effects of market-to-book ratio, earnings yield, and beta.

Results

The descriptive statistics revealed that the mean recommendation value is 2 (buy), indicating that consensus analysts tend to advise investors to purchase shares of the firms they follow. The mean (median) size-adjusted abnormal return generated by the full sample was 2.1% (-7.9%). Moreover, 1-year ahead mean and median cumulative and abnormal returns tended to vary inversely with recommendation levels, suggesting that consensus recommendations were not reliable indicators of future stock return performance.

After computing the PEIS scores, the researchers observed striking differences in the financial performance measures between firms in the top and bottom quintiles, with firms in the lowest quintile exhibiting 0.4% RNOA and almost no sales growth and those in the highest quintile exhibiting 29.6% RNOA and nearly 40% sales growth. 63.9% of firms in the highest quintile reported an earnings increase compared to only 51.3% of firms in the lowest quintile.

The scoring system was most effective in predicting performance within the “hold” category, with 66.3% of firm-years in the highest PEIS quintile generating an earnings increase the following year in contrast to only 42.5% of those in the lowest quintile. Applying PEIS with the fundamental strategy yielded positive changes in the abnormal returns across each of the consensus recommenda-

“In looking at consensus analysts recommendations, we found that they recommended buys and strong buys for 73.4% of the stocks they followed. That proportion seemed high, especially when they labeled another 25% of the stocks as holds and it wasn’t at all clear what that category was supposed to mean. My coauthor and I decided to start over with the fundamentals of financial analysis to see if we could build up a more useable mechanism to predict future changes earnings and pick which stocks to buy or sell. We discovered that our predicted earnings increase score significantly outperformed analysts consensus recommendations.

tion categories. Within the hold category, conditioning on PEIS increased abnormal returns from 6.8% to 22.4%.

For the full sample, the fundamental strategy using PEIS significantly outperformed the buy/sell strategy, with average annual abnormal returns of 9.8% using PEIS in contrast to -9.0% following consensus recommendations. Use of the regression showed that after controlling for other predictors of future returns, the buy/sell strategy yielded positive abnormal returns in 8 out of 12 years with an average abnormal return of 0.9%, while the fundamental strategy yielded positive abnormal returns in all but one year with an average abnormal return of 10.9%.

The unconditional hold strategy yielded an 8.5% average abnormal return, which was higher than expected given the ambiguity of the recommendation. The conditional hold strategy, however, yielded an average abnormal return of 19.3% and positive abnormal returns in all 12 years, outperforming the unconditional hold strategy by an average of 10.8 percentage points per year.

Business implications

The strong performance of the PEIS-reliant strategies as compared to consensus recommendations suggests that analysts are failing to fully incorporate the information available in financial statements into their recommendations.

The results of this study could be useful to analysts in improving their strategies for forming recommendations. PEIS could also be used directly by investors to improve their returns over a strategy of following consensus recommendations. Another potential application is to use the conditional hold strategy to fill in the information gap created by the unclear “hold” category.

By testing different means of predicting earnings increases, this study not only demonstrates ways to apply financial statement data to investment strategies but also challenges assumptions about the advice currently provided to investors.

James M. Wahlen and Matthew M. Weiland, “Can financial statement analysis beat consensus analysts’ recommendations?” *Review of Accounting Studies*, 16 (1), pages 89-115, 2011.

Applied Analytics in the Health Care System

Anwer Khan

Anwer Khan, Director, Deloitte Consulting, is an experienced health care and technology executive who holds seven nationally registered patents in the area of health care informatics.

anwkhan@deloitte.com

Aditya Sane

Aditya Sane is Manager of the Deloitte Analytics Institute. He holds patents for algorithms in medical devices, and has published scientific papers in signal processing techniques.

asane@deloitte.com

Vik Wadhvani

Vik Wadhvani, Senior Manager, Deloitte Consulting, has extensive experience in managing large scale IT and business transformation efforts. Wadhvani obtained a business analytics certification from Kelley this year.

vwadhvani@deloitte.com

Mark Zozulia

Mark Zozulia, Principal, Deloitte Consulting, is the Executive Sponsor for Indiana University Recruiting and a National leader in their Information Management practice.

mzozulia@deloitte.com



Total health care spending in the U.S. — by individuals, corporations, and federal, state, and local governments — continues to amount to a significant proportion of the GDP, and is expected to reach 19.5% of GDP by 2017. Individual spending averaged \$8,233 per year according to a recent report by the Organization for Economic Co-operation and Development. As such, all sectors of the health care industry are actively engaging in addressing cost and quality challenges using business analytics. Health care providers and related hospital-like settings must improve quality while reducing costs. Managed care and health insurance payers also struggle to reduce costs, but attempt to differentiate themselves by launching and administering disease management and wellness programs more efficiently and effectively than their competitors. Life science organizations, meanwhile, are trying to stay

ahead of patent cliffs by searching for derivative products across new therapeutic classes to serve more targeted cohorts.

A number of ways in which business analytics is currently applied across different fields of health care include:

- **Health Care Providers:** Workflow improvement, readmission prediction and management, reducing drug interactions, supply chain optimization, and patient load forecasting.
- **Health Insurance Payers:** Targeted care management and wellness programs; readmission reduction; targeted marketing; formulary optimization; network design; provider profiling and quality based reimbursements; and fraud, waste, and abuse detection.



- **Life Science Organizations:** Comparative effectiveness, signal detection, accelerated drug approvals with phase 4 virtual trials, clinical trial recruitment and enrollment, and translational medicine collaboration.
- **Public Health Agencies:** Disease outbreak prediction; chronic conditions projection; adverse events monitoring; and fraud, waste and abuse detection (for Medicare/Medicaid).

Health care data proliferation

Business analytics goes beyond traditional operational reporting to offer health care organizations the detailed insights into trends and predictors needed to address these and other challenges. Through analytical models and advanced data visualization, applied analytics can provide a range of approaches and solutions from looking backward to evaluate performance to forward-looking scenario planning and predictive modeling. The digital health care ecosystem uses and creates an endless stream of information from a large number of sources, such as medical records, claims systems, clinical registries, financial and payment reimbursement systems, imaging applications, and home health devices, just to name a few. Analytic methods such as classification, logistic regression, decision trees, neural networks, segmentation, time series, sequence models, outlier detection, and natural language processing are some of the tools applied to create actionable recommendations from this vast universe of data. The goal of applied health care analytics, however, is not to analyze everything, but rather to analyze the right things at the right time.

Patient readmissions

Taking inpatient readmissions as an example, research shows that nearly 20% of inpatient admissions result in readmission. Approximately 90% of readmissions are preventable. These unplanned readmissions have an estimated \$42 billion national cost. Moreover, financial penalties and bonuses based on quality outcomes are now being enforced. For instance, Medicare payments for readmissions in select areas will be reduced beginning October 2012. Providers will assume part of the financial risk, which averages \$16,000 per readmission.

Recently, our team used business analytics to create a readmission model using a mathematical algorithm pitted against a set of iterative decision trees that calculated the propensity of readmission by disease class. A multi-facility hospital system in the western U.S. customized this model to its own data and was able to identify 4,000 cases for intervention. After realigning case manager workflows to target the intervention cases, this hospital system projects an incremental net savings of more than \$1 million per facility.

Conclusion

Business analytics will play a crucial role in improving health care outcomes going forward. Ultimately, such improvement will depend on (a) solid understanding, integration, and management of existing and newer data sources; (b) application of advanced analytical models and techniques; and (c) integration of analytical outputs with existing operational processes (e.g. hospital care management workflows). To maximize effectiveness and reduce problems like preventable readmissions, health care organizations must invest in optimizing their analytics infrastructure and exploring emerging architectures to support the utilization of applied analytics.

Unblocking Patient Flow



Kurt M. Bretthauer

Professor of Operations & Decision Technologies;
Kimball Faculty Fellow

kbrettha@indiana.edu

Hospitals face the dual imperative of delivering the most effective medical care while attempting to minimize costs in an environment of limited resources and increasing expenditures. One crucial determination is the mix of inpatient beds provided for different levels of care. Viewed in isolation, post-acute or skilled nursing units may not appear as vital or as cost-effective as more intensive treatment units. If these post-acute care units fill, however, patient flow will be blocked out of – and into – more acute care. With this study, the researchers developed a model for minimizing this blocking while also addressing a variety of management considerations such as overall revenue, patient acuity, and the objective of not turning patients away from the hospital.

Statement of the problem

Within a hospital setting, patients typically move through several levels or units of care. Hospitals must accurately predict patient flow in order to determine the appropriate number of beds in each service unit. Too many beds in a given unit waste valuable resources, while too few can result in patients being turned away or, conversely, retained at a higher level of care than they require. Modeling patient flow is difficult, however, not only because patients vary in their length of stay but also because their movements among the different treatment levels are not always sequential. While many will begin in critical care and move toward post-acute treatment one step at a time, others will enter or leave the system from its midpoints and/or return to more acute care after spending time in less-intensive treatment. Hospital administrators must also take into consideration the relative importance their organization places on keeping patients flowing throughout the system, treating the sickest patients, producing as much revenue as possible, and not turning patients away.

Data sources used

The researchers based their model on a large U.S. hospital system on the West Coast, which could be segmented into four units of inpatient treatment: Intensive Care (ICU), Step Down (SD), Acute Care (AC), and Post-Acute Care (PAC). Throughout the period of data collection, there were 28, 76, 125, and 56 beds in the respective units. The data contained detailed information on

patient arrivals, lengths of stay at each unit, and patient routing.

Analytic techniques

In order to build the patient flow model, the researchers began with an n -stage tandem capacitated queuing system in which patients sequentially visited each stage of treatment, capturing the blocking in patient flow that would occur at each stage within such a system. The next phase of analysis involved developing a heuristic to estimate the amount of blocking in the system, and testing the researchers' heuristic against the exact solution and against other heuristics from the literature. For these analyses, the researchers used the MATLAB computing environment.

Using the tandem heuristic as a building block, the researchers went on to model the more complex hospital setting in which patients may move nonsequentially among the units (general patient routing). This algorithm became the Capacity Allocation Model, which could be used to solve for the minimum probability of blockages at each stage of service, subject to a budget constraint. The model also allowed for different weights to be placed on the various hospital units in order to capture the relative importance to administrators of admitting external patient arrivals, freeing up capacity in the most-intensive care units, and affecting revenue.

Applying the model to the data from the U.S. hospital, the researchers found that the blocking probabilities obtained from their heuristic method compared well with those obtained from the MedModel simulation package. Because the main concern of the hospital was the number of beds in the PAC unit, the researchers tested blocking probabilities with 105, 90, 70, and 56 PAC beds, which were the numbers of beds the unit had at different times during the previous decade. The researchers used the heuristic to solve the optimization problem under different weighted scenarios: minimizing blocking equally at all stages (Scenario I); minimizing blocking with priority weighted toward the most intensive units (Scenario II); minimizing the number of patients turned away from the hospital (Scenario III); minimizing per-server per-day costs, an inverse proxy for revenues (Scenario



"I spent a sabbatical with a U.S. hospital on the West Coast that had been reducing the number of beds in their post-acute care unit as a cost-saving measure. The problem came when those beds were full and the hospital was unable to move recovering patients out of the more intensive treatment units. I wanted to develop a model for optimal capacity allocation at each stage of treatment that would consider how changes in one unit would affect patient flow throughout the system.

IV); and minimizing the revenue lost from turning patients away from the hospital (Scenario V).

Lastly, the researchers conducted a sensitivity analysis to determine the robustness of the model against varying parameters, specifically those of patient arrival rate and budget.

Results

The performance of the researchers' tandem system heuristic for estimating blocking was found to be robust against the exact solution, and to represent a significant improvement in accuracy over existing heuristics from the literature. By taking into account settings with severely constrained capacity and those in which there is no buffer or waiting room in front of the various stages, the new heuristic outperformed prior approximation methods.

When applied to the hospital data, the model indicated that the initial reduction in PAC beds from 105 to 90 had hardly any effect on service, increasing blocking from 0.0% to just 0.3%. The second round of reductions to 70 beds, however, not only led to a significant increase in PAC blocking but also to a slight increase in blocking probabilities upstream in the AC and SD units. The final reduction to 56 beds, moreover, led to significant blocking problems in the PAC along with large increases in the blocking probabilities of the AC and SD units. This final cut appeared to decrease blocking probability to the ICU, but further analysis revealed that this reduction was due to external patient arrivals at the SD and AC being turned away, reducing overall patient flow throughout the hospital.

Solving for the different management objectives under consideration, the researchers observed that the optimal capacity for all five scenarios suggested an increase in PAC capacity. Additionally, all but Scenario III (minimizing number of patients turned away) also called for an increase in ICU beds. The resources necessary for these transitions came from decreases in SD and/or AC capacity.

Interestingly, the number of PAC beds was close to optimal for four out of the five scenarios before its last reduction from 70 to 56 beds, indicating that the final cut had significantly increased blocking at the hospital.

The sensitivity analysis revealed that the model was robust unless patient arrival rate increased drastically. In that scenario, as in that of a significant increase to the hospital's budget, it became advantageous to shift more resources upstream toward the more-acute care units.

Business implications

Within a hospital setting, this research provides important insight into the value post-acute care provides by reducing blocking in other units. Hospitals can use their own patient flow data to apply the Capacity Allocation Model and determine their optimal mix of beds while taking into account their particular budget constraints and specific management objectives.

The model can be applied to other settings as well, enabling an analysis of flow between different units or stations within such sectors as manufacturing, customer service, or case management. In these and other multi-stage treatment or production situations, managers and administrators can use the Capacity Allocation Model to determine the optimal number of servers to minimize blocking, maximize revenue, or avoid turning people or orders away.

This research demonstrates how analytics can be used to solve complex business problems using highly accurate simplified models. Additionally, these models can be used to determine solutions that meet a wide range of management objectives.

Kurt M. Bretthauer, H. Sebastian Heese, Hubert Pun, and Edwin Coe, "Blocking in Healthcare Operations: A New Heuristic and an Application," *Production and Operations Management*, 20 (3), pages 375-391, 2011.

Leveraging Data in the Public Sector

Tom Preston

Tom Preston leads Booz Allen's Enterprise Resource Planning (ERP) Advisory Services and oversees teams supporting Army, Navy, and Air Force ERP projects. He is a certified SAP Integration and Business Process Expert with more than 20 years' experience leading large-scale financial systems and business intelligence projects across Department of Defense, civilian, and intelligence agencies. Preston is also a graduate and the executive sponsor of the Booz Allen/Indiana University Kelley School of Business ERP and Business Analytics graduate certificate programs.

preston_thomas@bah.com



Business analytics has enormous potential to increase efficiencies and improve outcomes in government operations. The federal government alone has a budget of \$2 trillion and annual expenditures of approximately \$3 trillion; the Department of Defense, even before war supplementals, allocates some \$520 billion each year. Federal agencies are always looking for ways to save money by finding more efficient ways to operate. For example, instead of relying on the historical cost of a product or capability, the government is leveraging analytics to model and determine what something should cost and then using real-time dashboards and visualization techniques to assess their performance against industry best practices. In addition to data on these expenses, sensors and radio frequency identification (RFID) are adding date/time/location stamps to everything from shipped packages to key components of vehicles and aircraft. These many tracking systems offer rich material for business analytics, and government departments are beginning to implement such strategies as:

- Stochastic simulations and optimizations to approach what-if scenarios, from assessing financial stimulus and money supply/interest rate impacts on the U.S. economy to modeling the recruiting efficiency of the Armed Forces in meeting future military needs.
- Visualization tools that graphically demonstrate relationships between program expenditures and mission outcomes. For example, illustrating how a vaccine program is effectively protecting a city, county, or state by graphically depicting vaccine spending and flu cases with contrasting images depicting the amount of spending by location overlaid with population density.
- Text data analysis to incorporate public feedback via social media – for example, information on how citizens feel about the vaccine program as well as their ideas for improving disease prevention.



are required to develop actionable recommendations. For example, visualization software may show you that a certain area of the country has a higher-than-normal incidence of traffic delays for truck shipments. This recognition can motivate action, but the graphic won't tell you what factors are contributing to the problem, or what the widespread and long-term results would be after implementing different solutions.

With respect to mobile devices, their ubiquity creates potentially limitless opportunities for both data collection and analysis. Wherever people go, analytics can go, so long as software development is focused on preserving computational capability both on- and offline. The apps we leverage for analysis, which will include everything

from Microsoft Excel to the most advanced business intelligence tools, must provide real-time analytic capabilities and the ability to access large information stores via parallel processing and in-memory databases. These apps could assist with everything from mapping sewer lines to measuring ambient noise levels, but must be designed to function in areas without cellular service, with the capability to sync back to databases when the user goes online.

Crowd sourcing also promises a wealth of actionable information. Because it captures multiple individual viewpoints, this type of data can be a particularly informative resource. For example, I am a frequent business traveler, and I rarely go to a restaurant that does not earn positive visitor reviews. Public agencies should also find ways to leverage social media to identify and address problems, which could include everything from reporting potholes to suggesting new services. In addition to gathering feedback from people in their service area, crowd sourcing can be a means of tapping into the expertise of researchers and professionals who are eager to help but don't always have a conduit for their ideas. These tools can also contribute to constituent buy-in, so long as the organization actively responds to and makes use of this feedback.

Conclusion

Looking forward, these and other analytic tools offer promising opportunities for the public sector to address the problem of "doing more with less." In order to fully leverage these capabilities, however, decision makers will need to first of all invest in understanding business analytics enough to trust its conclusions. Secondly, the link must be established between using visualization tools to identify problem areas and following through by implementing the complementary analytic techniques and tools that can deliver solutions. Finally, departments and organizations can start to explore new technologies in the form of mobile apps and crowd sourcing, which will enable them to get more information and innovative solutions directly from the people and areas they serve.

Modeling multidimensional problems

The challenge, however, is that the overwhelming amount of information can sometimes paralyze departments instead of guiding them. The complexity of public sector issues and structures makes every problem multidimensional, with each decision carrying multiple implications for a variety of programs and stakeholders. Modeling these problems in their entirety can seem impossibly daunting.

Because of this complexity and the massive amount of data these systems generate, the temptation is to fall back on the "salami cut" strategy of making a uniform percentage budget reduction across all departments. This approach has the political advantage of appearing not to favor any one area over another. Unfortunately, it's almost always a mistake, because it addresses only the cost and not the impact of the various departments and programs. A better strategy would be for agencies to identify ways to reallocate funding and share workload across departments, deploying a shared-services model to minimize program impacts. In order to do this, however, the various stakeholders must openly share information and perform the multi-dimensional analysis required to find effective solutions.

Business analytics, when utilized to its full potential, cannot only reduce waste and conserve resources, but also strengthen an organization's ability to achieve its mission. The greatest opportunity that business analytics presents in the public sector is the potential to uncover powerful levers for positive change. Several new trends are enhancing this potential, such as visualization tools, mobile devices, and crowd sourcing.

The 21st century business analytics toolkit

The newest visualization software takes advantage of large-scale memory computing to graphically represent large volumes of data. These tools are designed to demonstrate relationships and comparisons in an intuitive manner, so they can be understood without the need for extensive training in analytics. It's important to keep in mind, however, that additional methods and techniques

Vehicle Replacement in Humanitarian Operations



Alfonso J. Pedraza-Martinez

Assistant Professor of Operations & Decision Technologies

alpedraz@indiana.edu

The International Committee of the Red Cross approached us with the question of whether their vehicle replacement policy was cost effective and gave us access to their data. My background is in humanitarian logistics, so I was glad to have the opportunity to apply empirical data to this type of problem instead of using only a theoretical model. We learned that what works in the commercial sector is not necessarily valid for the humanitarian setting, and we were able to give the Red Cross recommendations that were tailored to their particular needs.



Transportation represents an enormous overhead cost for humanitarian agencies, second only to personnel. The International Committee of the Red Cross (ICRC) relies on its fleet of 4x4 vehicles to transport people and aid across territories that are often in conflict and roads that are rarely paved. Because the reliability of the fleet is critical to the ICRC's mission, the organization's vehicle replacement policy is a topic of major importance.

With this study, the researchers examine the ICRC's Standard Replacement Policy (SRP) in the context of the organization's data on vehicle usage and replacement. They find that the current SRP is not being followed due to a misalignment of incentives between the organization's headquarters (HQ) in charge of purchasing the vehicles and the national delega-

tions (ND) that use them. Based on the usage and sales data, the researchers propose a new policy yielding cost savings.

Statement of the problem

The ICRC operates a fleet of 1,700 4x4 vehicles used in more than 80 countries. The organization upholds the manufacturer's suggested replacement policy of 5 years or 150,000 km, whichever comes first. These specifications, however, refer to commercial use under "normal" conditions, whereas ICRC vehicles are likely to be driven in rural areas of developing countries where roads are rough or nonexistent. The researchers set out to determine 1) whether the National Delegations were following the present SRP and 2) whether the SRP was optimal from a cost perspective given the actual conditions under which the vehicles were being used.

Data sources used

The researchers had access to data on ICRC fleets in Afghanistan, Ethiopia, Georgia, and Sudan. The data covered the period from 2002 to 2006 and pertained to procurement and sales, operating costs, monthly mileage, and accidents. The researchers also collected qualitative data on vehicle use and replacement practices through interviews with HQ staff and regional- and national-level logistics.

Analytic techniques

The nature of the quantitative data collection, which was conducted by ICRC staff in the field under emergency conditions, required a lengthy cleaning process during which the researchers verified each record. The next step was to create descriptive statistics for each of the four categories of quantitative data. The researchers plotted monthly costs as a function of age and monthly mileage over the course of the vehicles' lifecycles.

To determine whether ND were following the SRP, the researchers plotted the age and odometer reading of each vehicle at the time of replacement. The drivers of vehicle replacement were determined through a binary logistic model considering the independent variables of age, odometer reading, and accidents.

For their parameter estimation, the researchers created functions for preventive maintenance and miscellaneous costs and determined the drivers of salvage value using OLS regressions in Stata software. These functions were used to create an optimal replacement model based on a dynamic programming algorithm. The researchers used a C++ application to solve the model. Stochastic simulations were used to check the results' robustness.

Throughout the study, the qualitative information gained from the interviews was used to inform the model.

Results

The descriptive statistics revealed that the vehicles were considerably older than the SRP would prescribe, with median ages of 5 years in three out of four countries. Monthly mileage records showed decreased use over the lifespan of the vehicles despite rigorous standards ensuring continued reliability, which suggested excessive fleet size.

Records of the age and odometer reading of the vehicles at replacement clearly reveal that ND were not following the SRP. Only 5% of vehicles were sold according to the policy, with more than 50% of vehicles surpassing both 5 years and 150,000 km before being sold.

Combining the statistical analysis with information from the qualitative interviews, the researchers observed that the unique economic drivers of the humanitarian setting were prompting ND to keep the vehicles longer than the SRP prescribed. HQ was responsible for purchasing the vehicles and received the vehicles' salvage value at the time of sale. ND were responsible for paying depreciation during their usage of the vehicles, but only for the first five years. At that point, the policy indicated that ND

should sell the vehicles, but because they were no longer paying a monthly fee to HQ and would not receive the vehicles' salvage value, there was little incentive to do so. The evidence of fleet inflation further suggested that the cost analysis of HQ logisticians did not align with the practical considerations of ND.

Based on the binary logistic model, age and odometer reading appeared to be the drivers of the replacement decision. The 150,000 km cutoff was the approximate average for odometer reading, which was the replacement mileage suggested by the manufacturer for commercial fleets, but the corollary age at which ND were replacing vehicles was eight years.

The optimization model, which captured the increase in maintenance costs under the actual conditions in which the vehicles were being driven, the low purchasing cost at which the ICRC was able to procure the vehicles under a special agreement with the manufacturer, and the regression's result that odometer reading was the only significant driver of salvage value, revealed that by replacing the vehicles at 100,000 km instead of 150,000 km, the ICRC would save 8.7% on operating costs.

The results were robust to significant variation in purchase price, salvage value, and maintenance costs. Very large increases in maintenance costs or salvage value or very large decreases in purchase cost would indicate earlier replacement. Conversely, very large decreases in maintenance costs or salvage value or very large increases in purchase cost would argue for later replacement.

Although the 100,000 km policy would be considerably more cost effective for the ICRC as a whole, the misalignment of incentives remains between HQ and ND. The researchers note that until the incentives problem is resolved, the new policy may not be implemented at the ND level.

Business implications

With the ICRC serving as the benchmark in fleet management for other humanitarian organizations such as the World Food Programme, World Vision International, and the International Federation of Red Cross and Red Crescent Societies, this study may serve to influence decision making for a significant portion of the 80,000 4x4 vehicles presently in service in the international humanitarian sector.

Moreover, this study demonstrates that policies developed within the commercial sector do not always translate successfully to humanitarian operations. The importance of collecting and analyzing data specific to the field is revealed by the use of the optimization model. This research also demonstrates how quantitative and qualitative data can be combined to offer meaningful insights into policies and their implementation.

Alfonso J. Pedraza-Martinez and Luk N. Van Wassenhove, "Vehicle Replacement in the International Committee of the Red Cross," *Production and Operations Management*, published online 2012.

Digital Analytics: Driven by Search and Social Media

Vaibhav Gardé

Vaibhav Gardé is a Marketing Principal in the Interactive Marketing practice at FedEx. He focuses on leveraging digital analytics to create insights that improve the effectiveness of marketing programs via appropriate design of experiments, metrics, technology, and targeting. He holds an MBA (Marketing & Finance majors) from the University of Arizona, Tucson and an MS (Management Information Systems) from Arizona State University, Tempe.

vgarde@fedex.com



Digital marketing today is driving rapid adoption of measurement techniques across devices and channels. According to Forrester Research, the shift to digital marketing is placing a premium on web intelligence: the collection, measurement, and utilization of multichannel digital data to drive mutually beneficial customer relationships. A 2011 Forrester survey shows that 84% of businesses are using web analytics technology. This market is expected to evolve into three main categories: Enterprise Marketing Platforms, Online Marketing Suites, and Digital Analytics Specialists. The technology helps track the influence of online channels on sales, customer experience, loyalty, and issue resolution, among other things.

Within this decade, however, the field of web analytics has begun changing so profoundly that it has taken on an entirely new identity. The leading industry standards group – the Web Analytics Association – recently changed its name to the

Digital Analytics Association. This shift in terminology reflects a tremendous growth in mobile, social, and other emerging digital channels. The “web page,” as we know it, may soon become obsolete.

The new customer interaction paradigm

SoLoMo, meaning Social Local Mobile, is the hot new growth area for customer interactions. Whether for marketing, payments, customer service, or news consumption, companies are moving from addressing the masses to serving individuals in highly customizable formats. These technologies, and the associated explosion in devices, are fundamentally changing the way we live and work.

This protean nature of digital analytics creates many challenges for marketers that can't be solved with outdated “web analytics” methods. The new customer interaction paradigm, however, also generates new opportunities for continuous, relevant



engagement. Marketers are now closer to the holy grail of marketing: “serving the right message at the right time to the right person at the right place.”

Most consumers today – all across the world – rely heavily on online search to discover, research, and buy products and services. Search engines like Google, Bing, and Yahoo frequently modify their algorithms in an effort to serve up the “best” results. These results are increasingly based not only on the user’s search key words but also on user parameters like location, browser, search history, device, and, increasingly, social media interactions. Today’s search results are materially affected by user ratings, friends’ reviews, and other social signals.

Given the instantaneous results of online search, the cost of product comparison is lower than ever, with sites like PriceGrabber, Bizrate, and a host of others offering comparison shopping at the user’s fingertips. Online marketplaces like Amazon and eBay have made it extremely easy to search for specific products and buy at the lowest price, from the highest-rated seller, or based on other criteria like location of seller, shipping costs, or product reviews.

New career opportunities for digital business analysts

The net result of all this digitization is that incredible amounts of data are being generated each minute, 24/7/365, across many platforms, devices, and languages. All of this data gets stored for in-depth analysis and site optimization to enable companies to better meet their competitive objectives in the realms of profits, sales, donations, readership, and more.

As a result of this data explosion, job titles like “Chief Data Scientist,” “Social Media Strategist,” and “Chief Content Officer” are becoming increasingly common. To fill these roles, companies

are looking for candidates with the right mix of technical skills, ability to continually learn and adapt, and a firm grounding in business principles.

In the area of marketing analytics, it is critical to understand how to deploy paid search, optimize for organic search, and leverage both social media and content (including text, images, and video). This is not your grandfather’s marketing; it is heavily driven by data and technology. The ability to quickly make sense of the vast amounts of data, generate insights that galvanize action, and continually create “testable” scenarios will be paramount to employers going forward. Drawing insights from “Big Data” and creating accurate attribution models to optimize performance of digital channels will be critical factors to improve profitability of programs.

Conclusion

Digital analytics will continue to be a very exciting – and demanding – field for the foreseeable future. Talent in this area is going to be very hard to find (and keep). But for folks who understand data, thrive on learning new technologies, and are curious about ways to solve the ever-evolving problems of doing business in tomorrow’s digital world, the sky is the limit.

Forrester research reports: “Forrester Wave Web Analytics Q4 2011” and “Marketing Technology Adoption 2011.”

Estimating Consumer Search Costs in Online Markets



Babur De los Santos

Assistant Professor of Business Economics

babur@indiana.edu



Matthijs R. Wildenbeest

Assistant Professor of Business Economics

mwildenb@indiana.edu

Theoretical models of consumer search behavior offer two primary frameworks: a “fixed sample size” search model, in which customers cost-compare across a fixed, predetermined number of retailers, and a “sequential” search model, in which customers investigate one retailer after another until they find an acceptable price and make a purchase. Both models assume that searching has a cost to the consumer, which factors into the decision to buy or continue searching.

These models were developed in the 1960s and ’70s, however, prior to the advent of the Internet. How do today’s consumers behave online, when searching requires only a few keystrokes?

Surprisingly, the researchers find that in the context of online book shopping, consumers appear to assign a high cost to searching. Most customers did not search at all, but rather purchased from Amazon without conducting any between-store price comparisons. Among consumers who did browse, the researchers found that online consumer search behavior more closely resembles the fixed sample size model.

Statement of the problem

To what extent does online consumer search behavior follow classical search models? Does the sequential search model favored in the literature fit with online search behavior? If not, can browsing and transaction data be used to construct a new model for consumer search in the context of e-commerce?

Data sources used

Data from the ComScore Web-Behavior Panel provided detailed information on browsing and transactions for 100,000 Internet users in 2002 and 52,028 users in 2004, drawn at random by ComScore from a universe of more than 1.5 million global users. The total set of online bookstore transactions featured 7,558 observations in 2002 and 8,020 observations in 2004 across 15 online bookstores. Linking browsing history to the purchase by including search activity up to seven days prior to the transaction resulted in a sample of 18,350 search observations in 2002 and 25,556 in 2004.

Analytic techniques

The researchers analyzed each consumer’s browsing behavior and identified the bookstores visited and prices observed prior to each transaction. Then, the researchers tested whether a sequential search model fits the observed behavior. To test the model’s recall hypothesis that a customer should not return to a previously visited store unless she has sampled all known stores, they computed percentages of purchases that involved consumer recall or an exhaustive search.

To test the price dependence hypothesis that customers continue searching when encountering a relatively high price and cease searching when encountering a relatively low price, the researchers used a logit model. The dependent variable for the regression reflects the decision to visit only one retailer or continue the search to other retailers. The researchers ran several specifications to account for high/low search costs, consumers, loyalty, and consumer fixed effects.

A third sequential search model test incorporates product differentiation to investigate the hypothesis that consumers are more likely to continue searching if the price of a book is relatively high within the store’s price distribution over time for that book. The researchers used a regression of the number of stores visited by consumers on the within-store relative price.

Next, the researchers constructed a fixed sample size search model reflecting heterogeneity in consumer preferences. Starting with a utility specification reflecting both the customer’s store preference and that store’s price for the book, they construct an equation that also incorporates search costs and a stochastic noise term reflecting errors in the individual’s assessment of expected gains. To estimate the model, the researchers used a log-likelihood function, estimating parameters through a maximum simulated likelihood procedure.

Finally, the researchers constructed a multinomial logit demand model assuming consumers had sampled all stores to measure the observed difference in estimated price coefficients between the search model and a model that assumes consumers already know all prices.

“Previous research on consumer search behavior had to rely on laboratory experiments, but e-commerce provides a new opportunity to look at real-world data. We decided to investigate book shopping in our study because online bookstores represent a fairly mature marketplace. There were a number of surprises – the data suggested that consumers were searching much less than one might expect given the ease of the Internet, and even when they did search, they did not always choose the least-expensive option.”

–Matthijs R. Wildenbeest



Results

The online book market was found to be highly concentrated, with two dominant bookstores – Amazon and Barnes and Noble – capturing 83% of the market. Amazon alone accounted for 66% of book sales while Barnes and Noble was a distant second with 17%. Moreover, Amazon was visited in 74% of transactions, and in only 17% of these transactions did buyers visit any other bookstore. All together, only 25% of transactions followed visits to more than one store. These low levels of search make it difficult to model search behavior, as either model could fit: prices may be low enough for sequentially searching consumers to cease their search immediately, or search costs may be high enough that consumers set a fixed sample size of 1.

Interestingly, consumers did not always buy from the lowest-price store they visited: in 37% of transactions following visits to multiple sites, consumers purchased a higher-priced copy of the book. The average price difference between the transaction price and the lowest price encountered was \$1.99; between purchase price and lowest price available online, the average difference was \$2.60.

In testing the sequential search model, the researchers discovered that 38% of consumers recalled a previously visited store. Further, 42% of these recalling customers did not visit all bookstores of which they were known to be aware. The researchers also found that the first observed price did not affect the decision to continue searching. They found similar results when allowing for consumers to have idiosyncratic preferences for a particular retailer. These results suggest that a fixed sample search model better characterizes online consumer search behavior in this market.

Estimating the fixed sample size model with books for which the sample contained at least 20 transactions, the researchers calculated separate price coefficients for three income groups (<\$35,000, \$35,000 - \$75,000, and >\$75,000), finding that the magnitude of the price coefficient was largest for the lowest-income group. Normalized for these price coefficients, the estimated search cost was found to average \$1.35. Having a

broadband connection decreased search costs, as did having additional household members.

With respect to store fixed effects, Amazon had the highest fixed effect, with preference for shopping at Amazon appearing to account for \$3.89 more value than Barnes and Noble and \$7.53 more value than the top five “book club” sites. Price changes at Amazon, however, had a substantial impact on competitors’ market shares. As a comparison, the multinomial logit demand model assuming full price information indicated higher search costs and price elasticity.

Business implications

This study offers a number of intriguing insights for online retailers. Even in the absence of a physical store environment, customers clearly value their store preferences and will pay a premium in order to shop at their favored store. Additionally, despite the apparent ease of online price comparisons, customers are disinclined to conduct extensive searches. Most customers do not search at all, but rather visit their preferred retailer and make a purchase.

The fixed sample size model presented by the researchers could be used to model online consumer behavior and price elasticities in a number of markets, not only within retail but also for services such as health insurance in the U.S. or electricity in Europe.

As an example of the applications of business analytics, this study demonstrates the value of using empirical data to test theoretical models. Sometimes, as seen here, the data will pose a significant challenge to accepted models of consumer behavior.

Babur De los Santos, Ali Hortaçsu, and Matthijs R. Wildenbeest, “Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior,” *American Economic Review*, forthcoming.

IBA Affiliated Faculty



Frank Acito

Professor of Marketing; Max Barney Faculty Fellow; Co-Director, Institute for Business Analytics



Herman Aguinis

Professor of Organizational Behavior & Human Resources; Dean's Research Professor; Director, Institute for Global Organizational Effectiveness



Goker Aydin

Associate Professor of Operations & Decision Technologies



Hillol Bala

Assistant Professor of Information Systems



J. Doug Blocher

Chairperson and Associate Professor of Operations & Decision Technologies; Arthur M. Weimer Faculty Fellow



Kurt M. Bretthauer

Professor of Operations & Decision Technologies; Kimball Faculty Fellow



Raymond R. Burke

E.W. Kelley Chair of Business Administration; Professor of Marketing; Director, Customer Interface Laboratory



Kyle Cattani

Associate Professor of Operations Management; W.W. Grainger, Inc. Faculty Fellow



Sandeep Chandukala

Assistant Professor of Marketing; 3M Jr. Faculty Fellow



H. Sebastian (Seb) Heese

Associate Professor of Operations Management



Randy Heron

Professor of Finance; Roger & Barbara Schmenner Faculty Fellow



F. Robert Jacobs

Professor of Operations Management; Chase Faculty Fellow



Vijay Khatri

Associate Professor of Information Systems; Arthur M. Weimer Faculty Fellow; Co-Director, Institute for Business Analytics



Haizhen Lin

Assistant Professor of Business Economics



Philip T. Powell

Chairperson of Kelley Direct; Clinical Associate Professor of Business Economics and Public Policy



Anne P. Massey

Chairperson, Doctoral Programs;
Professor of Information Systems;
Dean's Research Professor



Ash Soni

Associate Dean of Information
Technology; Professor of Operations &
Decision Technologies; ArcelorMittal
Faculty Fellow



Alfonso J. Pedraza-Martinez

Assistant Professor of Operations
& Decision Technologies



Gilvan "Gil" C. Souza

Associate Professor of
Operations Management



Jeff Prince

Associate Professor of
Business Economics



Ramesh Venkataraman

Chair of the MSIS Program; Associate Pro-
fessor of Information Systems; Whirlpool
Corporation Faculty Fellow



Babur De los Santos

Assistant Professor of
Business Economics



**Munirpallam
Venkataramanan**

Associate Dean of Academic Programs;
Professor of Decision Sciences; Jack R.
Wentworth Professor



Kim Saxton

Clinical Assistant Professor
of Marketing



James M. Wahlen

Professor of Accounting;
James R. Hodge Chair



Scott B. Smart

Associate Chair of the Full-Time MBA
Program; Clinical Associate Professor
of Finance; Whirlpool Finance Faculty
Fellow and Director of Corporate
Finance Academy



Rockney G. Walters

Professor of Marketing



Daniel C. Smith

Dean of the Kelley School of Business;
Professor of Marketing; Clare W. Barker Chair



Matthijs R. Wildenbeest

Assistant Professor of
Business Economics



Noah Stoffman

Assistant Professor of Finance



Wayne L. Winston

Professor of Operations & Decision
Technologies; John & Esther Reese
Professorship



About Us

The Kelley Institute for Business Analytics uses the resources of the prestigious Kelley School of Business to produce insightful research and train professionals who can excel in this exciting new field.

What Is Business Analytics?

Simply put, it's using data to make better business decisions. And it's becoming big business.

For years, companies have collected data about their practices and consumers. Now, thanks to inexpensive computing, more companies are putting their data to work — using techniques such as predictive analytics, optimization, and simulation to make fact-based decisions that improve productivity, increase profits, and create a competitive advantage.

Kelley: Leading the Way

To make the most of business analytics, companies need innovative ideas and well-trained professionals. That's where Kelley comes in.

One of just a few business analytics programs nationwide, Kelley's IBA supports:

- An academic program that prepares students to solve business problems using analytics
- Corporate partnerships that shape Kelley's understanding of analytics and help companies tap into Kelley's talent
- Cross-disciplinary research by Kelley's expert faculty
- Seminars, conferences, a speaker series, and case competitions that bring together professionals, faculty, and students

Contact Us

<http://kelley.iu.edu/iba>

kiba@indiana.edu

Linkedin: <http://tinyurl.com/linkedin-kiba>

Youtube: <http://tinyurl.com/youtube-kiba>

