UNCW
UNIVERSITY of NORTH CAROLINA WILMINGTON
DEPARTMENT of ECONOMICS AND FINANCE
601 South College Road
Wilmington, NC 28403-5945

ACADEMY of ECONOMICS and FINANCE

PAPERS AND PROCEEDINGS

54th
ANNUAL MEETING

CHARLESTON, SC
FEBRUARY 8-11, 2017

editors
DR. ADAM JONES
DR. CLAY MOFFETT
UNC WILMINGTON

Academy of Economics and Finance

Papers and Proceedings

Volume 41, 2017

Fifty-Fourth Annual Meeting
Charleston, SC
February 8-11, 2017

Program Arranged by Dr. Shahdad Naghshpour
University of Southern Mississippi

Editors

Dr. Adam T. Jones and Dr. Clay M. Moffett
University of North Carolina Wilmington

2016-2017 AEF Leadership

President, Mary Kassis, University of West Georgia

President Elect, Shahdad Naghshpour, University of Southern Mississippi

First Vice President, Michael Toma, Armstrong State University

Second Vice President, Olivier Maisondieu Laforge, University of Nebraska - Omaha

Secretary, Kath Henebry, University of Nebraska Omaha

Treasurer, Richard M. Vogel, State University of New York at Farmingdale

Local Arrangements Chair, Jocelyn Evans, College of Charleston

Directors

Michael McLain (2017), Hampton University

Jason Beck (2017), Armstrong Atlantic State University

Mary Funck (2018), Sam Houston State University

David Jackson (2018), University of Texas - Pan American

Stephen Conroy (2019), University of San Diego

Adeba Mussa (2019), State University of New York at Farmingdale

Adam Jones (2020), University of North Carolina – Wilmington

Marc Sardy, Rollins College, FL (2020)

Recent Presidents

Edward J. Graham, University of North Carolina Wilmington

M. Douglas Berg, Sam Houston State University

Jocelyn Evans, College of Charleston

Robert T. Burrus Jr., University of North Carolina Wilmington

Steven Jones, Samford University

Robert Stretcher, Sam Houston State University

Richard McGrath, Armstrong Atlantic State University

Bill Sackley, University of North Carolina Wilmington

David Boldt, University of West Georgia

Albert DePrince, Jr., Middle Tennessee State University

Graham Mitenko, University of Nebraska - Omaha

Bichaka Fayissa, Middle Tennessee State University

Editors, AEF Papers and Proceedings

Adam T. Jones, University of North Carolina Wilmington
Clay M. Moffett, University of North Carolina Wilmington

Editor, Journal of Economics and Finance

James Payne -- University of New Orleans

Editor, Journal of Economics and Finance Education

Joshua Hall – University of West Virginia
Richard Cebula – Jacksonville University

Editors, AEF Journal

Adam T. Jones, University of North Carolina Wilmington
Clay M. Moffett, University of North Carolina Wilmington

Webmaster

Doug Berg -- Sam Houston State University

Academy of Economics and Finance

Volume 41 **Papers and Proceedings** **2017**

Table of Contents

The Fear Index and the Effectiveness of Accounting Screens in Selecting Value Stocks

Melissa K. Woodley, Creighton University
Steven T. Jones and James P. Reburn, Samford University

Abstract

In prior research, we have demonstrated an apparent deterioration in the ability of the Piotroski (2000) accounting screens to differentiate future winners from future losers among value stocks. One possible interpretation of that result -- consistent with our findings, but not proven by those findings -- was that we were witnessing a classic example of a good model gone bad.

This paper explores an alternative explanation: namely, that the effectiveness of the Piotroski model varies based on the level of the CBOE Volatility Index (the "VIX"). We find that when VIX is low, the Piotroski (2000) model performs extremely well in distinguishing those value stocks that are likely to perform well over the next year from those that are likely to perform poorly. The performance of the Piotroski model is less impressive when the VIX is middling, and less impressive still when the VIX is high.

Thus, the effectiveness of the Piotroski (2000) model appears to be greatest when the fear index is at its lowest, and vice versa.

Introduction, Literature Review, and Motivation

In our prior work, we examined the Piotroski model, which relies on financial statement variables to distinguish future winners from future losers among stocks with high book-to-market ratios (Woodley, Jones, and Reburn, 2011). Our findings confirmed Piotroski's (2000) findings during his test period of 1976-1996. However, we also found that during the ensuing years of 1997-2008, the Piotroski model essentially worked in reverse, as the stocks selected using his methodology actually underperformed other value stocks during that later period.

One potential explanation of this finding is simply that a good model has gone bad. Market efficiency advocates might well argue that stock prices now incorporate the Piotroski fundamentals into security prices, such that it is no longer possible to earn abnormal returns using his screens.

In this paper, we explore an alternative explanation for the lack of effectiveness of the Piotroski model during more recent time periods. More specifically, we ask whether the effectiveness of the Piotroski model is sensitive to investor sentiment.

If the effectiveness of the Piotroski model is indeed sensitive to investor sentiment, then the observed decline in the model's effectiveness may have resulted not from a permanent deterioration in the predictive power of the accounting screens themselves, but rather from differences in investor sentiment during the two periods tested. If so, then the model might be expected to once again produce positive results, if and when investor sentiment is more similar to what it was during the Piotroski (2000) test period than to what it was during our subsequent test period. Otherwise, it is at least possible that the Piotroski model simply does not work anymore.

One common measure of investor sentiment is the Volatility Index, or VIX. This index was first introduced by Whaley (1993) as a means to measure market volatility, and has been provided by the CBOE since 1993. It measures the near-term volatility implied by stock index option prices. Higher levels of VIX are associated with increased investor fear and changes in risk premia (Durand, Lim, and Zumwalt, 2011). The VIX is often used as a means of measuring investor fear, as seen in Copeland and Copeland (1999), Whaley (2009), Boscaljon, Filbeck, and Zhao (2011), Qadan and Cohen (2011), Mian and Sankaraguruswamy (2012), and Fernandes, Medeiros, and Scharth (2014). In fact, the VIX is colloquially known as the "investor fear gauge," i.e., as a barometer of investor anxiety (Whaley, 2009); the term "fear index" is also in rather common use.

Previous studies, including Copeland and Copeland (1999) and Boscaljon, Filbeck, and Zhao (2010), have found merit in market timing approaches of asset allocation during times of investor fear as measured by the VIX. Copeland and Copeland examined a market timing asset allocation approach based on the change in the VIX index, using data from May 1981 through September 1997. Their findings suggest that when the VIX is high, value stocks outperform growth stocks. On the other hand, their findings suggest that when VIX is low, value stocks underperform relative to growth stocks.

Subsequent to the test period examined by Copeland and Copeland (1999), the inputs to calculate the VIX were changed. The original VIX was calculated using the S&P 100 Index. On September 22, 2003, the Chicago Board of Options Exchange (CBOE) made two fundamental changes to the calculation of the VIX Index. The first was to substitute the S&P 500 Index for the S&P 100 Index. The second was to incorporate out-of-the-money options in the determination of the VIX. Whaley (2009) suggests that these changes provide better information about market volatility and make the VIX less sensitive to any one option price.

Boscaljon, Filbeck, and Zhao (2010) extended the work of Copeland and Copeland (1999), applying the new VIX measurement to data from 1990 to 2008. Although their results are to some extent consistent with the findings of Copeland and Copeland, the results are statistically significant only for holding periods longer than 30 days. For these longer holding periods, rebalancing between value and growth stocks based on changes in the VIX appears to enhance returns.

We investigate whether the level of the VIX can be used to predict the effectiveness of the Piotroski (2000) model in distinguishing future winners from future losers among value stocks.

Data and Methodology

Piotroski (2000) develops a method of scoring the financial strength of value stocks on a 0-9 scale based on nine fundamental indicators divided into three categories: profitability; leverage, liquidity, and sources of funds; and operating efficiency. For each of the nine measures, a firm receives one point if the measure is positive for firm health, zero otherwise.

Profitability is measured in terms of level, change, and earnings quality using return on assets, cash flow from operations, and accruals. Specifically, return on assets is defined as net income before extraordinary items scaled by beginning of year total assets. If ROA is greater than 0, then F_ROA equals 1; otherwise, F_ROA equal 0. Similarly, cash flow from operations is defined as operating cash flow scaled by beginning of year total assets. If CFO is greater than zero, then F_CFO equals 1; if not, F_CFO equals 0. Change in profitability is measured by the year-over-year change in ROA. If the year-over-year change in ROA is positive, F_dROA equals one; otherwise, F_dROA equals zero. Finally, earnings quality is measured by comparing ROA and CFO. If CFO is greater than ROA, then F_ACCRUAL is equals 1; otherwise, F_ACCRUAL equals zero.

Leverage, liquidity and source of funds are measured using the debt ratio, current ratio, and equity offerings. The debt ratio is defined as long-term debt scaled by the average of beginning of year total assets and end of year total assets. If the debt ratio declined compared to the previous year, then F_dLEVER equals one; else, F_dLEVER equals zero. Liquidity is measured by the current ratio, defined as total current assets scaled by total current liabilities. If the current ratio increased compared to the previous year, then F_dLIQUID equals one; if not, F_dLIQUID equals zero. Equity offerings are a negative signal for value firms because value firms by definition have a low equity price relative to book value. For a value firm, raising new funds via equity offerings is an indication that the firm is unable to generate sufficient earnings from operations to satisfy funding needs and is thus forced to issue potentially underpriced equity. Because small increases in equity could occur as the firm issues shares for stock options or other non-funding related reasons, we follow Greenwood and Hanson (2012) and classify firms as equity issuers if equity increased by at least 10% based on the change in the adjusted number of shares reported in CRSP. If equity outstanding increased by less than 10%, then EQ_OFFER equals one; otherwise, F_EQ_Offer equals zero.

Operating efficiency is measured by the change in gross margin (gross profit scaled by total revenue) and asset turnover (total revenue scaled by beginning of year total assets). If the gross margin increased compared to the previous fiscal year, F_dMARGIN equals one; if not, F_dMARGIN equals zero. Similarly, if asset turnover increased then F_dTURN equals one; otherwise, F_dTURN equals zero.

We take as our initial sample all firms in CRSP that have the necessary fiscal year-end financial statement data reported in Compustat to compute the nine F-Score variables. Following Piotroski (2000), we define value stocks as the top quintile of book-to-market firms for the fiscal year prior to portfolio formation, and estimate annual returns from the beginning of the fifth month after the fiscal year end to allow sufficient time for the financial results to be public. So, for instance, if the firm reported on a calendar year basis and the calendar year in question is 2013, its one-year performance would then be measured for the year beginning May 1, 2014 and ending April 30, 2015. Raw return is the buy and hold return for the year, while market-adjusted performance is raw return less the buy and hold return for the CRSP value-weighted index over the same period.

We collect daily VIX levels from the Federal Reserve Bank of St. Louis. VIX reporting began in January 1990; accordingly, we limit our sample to fiscal year 1989 through fiscal year 2013. This yields a final sample of 19,200 firm/fiscal year observations representing 6,431 individual firms.

Results

For purposes of brevity, in this proceedings version of the paper the tables described in this section are omitted.

Sample Descriptives

Table 1 provides summary data for our sample. Panel A compares sample financial characteristics for all 19,200 firm/fiscal year observations against the universe of 95,955 firm/fiscal year observations from which they are drawn. Compared to this broader universe, the sample firms tend to be smaller in asset size and to have a smaller market value of equity. The firms in the "value" sample also tend to be less profitable than the universe of firms.

Panel B provides a comparison of the financial characteristics of the "High Score" firms (those with an F_Score of 8 or 9) to those of the set of value firms as a whole. Likewise, Panel C compares the financial characteristics of the "Low Score" firms (those with an F_Score of 0 or 1) to those of the set of value firms as a whole. It is of course no surprise that the High Score firms will have better outcomes on the individual items that make up the F_Score, and vice versa. In addition, though, it is worth noting that compared to the overall sample of value firms, High Score firms tend to be larger, and also tend to have higher book-to-market ratios. In other words, within the set of value firms, the High Score firms as a group tend to look relatively more like the broader universe of firms.

Overall Model Results

Table 2 provides, for the overall sample of 19,200 firm/fiscal years, the results of the same basic comparisons used in the original Piotroski (2000) tests. Results are, first of all, calculated for both raw returns and market-adjusted returns. Within each of these sets of calculations, we compare the results of High Score firms to both the results of the sample as a whole ("High-All") and the results of the Low Score firms ("High-Low"). Within each of these four sets of comparisons, we perform three tests for statistically significant differences: differences in mean returns, differences in median returns, and differences in the percentage of positive returns. This produces a total of 12 statistical tests regarding differences in returns.

For both the raw returns and the market-adjusted returns, the High Score firms have a lower mean return than do the sample of value firms as a whole; however, neither of these differences are statistically significant. High Score firms have a higher mean raw return and a higher mean market-adjusted return than do Low Score firms, with these results easily attaining significance at the 5% level. And, on all 8 of the remaining tests, High Score firms produce superior results at levels well below 1%. The overall results for the model are highly impressive, with the weakest results coming from comparisons of mean returns.

A careful look at the more detailed data in the table provides an interesting explanation of why the High Score firms produce better relative results on, for instance, the median return than they do on the mean return. In all four sets of comparisons (High Score firms vs. both Low Score firms and the overall set of value firms, for both raw returns and market-adjusted returns), there is a clear pattern in which the relative performance of the High Score firms is much stronger at the lower end of the distribution than at the higher end of the distribution.

For instance, the first comparison displayed contrasts raw returns for the High Score firms against those of the overall sample. The 10^{th} percentile of High Score firms outperforms the 10^{th} percentile of the overall value firm sample by 19.14%. The 25^{th} percentile outperforms by 13.43%; the median outperforms by 7.31%; the 75^{th} percentile underperforms by 4.00%; and, the 90^{th} percentile underperforms by 22.58%. In other words, a High Score firm is relatively less likely than is the average value firm to produce either extremely poor returns or extremely good returns.

Impact of Volatility on Model Effectiveness

Having performed the preliminary tests over the entire sample period described above, we now turn to the primary issue in this paper: does the effectiveness of the Piotroski (2000) model vary based on the level of investor uncertainty, as measured by the VIX?

Table 3 displays the results when we apply the same tests described in our discussion of Table 2 to those specific periods of time when implied volatility is relatively low, and in particular when the VIX is below 20. The results are overwhelmingly favorable to the model: in all 12 comparisons, the High Score firms outperform, with p-values that are significant at the 5% level or better in every case, and significant at the 1% level or better in 11 of the 12 comparisons.

Even here, the results for the comparisons of means are not as strong as the results for the other comparisons, all 8 of which are significant at a level below 0.01%. The reason appears to be the existence of the same pattern described above regarding the sample as a whole: the relative performance of High Score forms is strongest at the bottom of the distribution, and declines as we move toward the top of the distribution. The pattern is quite clear in all four sets of comparisons, and indeed is monotonic in three of the four.

Table 4 provides an analogous set of comparisons when the level of volatility is considered "medium," based on a VIX level between 20 and 30, inclusive. The results here still appear, on balance, favorable toward the model; but, they are not nearly as strong or as straightforward as in those time periods in which the VIX level is below 20.

In the four comparisons of the mean return, the High Score firms outperform in two instances (neither of them significant), and underperform in two instances (one of them significant at the 10% level, and in fact nearly at the 5% level). The High Score firms outperform in all four comparisons of the median return (significantly so at the 10% level or better in three instances), and in all four comparisons of the percentage of positive returns (significant at the 5% level or better in two instances).

Yet again, the relatively weaker performance on the mean comparisons as opposed to the other comparisons can be explained by a noticeable, if not always monotonic, deterioration in the effectiveness of the model as we progress from the lower end of the returns distribution to the higher end of the distribution.

Finally, Table 5 displays a set of comparisons when the level of volatility is considered "high," based on a VIX level above 30. Here, the results are not at all favorable for the Piotroski (2000) model. The High Score firms underperform on 11 of the 12 comparisons. In the case of the mean return, this underperformance is significant at the 1% level in two of the four instances, and the underperformance on the median return is significant at the 10% level in one of the four. None of the comparisons of the percentage of positive returns are statistically significant at the 10% level, and this particular set of comparisons includes the only one in which the High Score firms outperform at all.

Consistent with the patterns noted in all prior comparisons, the model's effectiveness declines as we move from the lower end of the distribution to the upper end. The High Score firms outperform at the 10^{th} percentile in all four comparisons, and at the 25^{th} percentile in three of the four. At the median, as noted above, the High Score firms underperform. And, at the 75^{th} and 90^{th} percentiles, the High Score firms dramatically underperform.

Thus, the one pattern that is consistent across all three subsets of volatility levels is that the model is most effective at limiting the downside, and least effective at generating upside. In terms of the model's "overall" performance, the model is clearly quite effective during low VIX periods, arguably of some use during medium VIX periods, and distinctly counter-productive during high VIX periods.

Conclusions and Future Research Direction

Our prior work (Woodley, Jones, and Reburn, 2011) noted a sharp decline over time in the effectiveness of the Piotroski (2000) model. One potential explanation for this is that the model's use by investors has affected asset pricing in such a way that a good model has gone bad.

This paper's results provide a potential alternative to that explanation. We find that the model has performed best when the CBOE Volatility Index (VIX) has been low, and worst when VIX has been high. If that pattern were to persist over time, this could be good news for those seeking to make use of the Piotroski (2000) model. The reason for that is that those times when the VIX is low have been found to be those periods in which value stocks as a whole perform worst relative to growth stocks. Thus, periods when the VIX is low might be argued to be the very periods in which it is most necessary for those investing within the set of value stocks to distinguish future winners from future losers within that group. While the Piotroski model performs relatively poorly during periods when the VIX is high, one might argue that these are the very periods in which it is least important to distinguish future winners from future losers among value stocks; a broad "value" strategy might very well do nicely during those periods.

In addition, our findings show a consistent pattern in which the Piotroski (2000) model appears to lead to the selection of a less risky portfolio of stocks, relative to a broader portfolio of value stocks. Specifically, the stocks selected by the Piotroski model consistently outperform other value stocks at the lower end of the distribution; but, the relative performance of these stocks declines in a monotonic or near-monotonic fashion as we move from the lower end of the distribution to the upper end. In low-VIX periods, this apparent lessening of volatility is paired with an improvement in average returns, a clear win-win for investors. In high-VIX periods, on the other hand, investors pay dearly for the reduction in risk in the form of greatly reduced average returns.

We should be careful to note that this paper in no way proves that the aforementioned finding of a decline over time in the usefulness of the Piotroski model is actually a result of increased volatility in the markets during that same period of time. While this would seem to be one plausible explanation, it is possible that the converse is true. In other words, it is possible that a good model truly has gone bad, and this paper's finding of lower performance during period of higher volatility is simply a result of the fact that most of the periods of higher volatility happen to have occurred after the model lost its effectiveness.

If it turns out that at some point there is an extended period of time over which the VIX returns to levels that are moderate or low by historical standards, this would provide a ripe environment for comparing the usefulness of these two explanations. If, during such a period, the Piotroski (2000) model once again were to produce strong results, then the apparent weakening of the model that we have witnessed in recent years might well be explained as having been an artifact of the high VIX levels that have predominated during that time span. If, however, the model continues to produce relatively weak results, even in a low-to-moderate VIX environment, this would seem to provide at least some indication that a good model truly has gone bad.

References

Boscaljon, Brian, Greg Filbeck, and Xin Zhao. 2011. "Market Timing Using the VIX for Style Rotation." *Financial Services Review*, 20(1), 35-44.

Chen, Hsiu-Lang, and Werner De Bondt. September 2004. "Style Momentum Within the S&P-500 Index." *Journal of Empirical Finance* 11(4), 483-507.

Copeland, Maggie. M., and Thomas E. Copeland. March/April 1999. "Market Timing: Style and Size Rotation Using the VIX." *Financial Analysts Journal* 55(2), 73-81.

Durand, Robert B., Dominic Lim, and J. Kenton Zumwalt. Summer 2011. Fear and the Fama-French Factors. *Financial Management* 40(2), 409-426.

Fernandes, Marcelo, Marcelo C. Medeiros, and Marcel Scharth. March 2014. Modeling and Predicting the CBOE Market Volatility Index. *Journal of Banking & Finance* 40(1), 1-10.

Fu, Chengbo, Gady Jacoby, and Yan Wang. November 2015. "Investor Sentiment and Portfolio Selection. *Finance Research Letters* 15, 266-273.

Greenwood, Robin, and Samuel G. Hanson. April 2012. "Share Issuance and Factor Timing." *Journal of Finance* 67(2), 761-798.

Mian, G. Mujtabe, and Srinivasan Sankaraguruswamy. July 2012. "Investor Sentiment and Stock Market Response to Earnings News." *The Accounting Review* 87(4), 1357-1384.

Moran, Matthew. October 1, 2004. "Taking a Ride on the Volatile Side." *Journal of Indexes* 4, 11-23.

Piotroski, Joseph D. 2000. "Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers." *Journal of Accounting Research* 38(Supplement): 1-41.

Qadan, Mahmod, and Gil Cohen. January 2011. "Is It Profitable to Invest According to the VIX Fear Index?" *Journal of Modern Accounting and Auditing*, 7(1), 86-90.

Whaley, Robert E. February 2009. "Understanding the VIX." *Journal of Portfolio Management* 35(3): 98-105.

Woodley, Melissa K., Steven T. Jones, and James P. Reburn. December 2011. "Value Stocks and Accounting Screens: Has a Good Rule Gone Bad?" *Journal of Accounting and Finance* 11(4): 87-104.

Estimating the Life-Cycle of Fonts

Mary C. Pflug, Monotype Inc.
Richard A. Lewin, Rollins College

Abstract

We analyze the global font industry to determine their anticipated lifecycle based on an analysis of accounting data from Monotype Inc., a global company specializing in the production and licensing of fonts and other intellectual property to software designers. Our analysis highlights amortization as a key business parameter that may predict future acquisition trends and the balance of market power between foundries and distributors. We conclude by exploring the impact on shareholder wealth from amortization decisions made by commercial font companies. This supports our claim that strategic acquisitions may well follow hand-in-hand with the adoption of more aggressive amortization decisions.

Introduction

Digital products are now an integral part of everyday life with people interacting with a myriad of items that could be classified as a digital product. Every aspect of computer software is a commodity, protected by some form of recognized intellectual property right (Landes and Posner, 2003). Even offline, everyone interacts with digital products. Digital products that have an economic value are both expensive to produce, and legally protectable (Pavlus, 2014). Rich media functionality has now migrated onto mobile platforms, driving the need for and adoption of scalable text on phones, tablets and other similar mobile devices.

Naturally, digital products must be considered differently from physical products. Indeed, digital products face different challenges than physical products, are marketed and sold to consumers using different techniques, and must be heavily protected by legal support systems. Behind most tangible products are patented ideas. But dilemmas arise when there is no tangible form to a digital product, because a) the law is not as clear about protecting digital products (Copyright Office, 1988) and b) digital products are more easily copied, pirated, and distributed illegally to would-be consumers through the Internet. This presents unique challenges to those in the business of digital products (Martinez 1997). These challenges must be confronted however since most - if not all - businesses financially interact with digital products as an integral part of business operations and daily consumer interactions. This research focuses on the digital product industry of fonts, as an example of a digital product.

Most people are already familiar with fonts; specifically, people recognize fonts whenever using text editing software where they can browse a list of standard fonts. Beyond this, fonts are subliminal and occur in hundreds of instances in everyday life, so commonplace in fact that they are practically invisible. Fonts, also commonly known as typefaces, are a set of letters, numbers, and symbols that all have the same visual characteristics. To be precise, a typeface describes the visual characteristics that make up the look of a set of characters (letters, numbers, and symbols) in all weights (thickness of letters) and sizes (8pt, 12pt, 24pt, etc.). This was historically important when items were printed with metal type. The word "font" is derived from the Latin for pouring metal, "fundere") where there was a fundamental difference between font (the specific set of letters that make up the Italic version of Times New Roman in 12pt size, for example) versus the entire design of the Times New Roman palette that is made up of the regular, italic and bold versions in all different sizes, (Griffin, 2015). Today, it is commonplace to use either term interchangeably.

Fonts matter due to their visual and technical nature. When two products are comparable in function, the consumer's selection of a product is often subtly influenced by design and use of fonts on packaging (Childers and Jass, 2002). Consumers give credibility to movies, websites, articles, products, and more based on how professional they look, often determined by precisely which fonts are being used (Poffenberger and Franken, 1923). Fonts thus provide an important tool for product differentiation according to Henderson, *et.al.*, 2004. This was exemplified by the extraordinary lengths to which Steve Jobs went at Apple to ensure resounding consumer appeal was consistently maintained through the careful control of typefaces (Kidwell, 2015). On the technical side, new font technologies are being developed constantly to cater to rapidly advancing technology and computer needs. Rich media functionality has now migrated onto mobile platforms, driving the need for and adoption of scalable text on phones, tablets and other similar mobile devices. Digital ads, HTML 5, and CSS all drive needs for new types of web fonts and licenses.

Font Industry Analysis

While this paper will specifically focus on fonts, the results can be more generally applied to a panoply of other types of digital products. We begin with an industry analysis of the font industry, using data from major distributors and sales platforms, primarily Monotype and Adobe. This analysis determines important financial and business trends that predict future market behavior: big picture acquisition trends and the role of foundries in relation to distributors.

The font industry consists of many different players providing fonts to the market: large corporations that create new fonts and also own licenses to historical typefaces, distributors who make profit from selling the work of designers, foundries that consist of one or more designers that release fonts through distributors and/or their own retail platforms, and individual designers whose products are available for sale online through distributors. Globally the font industry generates roughly $550 million in recurring annual revenues.

Table 1: Breakdown of Font Industry Annual Revenues Calculation*

2015 Revenues	USD	Reasoning
Monotype Creative Professional Sales	88,074,000	Monotype 10K - Creative Professional Sales
Cost of Sales	30,281,000	Monotype 10K - Royalties paid to designers
Deferred Revenue	10,086,000	Monotype 10K - Deferred designer royalties
Total Monotype Distributor Revenue	128,441,000	Creative Professional Sales + Cost of Sales + Deferred Revenue
Monotype OEM Revenue (2015)	104,345,000	Monotype 10K - OEM Sales
Adobe TypeKit Revenue (2015)	10,000,000	Estimate of TypeKit revenue from Owler
Morisawa Revenue (2015)	111,280,000	2015 Morisawa revenue from company website
Non-Monotype Distributor Sales	38,419,605	See Table 2 below for the calculation
Foundry Sales	83,430,302	50% of **all** distributor sales (Monotype & others)
Custom Type Design Sales	75,000,000	300 full-time type designers - $250k in custom fonts p.a., on average
Total Font Industry Global Revenues	**$550,915,908**	

Table 2: Calculations for estimating Non-Monotype Distributor Sales*

Value of Monotype Distributor Sales	$128,441,000	Total Monotype Font Sales Value
Units in Monotype Distributors	79,645	Sum of no. of families in Monotype Distributors
Value per unit (in Monotype distributor)	$1,612.67	128.441M / 79,645
Estimated % of Sales via Non-Monotype Distributors	60%	Estimated percentage
Value per unit (in Non-Monotype distributor)	$967.67	Monotype distributor VPU x 60%
Units in non-Monotype distributors	39,706	Sum of families in Non-Monotype Distributors
Value of Non-Monotype Distributor Sales	$38,419,605	1,612.67 x 39,706 x 60%

*As of January 24, 2016

According to the Type Foundry Archive, the countries with the most type foundries are in order of importance: United States of America, Germany, United Kingdom, Netherlands, France, Switzerland, Canada, Spain, Sweden, and the Czech Republic. Most would perceive the font sales industry to be monopolized, dominated by a single corporation: Monotype. Monotype Imaging Holdings Inc. whose mission statement is to Empower Expression and Engagement is the key player in the font licensing industry.

Monotype has several primary businesses: font licensing to consumers, licensing imaging software to large companies, producing custom font designs, and licensing font management software. With revenues of $192.4 million in 2015, and as the owner of five major distributors Monotype has considerable market power in the industry. The company owns 25,000 typefaces, 20 patents, 26 patents-pending, and has recently acquired many companies, amassing intellectual property. Monotype's primary customers are content creators, consumer device manufacturers, and independent software vendors.

Table 3: Major Industry Player: Monotype

Monotype Imaging Company Inc.	
Headquarters	Woburn, Massachusetts
Incorporated	Delaware
2015 Revenues	$192.4 million
Monotype Library	25,000+ typeface designs
Monotype distributors	127,000+ typefaces
Number of employees	494 employees
Total site visits in 2015	80 million, 200+ countries
Webfonts on Fonts.com	30 million

Font Acquisitions & Intellectual Property Ownership

Monotype owns the distributors MyFonts, Fonts.com, FontShop, Linotype, FontFont, and all of their corresponding websites. Additionally, they own several hefty libraries of fonts including the Monotype® Collection, Linotype® Collection, FontFont® Library, ITC® Collection, Ascender® Originals, and Bitstream® Library. This is where Monotype's power lies in the font industry; it owns a lot of intellectual property that people desire, and has the ability to easily and cheaply license these to end users. Now that it has assembled all of the major distributors, it is in a position to influence designers' royalty rates markedly. Monotype's acquisitions are designed to increase Monotype's intellectual property and international presence. Monotype also uses acquisitions to accumulate human capital and to diversify their current offerings. With very little debt and significant cash holdings ($87.5m in 2015), it is to be anticipated that Monotype will continue to make serial acquisitions in the foreseeable future. Since their licensing costs are generally fixed, the more intellectual property they can acquire and license, the more cost-effective their operation can become. Below in table 4 is a brief overview of selected significant acquisitions made by Monotype over the last decade.

Table 4: Monotype Font Acquisitions Since 2006

Name	**Date**	**Value**
Olapic	August 9, 2016	$130 million
Swyft Media	January 30, 2015	$27 million
FontShop	July 14, 2014	$14.8 million
Mark Boulton Design Limited	April 7, 2014	$0.8 million
Bitstream, Inc. (includes MyFonts)	March 19, 2012	$50 million
Design by Front Limited	November 1, 2012	$5.1 million
Ascender	December 8, 2010	$10.2 million
Planetweb, Inc.	December 10, 2009	$1.9 million
Linotype GmbH	August 1, 2006	$59.7 million
China Type Design	July 28, 2006	$4.8 million

By applying a DuPont analysis using Monotype's financials, one can observe the financial health of the company. Monotype's 2015 return on equity is 8.53%. In 2014, their ROE was 11.03%. As a comparison, for 2015 the average ROE for the software and programming industry was 15.53%[2] with Adobe's ROE at 8.99%.[3] Other components of the DuPont model reveal significant information about the company. With a high tax burden of 34.28%, it would be worth determining ways to reduce their taxes which could include a motive for further internationalization. With a low interest burden of 2.22%, Monotype is currently borrowing very little and could easily service more debt from existing cash flows to finance future acquisitions or more research and development. The operating profit margin for 2015 is 21.51%, a small decline from 2014. In general, an intellectual property company should be operating at around a 25% profit margin to be considered financially healthy. Adobe's operating profit margin for example is 28.01%[4] and the average operating profit margin for the entire software industry is

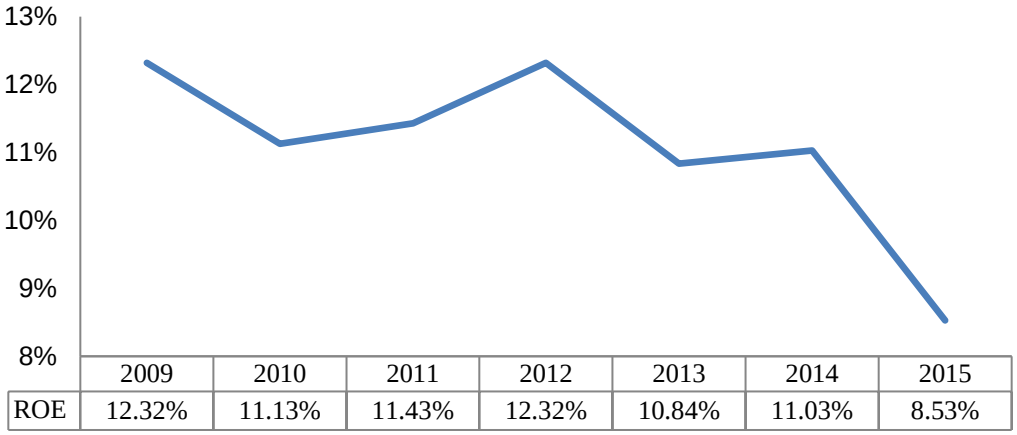
Table 5: Monotype DuPont Analysis

Return on Equity (ROE)	2014	2015
ROE = Net Profit / Equity	11.03%	8.53%
Tax Burden = Net Profit / Pretax Profit	65.85%	64.72%
Interest Burden = Pretax Profit / EBIT	97.95%	97.78%
Operating Profit Margin = EBIT / Sales	27.35%	21.51%
Asset Turnover Ratio = Sales / Assets	49.27%	49.11%
Leverage Ratio = Assets / Equity	126.89%	127.54%

29.82%[5]. Monotype's leverage ratio shows that the company has purchased slightly more of its assets by using debt rather than by issuing stock.

By looking at ROE historically (seen in Figure 1 below), one can see the effects of acquisitions on ROE. In 2010, the ROE declined sharply, the interest burden decreased, and the leverage ratio peaked in 2010, indicating that debt was taken on, interest rates being paid may have increased, and profits declined in the short term. We know that Monotype acquired Ascender in 2010 for $10.2 million, the most significant acquisition in its history at that time. The decline in ROE from 2012 to 2013 reflects the $50 million spent on Bitstream in 2012 and the decline from 2014 to 2015 reflects the acquisition of FontShop (2014). No doubt the future ROE figures for 2016 and beyond will show the effects of the major acquisition recently completed of Swyft Media (2015) and Olapic (2016).

Figure 1: Return on Equity (ROE) 2009-2015



	2009	2010	2011	2012	2013	2014	2015
ROE	12.32%	11.13%	11.43%	12.32%	10.84%	11.03%	8.53%

This declining trend shows the effect of many acquisitions in recent years. It takes time for companies to put acquired technology and human capital from acquisitions into productive use in a way that will be positively reflected in the ROE.

Table 6 below illustrates a breakdown of Monotype's competition by category. It is notable that there are three forms of competition that span all five of Monotype's primary business areas.

Financial Life-Cycle of Fonts

When Monotype acquires intellectual property, most often in the form of font collections and libraries, they must amortize them over time. Amortization is the spreading out of capital expenses for intangible assets over a specific period of time (usually over the asset's anticipated useful life) for accounting and tax purposes6. Amortization is just like the depreciation charge incurred on tangible assets, but applied to intangible assets instead (Dahmash et. al. 2009).

In most businesses amortization represents a de minimis impact on profitability, as most balance sheets house mainly tangible items. As they are not investments in tangible assets, most expenditures on intangible assets are not recognized as

Table 6: Monotype's Competition, by category

Font Licensing	Text rendering software providers	Font management software vendors	Custom Font Design	Printer Driver Software
• Adobe • Morisawa • Open Source Font providers (like Google Fonts) • Boutique font design foundries • Other font distributors like Creative Market	• Open source organizations like FreeType	• Extensis • Insider Software	• Boutique font design foundries • Independent professionals	• Software Imaging • Zoran • In-House Departments
All categories of competition:	• Adobe • In-house resources of larger customers • Illegally downloaded intellectual property			

corporate investments under US GAAP or IFRS accounting principles, associated R&D costs are expensed as incurred (see Chalmers, et al., 2008, for a discussion of IFRS versus GAAP treatment of intangible assets). In consequence, according to Ahmedb et. al. (2006) corporate profits may be understated in corporations that are investing a higher proportion of their cashflow in intangible assets. According to Nakamura, (2003) in aggregate this has caused price/earnings ratios in economies to rise over time, as the market value of stocks has risen relative to the tangible net worth or net book value measures.

The prominence of amortization in the accounts of Monotype represents a significant portion of the total annual charge, thus impacting reported profit levels concomitantly. Matolcsy and Wyatt (2006) document the negative association between accuracy and dispersion of analysts' earnings forecasts and the aggregate reported intangibles for listed companies. Each year, Monotype states the amount they are amortizing for that year in their report and accounts and then makes predictions for amortization expenses out into the future, as detailed in their respective accompanying notes. Cost of revenue also includes amortization of acquired technology, which are typically amortized over 8 to 15 years. For purposes of amortizing acquired technology, Monotype estimates the remaining useful life of the technology based upon various considerations, including knowledge of the technology and the way in which customers use it. The firm then applies the straight-line method to amortize acquired technology as there is no reliable evidence to expect any other pattern of amortization than a straight-line, which is considered to best reflect the expected economic usage.

We now examine how actual amortization charges have varied versus the prospective corporate forecasts detailed over the last 8 years of accounts. In Table 7, the dark turquoise cells represent the amount of amortization taken in that year. The green cells show the amortization expense leftover that needs to be spread out over the residual period of the intangible asset's useful life, or the 'thereafter'.

The cells between the dark turquoise and green cells are the actual amortization and predicted amortization for the five years after the date of the respective annual report. The light turquoise cells show where the actual values paid a year later have been substituted for the former past predictions. By dividing the residual by the average predicted yearly amortization expenses (green cell / average of the light blue cells), the result is a figure that shows the number of years anticipated beyond the current year for intangible assets remaining in place. By adding 6 to this number (to account for the 6 years that are not included in "thereafter") the result is the total number of years over which the assets are assumed to be amortized, or the number of years that the company expects the assets to "live," producing profit and revenue.

It is notable that the assets can still produce revenue far beyond their amortized lifespans. We note that the total amortization expense is declining over time concomitantly with the average font lifecycle shortening. This is somewhat counterintuitive given the pace of font industry concentration, as the value of an influential firm such as Monotype might be expected to reflect an increasing number of useful years estimated for its established intellectual property, in part reflecting reduced competition.

This also raises important issues for investors looking to estimate sustainable pass-through earnings out into the future, as this apparent divergence suggests they need to be especially careful in determining the appropriateness of current share prices. Table 7 shows the actual amortization versus the previously predicted amortization. The amortization figures in each year's annual report can be compared to the previous years' report to determine precisely how predicted amortization varied from actual amortization taken, as reported the following year. It is notable that Monotype consistently over-predicts the useful life of their assets. The convergence of the two lines in the graph is where the data for the past and future predictions is insufficiently distinct to be used as a comparison.

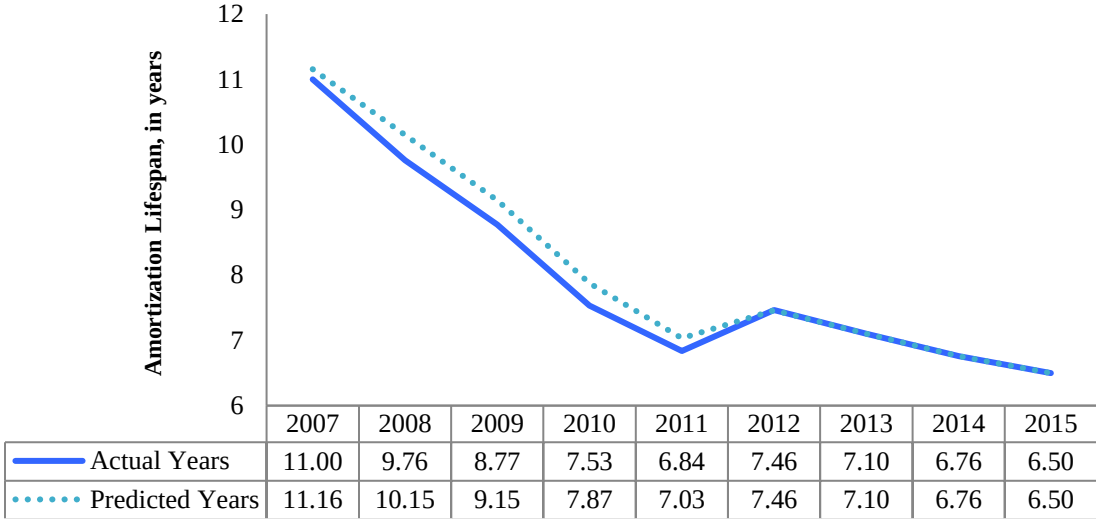
Table 7: Detailed Calculation of Amortization of Goodwill

2007	3,376								
2008	3,392	3,392							
2009	3,383	3,383	3,383						
2010	3,488	3,488	3,488	3,488					
2011	3,169	3,169	3,169	3,169	3,169				
2012	4,051	4,051	4,051	4,051	4,051	4,051			
2013	17,482	4,560	4,560	4,560	4,560	4,560	4,560		
2014		14,030	4,574	4,574	4,574	4,574	4,574	4,574	
2015			11,006	3,186	3,181	4,573	4,584	5,161	4,448
2016				5,991	2,780	4,172	4,183	4,760	4,687
2017					3,200	2,165	2,176	2,752	2,680
2018						5,866	2,072	2,649	2,576
2019							3,873	2,606	2,533
2020								2,725	1,686
2021									926
Average Prediction	3,497	3,730	3,968	3,908	3,829	4,009	3,518	3,586	2,791
Total Amortization	34,965	32,681	30,848	25,531	22,346	25,910	21,462	20,653	15,088
Future Depreciation	5.03	3.82	2.84	1.56	0.86	1.46	1.05	0.73	0.30
Total No. of years	11.03	9.82	8.84	7.56	6.86	7.46	7.05	6.73	6.30

To be accurate in comparing past and present, at least 5 years of data are required (since the predictions in the annual reports are being made for 5 years, excluding the current year's amortization charge). The average lifespan of a Monotype asset (most likely a font) is therefore calculated as 7.84 years calculated as a geometric mean. The average lifespan, using Monotype's predicted values, is 7.99 years. This means that in about 8 years, these assets will be entirely amortized and any income generated from these become surplus.

If a firm acquires software by buying another business, or its assets however, it must be amortized over 15 years using the straight-line depreciation method. Except for trademarks which are also amortized over 15 years, the IRS has not established any set time periods for the useful lives of intangible assets. It is thus left up to the taxpayer to determine the useful life of the asset which provides scope for creativity in accounting decisions.

Figure 2: Amortization of Goodwill (2007-2015)



	2007	2008	2009	2010	2011	2012	2013	2014	2015
Actual Years	11.00	9.76	8.77	7.53	6.84	7.46	7.10	6.76	6.50
Predicted Years	11.16	10.15	9.15	7.87	7.03	7.46	7.10	6.76	6.50

Conclusions

Based on the above analysis, which uses 8 years of accounting data, it appears that Monotype may be strengthening its intellectual property franchise by using amortization strategically on an accelerated basis. The company appears to be writing down the value of acquired assets within their annual accounts at a rate that is increasing over time. This suggests that the value of its intellectual property may decline over time which is at odds with the improvement in competitive standing versus the remaining industry. Most likely this is a tax deferral strategy being used in preparation for continued aggressive plans to acquire more companies in an industry that they already dominate, and where they are establishing powerful economic barriers to entry. This portends that they will be replacing the assets acquired previously that have already been fully amortized on an accelerated basis, as demonstrated in our preceding analysis.

Furthermore, we conclude that Monotype, like many other technology firms, consistently over-predicts the number of years it will take to amortize assets acquired, i.e. the lifespan in years of their intangible assets appears to be being adjusted to determine, or possibly smooth, measures of profitability. The motivation behind doing so is probably to defer taxes thereby releasing additional working capital for acquisitions. Clearly shareholder value is directly impacted by such decisions and the accounting treatments appear insufficient for shareholders to ascertain the true economic impacts. These are, at the very least, obfuscated by the selected accounting treatment of intangible assets.

Notes

[1] Views and opinions presented in this paper are solely those of the author(s) and do not necessarily represent those of Monotype Inc.

[2] CSI Market, 'Software & Programming Industry Management Effectiveness Information & Trends', CSIMarket.com, 2015. http://csimarket.com/Industry/industry_ManagementEffectiveness.php?ind=1011.

[3]Adobe Systems Inc. 'Adobe 2015 10K Annual Report', News release, November 27, 2015, Adobe Investor Relations. http://www.adobe.com/investor-relations/financial-documents.html.

[4]Adobe Systems Inc. "Adobe 2015 10Q Quarterly Report." News release, November 27, 2015. Adobe Investor Relations. http://www.adobe.com/investor-relations/financial-documents.html.

[5] Stern, NYU, 'Margins by Sector (US)' Operating and Net Margins, January 2016. http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/margin.html.

[6] http://www.investopedia.com/terms/a/amortization.asp.

References

Ahmedb, Kamran and John Goodwina. 2006. 'Longitudinal value relevance of earnings and intangible assets: Evidence from Australian firms', *Journal of International Accounting, Auditing & Taxation*, Vol.15, Issue 1, pp.72–91.

Chalmers, Keryn, Greg Clinch and Jayne M. Godfrey. 2008. 'Adoption of International Financial Reporting Standards: Impact on the Value Relevance of Intangible Assets', *Australian Accounting Review*, Vol.18, No.3, pp.237–247.

Childers, Terry L., and Jeffrey Jass. 2002. 'All Dressed Up with Something to Say: Effects of Typeface Semantic Associations on Brand Perceptions and Consumer Memory', *Journal of Consumer Psychology*, Vol.12, pp.93–106.

Copyright Office. 1988. 'Policy Decision on Copyrightability of Digitized Typefaces' Federal Register 53 (189): 38110–38113, *Office of the Federal Register*, ISSN 0097-6326.

Dahmash, Firas N., Robert B. Durand and John Watson. 2009. 'The value relevance and reliability of reported goodwill and identifiable intangible assets', *British Accounting Review*, Volume 41, Issue 2, pp.120–137.

Eisfeldt, Andrea and Dimitris Papanikolaou. 2014. "The Value and Ownership of Intangible Capital", *American Economic Review*, Papers & Proceedings, Vol 104, No.5, pp.189-194.

Griffin, P. 2015. 'Money Talks: A Brief History of Type Markets. "TypeCon 2015. Lecture conducted from Society of Typographic Aficionados, Denver, CO.

Henderson, Pamela W., Joan L. Giese, and Joseph A. Cote. 2004. "Impression Management Using Typeface Design." *Journal of Marketing*, Vol.68, pp.60–83.

Kidwell, Jim. 2015. Typographic Trends. Report http://blog.extensis.com/typography/typography-trends-get-the-full-report.php. Portland: Extensis.

Landes, William, and Richard Posner. 2003. The Economic Structure of Intellectual Property Law. Cambridge, Mass.: Harvard University Press.

Martinez, Frank J. 1997. "Protecting Type Font Design." *AIGA Journal of Graphic Design*, Vol.15, No.3, Accessed April 5, 2016. http://www.martinezgroup.com/articles/1997aigaprotectfont.html.

Matolcsy, Zoltan P. and Anne Wyatt. 2006. 'Capitalized intangibles and financial analysts', **Accounting and Finance**, Vol.46, No.3, pp.457-479.

Nakamura, Leonard. 2003. 'Intangible Investment and the New Economy' in Intangible assets, Values Measures and Risks, edited by John R. M. Hand & Baruch Lev, Oxford University Press.

Pavlus, John. 2014. 'Fonts Made for Writing, Not Reading', *Bloomberg Businessweek*, Issue 4395, pp.53-54.

Poffenberger, A. T., and R. B. Franken. 1923. "A Study of the Appropriateness of Type Faces", *Journal of Applied Psychology*, Vol.7, pp.312–323.

Short-Termism and the CAPM

Richard Lewin & Marc J. Sardy, Rollins College
& Stephen E. Satchell, University of Sydney

Abstract

Short-Termism is a concept used pejoratively to describe one of many alleged character failings of an investor. However, it is seldom defined in a formal context in which its presence/absence could be deduced from a statistical model. We provide both a definition based on duration and a model to estimate and test. We consider an extension of the traditional capital asset pricing model (CAPM) to incorporate duration as a general measure of the price volatility of bonds or equities. We outline our theory deriving an extended CAPM framework to account for duration explicitly, alongside implications for inclusion in asset management.

Introduction

This paper considers an extension of the Sharp (1964), Linter (1965) & Mossin (1966) capital asset pricing model (hereinafter referred to as CAPM) to incorporate duration, as a general measure of the price sensitivity of a bond or equity to its discount factor. Duration is a widely-used measure to quantify and control interest rate risk exposure in asset management. Given the extremely low interest rate environment globally, investor portfolios are more vulnerable than ever to minor interest rate perturbations, which may be mitigated by a duration approach. For pension schemes with long-dated liabilities, it is natural that they would hold high duration assets. Whilst this is dominant in fixed income markets, more recent extensions of duration have been derived as an interest rate risk measure appropriate to equity securities, see for example Leibowitz *et al.* (2010) and Schroder and Esterer (2016). On the basis of the equity duration literature, this paper outlines the theory behind deriving an extended CAPM framework, to account for duration explicitly as an additional factor parameter.

A priori we first consider several data issues relating to an empirical verification of any such model. As a starting point, we could look to take some 5 years or 60 months (in either weeks or days) of data, taking equity and bond indices weighted for the market, for example by 0.6 and 0.4, to reflect standard institutional weightings. Bond data characteristics could then be assumed to mimic the behaviour of bullet bonds. The measure of duration for pure discount bonds, in the absence of coupon payments, is equivalent to their term to maturity. We could subsequently specify our set of assets, where $i = 1 \ldots N$, and further assume from the above condition that our N assets have a well-defined non-negative duration d_i, (again $i = 1....N$), mimicking the behaviour of zero coupon bonds. We can then estimate $\mu_i - r_f$, β_{im} and d_i from available data, where μ_i, r_f, β_{im} and d_i are defined as the mean return on asset i, the risk free rate, the covariance of asset i with the market m, divided by the variance of the market, and the duration of asset i respectively.

The duration measures we would derive still of course imply some relationship between duration and the established view of market participants, as alluded to in the discussion found in Lewin *et al.* (2007). Our contribution here is to consider the reformulation of these measures to directly test such hypotheses. Thus, we aim to rationalise some of the clientele effects observed in financial markets, as well as conducting research into the underlying reasons behind different institutional requirements for specific target-portfolio durations. The particular desired duration for any individually specified portfolio is likely to be either longer or shorter than the duration, ordinarily implied when considering a traditional market equilibrium position. The advantages of thinking that an investor's expected utility is defined over duration is that we can produce a meaningful definition of short-termism.

A further important motivation is provided by Merton (1973) who finds in his ICAPM model that "expected returns on risky assets may differ from the riskless rate even when they have no systematic or market risk", and discusses (pg.885, section 9) if this is some form of long-term bond; whilst he does not pinpoint duration explicitly, the introduction of a 'duration' asset seems entirely compatible with his arguments.

Methodology

For this to be consistent with market equilibrium, we first invoke a separation theorem argument to provide for a duration-based approach to asset management. In examining such evidence empirically, we might necessarily discover the institutional clientele effects required for holding differing duration-based portfolio positions in an equilibrium framework. As is usual for such CAPM type models, we could then assume a representative agent and define an indirect expected utility function as V,

which is dependent upon expected excess returns, volatility and duration, as defined by $\mu_m - r_f$, σ_m and d_m (where the subscript *m* denotes the market portfolio).

$$\text{Let derived utility } (V) = V(\mu_m - r_f, \sigma_m, d_m), \text{ where } \bar{\mu}_m = \mu_m - r_f = \sum_{j=1}^{N} x_j(\mu_j - r_f), \tag{1}$$

For *N* risky assets, the riskless asset having a zero duration and corresponding to cash ,

$$\text{Also, } \sigma_m^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{ij}, \text{ and duration } d_m = \sum_{j=1}^{N} x_j d_j \tag{2}$$

Here we thus have a representative agent with three-fund money separation.

The above allows us to provide a formal definition of short-termism.

Definition of Short-Termism

We define as follows:

A short-termist market (or individual) as one whereby $\frac{dV}{dd_m}<0$;

A long-termist market (or individual) as one where $\frac{dV}{dd_m} > 0$.

We could also have term neutral investors where $\frac{dV}{dd_m} = 0$.

We note that for an individual *p*, we would have W $(\mu_p - r_f, \sigma_p, d_p)$. This represents the expected utility of investor *p*, expressed in terms of moments of their optimal portfolio *p* and the same definitions apply. We might expect some pension fund managers to be long-termist and day traders to be short-termist. The overall position of the market would depend on aggregating different types of investors.

We now consider the implications of this model.

$$\text{At the optimum } \mathrm{d}V = 0, \text{ correspondingly } \frac{\mathrm{d}V}{\mathrm{d}x_i} = 0 \tag{3}$$

$$\text{So, } \frac{\mathrm{d}V}{\mathrm{d}x_i} = \frac{\mathrm{d}V}{\mathrm{d}\bar{\mu}_m}.(\mu_i - r_f) + \frac{\mathrm{d}V}{\mathrm{d}\sigma_m}\left(\frac{\sum X_i \sigma_{ij}}{\sigma_m}\right) + \frac{\mathrm{d}V}{\mathrm{d}\,d_m}(d_i) = 0 \tag{4}$$

$$\text{Therefore rearranging, } \mu_i - r_f = \left(\frac{\frac{-\mathrm{d}V}{\mathrm{d}\sigma_m}}{\frac{\mathrm{d}V}{\mathrm{d}\bar{\mu}_m}}\right).\left(\frac{\sigma_{im}}{\sigma_m}\right) + \left(\frac{\frac{-\mathrm{d}V}{\mathrm{d}d_m}}{\frac{\mathrm{d}V}{\mathrm{d}\bar{\mu}_m}}\right)(d_i) \tag{5}$$

$$\text{Finally, we arrive at, } \mu_i - r_f = \left(\frac{\mathrm{d}\bar{\mu}_m}{\mathrm{d}\sigma_m}\sigma_m\right)\beta_{im} + \left(\frac{\mathrm{d}\bar{\mu}_m}{\mathrm{d}d_m}d_m\right).\left(\frac{d_i}{d_m}\right), \tag{6}$$

$$\text{Since at the optimum, } \left[\frac{\frac{-\mathrm{d}V}{\mathrm{d}\,a}}{\frac{\mathrm{d}V}{\mathrm{d}\,b}} = \frac{\mathrm{d}\,b}{\mathrm{d}\,a}\right] \text{ for any } a, b. \tag{7}$$

We present the above result in equation (6) as Theorem 1.

Theorem 1

In a representative agent equilibrium, given a one period world with returns, duration and volatility defined for each asset, equilibrium expected excess returns are linear in covariance β_i, and in duration d_i, where α_1 and α_2 are defined in the proof accordingly. Given equation (6) above, this result follows upon substitution and rearrangement.

$$\text{i.e. } \mu_i - r_f = \alpha_1 \beta_{im} + \alpha_2 d_i \text{, let } \alpha_1 = \frac{\mathrm{d}\bar{\mu}_m}{\mathrm{d}\sigma_m}.\sigma_m \text{ and } \alpha_2 = \frac{\mathrm{d}\bar{\mu}_m}{\mathrm{d}d_m} d_m . \frac{1}{d_m} \tag{8}$$

The testable implications of this model concern the signs of α_1 and α_2. From the above we expect $\alpha_1 > 0$, since $\mathrm{d}\mu_m / \mathrm{d}\sigma_m > 0$. However, the sign of α_2 is ambiguous. One consideration here is yield curve related, where we might argue that $\alpha_2 > 0$, if the yield curve is upward sloping, but $\alpha_2 < 0$ if it is inverted. So in fact, investor preferences for any duration component within their portfolios may be related to the envisaged shape of the yield curve, as well as their respective investment horizon. Of course, should our representative agent be a short-term investor, we might expect $\mathrm{d}\mu_m / \mathrm{d}d_m < 0$, but if he were a longer-term investor, we might equally expect $\mathrm{d}\mu_m / \mathrm{d}d_m > 0$. In our example, the representative agent's utility specification will depend upon the expected excess returns, the implied volatility and the duration. The typical UK/US investor could be thought of as a large pension fund whose liabilities are relatively long-term. Thus, we might envisage a preference for assets characterised by typically longer durations.

There is, however, another alternative and more populist literature, exemplified by Pickens (1986) that refutes the preceding argument, advocating that investors are in fact particularly short-termist in their approach. For a rigorous academic version that adds considerably to this area of debate, as well as providing a useful UK interpretation, see Miles (1993) and (1995) and the related notes appearing in Damant and Satchell (1995). For the details of a parliamentary discussion, which is certainly less academic, see the available references in Hansard[1]. For a more polemical treatment, the reader is referred to the arguments presented in Hutton's books (1996) and (1997). His core analysis that the increasingly market-oriented British model of capitalism was in trouble and needs to be reformed, to reduce the power of short-termist shareholders and strengthen that of other "stakeholders" in companies, including workers and the state. Irrespective of the manner in which our reader may elect to interpret the short-termist claims, as considered in any of these sources, our model permits these issues to be addressed objectively, by associating them with a value for α_2.

In order to extend the equilibrium conditions for our representative agent however, we would have to invoke a three-fund separation theorem. Our theory maintains that, cross-sectionally the asset risk premium will be related to beta as well as the duration of the asset relative to the duration of the market. We can therefore estimate β_{im} from historical market data. However, the duration of the market d_m presents more of a problem. In our three-funds CAPM framework, we would require to know three rates of return in order to determine a unique structure. We would need to know the return on the market portfolio, the rate of return on our riskless asset and the rate of return on any portfolio with a zero β, conditional upon having an associated non-zero duration. We delay a full treatment to Theorem 2.

If Theorem 1 applies to the entire market, taking $i = m$, $\beta_{mm} = 1$ and $d_i = d_m$, will allow for the following expression:

$$\mu_m - r_f = \alpha_1 + \alpha_2 d_m \tag{9}$$

$$\text{Thus } (\mu_m - r_f) = \alpha_1 \beta_{im} + ((\mu_m - r_f) - \alpha_1)\frac{d_i}{d_m} \tag{10}$$

$$= \alpha_1 \left(\beta_{im} - \frac{di}{dm} \right) + (\mu_m - r_f)\frac{d_i}{d_m} \text{ ,} \tag{11}$$

Therefore, in equation (11), we have arrived at an alternative representation.

Corollary 1

The empirical consequences of equation (11) *(for i = 1…. m.)* are that there should be some cross-sectional relationship between risk premia, the betas and the relative duration, d_i / d_m. Indeed, the sum of the two coefficients should equal the excess returns and will have testable consequences, suitable for empirical investigation.

An Extended CAPM with a Duration Bond Factor

Extensions to the above approach are possible. Suppose we have a fixed income instrument of duration d_m with zero exposure to the market and expected rate of return relative to r_f equal to$\mu_d - r_f$. Such an asset is broadly consistent with the Merton approach discussed at the end of section 1.

$$\text{Thus, } \mu_d = -r_f \frac{-(\frac{dV}{dd_m} d_m)}{\frac{dV}{d\mu_m}} \tag{12}$$

$$\text{It then follows from } \mu_i - r_f = \frac{-(\frac{dV\sigma_m}{d\sigma_m}\beta_{im} + \frac{dV}{dd_m} d_i)}{\frac{dV}{d\mu_m}} \tag{13}$$

$$\text{That } \mu_i - r_f = \frac{-(\frac{dV\sigma_m}{d\sigma_m}\beta_{im})}{\frac{dV}{d\mu_m}} + \frac{(\mu_d - r_f)d_i}{d_m} \mu_i - r_f = \beta_{im}(\frac{(\mu_d - r_f)}{d_m} + (\mu_m - r_f)) + \frac{(\mu_d - r_f)d_i}{d_m} \tag{14}$$

$$\mu_i - r_f = \beta_{im}(\mu_m - r_f) + \frac{(\mu_d - r_f)(d_i + \beta_{im})}{d_m} \tag{15}$$

Theorem 2

Assume the existence of a fixed income instrument of duration d_m with zero exposure to the market, an expected rate of return relative to r_f equal to $\mu_d - r_f$, and that the assumptions of Theorem 1 hold, therefore:

$$\mu_i - r_f = \beta_{im}(\mu_m - r_f) + \frac{(d_i + \beta_{im})}{d_m} (\mu_d - r_f) \tag{16}$$

Theorem 2 gives a stock risk premium consistent with a multi-factor model in which the factors are the market and the pure duration product whose duration is equal to the market portfolio. The fact that the CAPM might have a missing factor gives us the possibility of explaining the low beta anomaly. We recall that in a regression context where y is the dependent variable and x and z are the true regressors in the linear model:

$$y = \alpha + x\beta_1 + z\beta_2 + v \tag{17}$$

a regression of y on x only leads to the following outcome:

$$\text{plim } (\widehat{\beta_1}) = \beta_1 + \frac{cov(,x,z)}{var(x)} \beta_2 \tag{18}$$

Comparing (17) and (15), we might expect stocks to have underestimated betas if $cov(,x,z) < 0$ when β_2>0, which it should be in this context since $\frac{(d_i+\beta_{im})}{d_m}$>0.

This then reduces to the question of when might market duration and market returns be negatively correlated? Next, we outline plausible circumstances when this might occur. Suppose that the market portfolio undergoes a compositional change whereby the proportional capitalisation of small stocks relative to large stocks increases. If the size factor is working, market returns should go up. If small firms have smaller duration than large firms then we might expect a negative correlation. Issues that might matter to find a negative relationship between size and duration could be that small firms may have pay-outs out in the future relative to large firms(positive), that small firms are more likely to default(negative), that growth rates might be higher for small firms(positive).

We start by adapting the framework of Shaffer (2007), who links duration to firm default. Using a discounted cash flow approach with g being the growth rate of dividends, D the initial payment(which is also our size proxy, r the risk-neutral discount rate and p the probability of default, and ignoring some considerations not relevant to our argument, the Shaffer approach shows that the risk-neutral value of the firm F is given by $F = \frac{D(1+r)}{(r+p-g+pg)}$ and that the duration, d, is $d = \frac{(1-p)(1+g)}{(r+p-g+pg)}$ where d is defined as the absolute

value of the elasticity of F with respect to (1+r).As such, d is independent of D and hence is constant for all D. We note that there are other definitions of duration in the literature.

We now make further assumptions. Suppose that p = p (D) and $\frac{dp}{dD} < 0$. This captures the stylised fact that small firms are more likely to go bankrupt, see Mata and Portugal (1994) and included references. Likewise, we shall assume that g = g (D) and $\frac{dg}{dD} < 0$, which says that smaller firms have faster growing dividends.

Theorem 3

(i) Suppose that p = p (D) and $\frac{dp}{dD} < 0$ then $\frac{dLn(d)}{dD} < 0 \; if \; 2p + r < 1$.

(ii) Assume that g = g (D) and $\frac{dg}{dD} < 0$ then $\frac{dLn(d)}{dD} > 0 \; if \; 2p + r < 1$.

$$\text{Proof: } \ln(d) = \ln(1-p) + \ln(1+g) - \ln(r+p-g+pg) \tag{19}$$

Under 4(i), $\frac{dLn(d)}{dD} = \frac{-\frac{dp}{dD}}{(1-p)} + \frac{\frac{dp}{dD}(1-g)}{(r+p-g+pg)} = \frac{\frac{dp}{dD}(1-r-2p)}{(1-p)(r+p-g+pg)}$

Under 4(ii), $\frac{dLn(d)}{dD} = \frac{\frac{dg}{dD}}{(1+g)} - \frac{\frac{dg}{dD}(1-p)}{(r+p-g+pg)} = \frac{-\frac{dg}{dD}(1-r-2p)}{(1+g)(r+p-g+pg)}$

With annual data, *2p + r* is very likely to be less than 1; and evidence already cited supports $\frac{dp}{dD} < 0$.

However, whether a constantly growing firm grows faster starting from a smaller size seems unlikely. Empirical evidence suggesting such a phenomenon may be obtained by observing firms of different maturities over differing horizons. Thus, if we accept Theorem 3(i) as plausible, and are less inclined to accept Theorem 3(ii) and prepared to believe that higher duration stocks have higher market duration betas, then the analysis above would give reasons why small stocks might have too low market betas in a CAPM world where we ignore the duration factor.

We note, however, that some authors find different conclusions. To quote Schroder and Esterer (2016), "The small size effect is also present in the data, as small stocks have on average higher expected returns than large stocks, see panel A. Furthermore, the negative association between firm size and earnings growth suggests that small firms are expected to grow faster. There is a positive relation between equity duration and firm size, but a negative relation between cash flow duration and firm size. This is intuitive: small firms carry high risk premia, such that their cash flows are discounted at a high rate, leading to short equity durations. At the same time high expected growth rates of small firms result in long cash flow durations."

Close inspection of their definitions point to the fact that cash flow duration is just conventional equity duration with an exogenous discount rate, so that we recover the result we want with that additional assumption.

Theorem 4

Here we now consider what a minimum variance frontier might look like for a duration-constrained portfolio. This analysis is reminiscent of early work by Jean (1971) and (1973), as well as Ingersoll (1975). As before, we have *N* assets with rates of return *r*, where $r \sim (\mu, \Sigma)$ and the duration of the assets is denoted by *d*. Let the constrained expected returns be Π and constrained duration d_0. Our frontier is thus the locus of points given by:

$$(\Pi, d_0, \sigma^2), \text{ where } \sigma^2 = x' \sum x \tag{20}$$

Hence we solve the following optimization problem (where *e* is a vector of ones):

$$\underset{x}{Min} \; \left((x' \sum x) - \lambda_1 (x'e - 1) - \lambda_2 (x'\mu - \Pi) - \lambda_3 (x'd - d_0) \right). \tag{21}$$

Differentiating, we see that as before, we would get three-fund separation.

Alternatively, this can be written as:

$$\underset{x}{Min}\left(\frac{1}{2}x'\sum x-\lambda'(\mathrm{E}'x-\mathrm{E}_0)\right), \text{ where } \mathrm{E}(\text{an } n\times 3 \text{ matrix})=[e,\mu,d] \text{ and } \mathrm{E}_0=(1,\Pi,d_0)'. \tag{22}$$

The first order conditions imply that for $\hat{x}$, the solution becomes $\Sigma\hat{x}=\mathrm{E}\lambda$ and $\mathrm{E}'\hat{x}=\mathrm{E}_0$.

$$\text{Therefore } \hat{x}=\Sigma^{-1}\mathrm{E}\hat{\lambda} \text{ and } \mathrm{E}'\hat{x}=\mathrm{E}_0=(\mathrm{E}'\Sigma^{-1}\mathrm{E})+\hat{\lambda}. \tag{23}$$

$$\text{So } \hat{\lambda}=(\mathrm{E}'\Sigma^{-1}\mathrm{E})^{-1}\mathrm{E}_0 \text{ and therefore } \hat{x}=\Sigma^{-1}\mathrm{E}\,(\mathrm{E}'\Sigma^{-1}\mathrm{E})^{-1}\mathrm{E}_0. \tag{24}$$

$$\text{If we denote (the 3 by 1 matrix) } \theta=(\theta_1,\theta_2,\theta_3)'=(\mathrm{E}'\Sigma^{-1}\mathrm{E})^{-1}\mathrm{E}_0, \tag{25}$$

$$\hat{x}=\theta_1\Sigma^{-1}e+\theta_2\Sigma^{-1}\mu+\theta_3\Sigma^{-1}d \ . \tag{26}$$

Remark 1:

It is possible to compute expressions for θ_i, by computing the inverse of the (3 by 3) matrix $(\mathrm{E}'\Sigma^{-1}\mathrm{E})^{-1}$.

Remark 2

Equation (26) shows that the minimum variance frontier portfolio satisfies three-fund separation. As we vary Π and d_0, we will trace out a minimum variance frontier in three dimensions, ((σ, Π, d) space).

The equation for the efficient frontier will thus be given by,

$$\sigma^2=\mathrm{E}_0'\,(\mathrm{E}'\Sigma^{-1}\mathrm{E})^{-1}\mathrm{E}_0. \tag{27}$$

Equation (27) is clearly quadratic in σ, Π and d. Whilst it is valid to assume that $\mathrm{d}\sigma/\mathrm{d}\Pi$ is positive, it may not necessarily be the case that $\mathrm{d}\sigma/\mathrm{d}d_0$ can be signed. Thus, there is no geometrically determined market portfolio, but rather a region.

Conclusions

We have developed a framework to include a duration parameter within the CAPM equilibrium framework based on three fund separation. We could potentially extend this analysis further by assuming we have types of fund managers/ individuals as follows: (i) Duration neutral, but return-loving investor. (ii) Return neutral, but duration-loving pension fund. (iii) Other combinations are possible. These assumptions would readily allow us to generate an equilibrium in which we might be able to determine the *V* function. Thus our analysis allows us to define short-termism in a natural way. It can be used to compute the equity risk premium for individual stocks, and also give potential explanations for why small stocks can fall prey to the low beta anomaly, at least under certain circumstances.

Notes

[1] 'Budget Resolutions and Economic Situation', *Hansard*, 7th July 1997, pp.688-690, and the Finance Bill, *Hansard*, 16th July 1997, pp.475-476.

References

Damant, David C. and Stephen. E Satchell, 'Testing for Short Termism in the UK Stock Market: A Comment', *Economic Journal*, Vol.105, No.432, September 1995, pp.1218-1223.

Fama, Eugene F. and Kenneth R. French, 'The Cross-Section of Expected Returns', *Journal of Finance*, 1992, 47, pp.427-465.

Fama, Eugene F. and Kenneth R. French, 'Common Risk Factors in the Returns on Stocks and Bonds', *Journal of Financial Economics*, 1993, 33, pp.3-56.

Hutton, Will N., *The State We're In, Why Britain Is in Crisis and How to Overcome It*, Vintage Publications, London, 1996.

Hutton, Will N., *The State To Come*, Vintage Publications, London, 1997.

Ingersoll, Jonathan E., 'Multidimensional Security Pricing', *Journal of Financial and Quantitative Analysis*, Vol.10, No.5, 1975, pp.785-798.

Jean, William H., 'The Extension of Portfolio Analysis to Three or More Parameters', *Journal of Financial and Quantitative Analysis*, Vol.6, No.1, January 1971, pp.505-515.

Jean, William H., 'More on Multidimensional Portfolio Analysis', *Journal of Financial and Quantitative Analysis*, Vol.8, No.3, June 1973, pp.475-490.

Leibowitz, Martin. L., A. Bova, and P. B. Hammond, Chapter 14 'Real Return Tents and Equity Durations', in *The Endowment Model of Investing: Return, Risk, and Diversification* , Hoboken, NJ: John Wiley and Sons, 2010, pp.213–223.

Lewin, Richard A., Marc J. Sardy, and Stephen E. Satchell, 'UK measures of firm-lived equity duration', *International Finance Review*, 7, 2007, pp.335–347.

Lintner, John, 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets', *Review of Economics and Statistics*, 47 (1), 1965, pp.13–37.

Mata, José and Portugal, Pedro, 'Life Duration of New Firms', *Journal of Industrial Economics*, Vol. 42, 1994, pp.227-245.

Merton, Robert .C. 'An Intertemporal Capital Asset Pricing Model', *Econometrica*, Vol.41, No 5, September 1973, pp.867-887.

Miles, David K., 'Testing for Short Termism in the UK Stock Market', *Economic Journal*, Vol.103, No.421, November 1993, pp.1379-1393.

Miles, David K., 'Testing for Short Termism in the UK Stock Market: A reply', *Economic Journal*, Vol. 105, No. 432, September 1995, pp.1224-1227.

Mossin, Jan 'Equilibrium in a Capital Asset Market', *Econometrica*, Vol. 34, No. 4, 1966, pp.768–783.

Pickens, T. Boone, 'Professions of a Short-Termer', *Harvard Business Review*, Vol.64, Issue 3, May-June 1986, pp.75-79.

Schröder, David and Florian Esterer, 'A New Measure of Equity and Cash Flow Duration: The Duration-Based Explanation of the Value Premium Revisited', *Journal of Money, Credit & Banking*, Volume 48, Issue 5, 2016, pp.857-900.

Shaffer, Sherrill L., 'Equity Duration and Convexity when Firms can Fail and Stagnate', *Finance Research Letters*, Vol.4, 2007, pp.233-241

Sharpe, William F., 'Capital asset prices: A theory of market equilibrium under conditions of risk', *Journal of Finance*, 19 (3), 1964, pp.425–442.

Spending Vs Savings: Does The Source Of A Cash Flow Matter?

Valrie Chambers, Stetson University
Eugene Bland and Marilyn Spencer, Texas A&M University – Corpus Christi

Abstract

Earlier research showed that tax rebate spending/saving choices depended on the rebate distribution - as a lump sum or a series of partial rebates totaling the same amount. This study examines whether people do use different mental accounts for different types of hypothetical revenue windfalls - work bonus, game show lottery winnings, or inheritance - rather than viewing these small revenue windfalls as fungible in their use consistent with neoclassical economics.

This study finds that the income source sometimes influenced the amount spent/saved. A greater percentage of game show winnings would be spent than a windfall from any of the other sources - lottery, work bonus, tax rebate, or inheritance, with a greater inheritance percentage spent than from a bonus or tax rebate.

The order of presentation also mattered. When distributions of monthly payments for a year were presented first, more would be saved than if the same amount of distribution were paid as one lump sum.

A respondent's general status as a spender or saver was highly significant in all regressions. As suggested by Spencer and Chambers (2012), the consumer's spending/saving default is important. Those indicating that they would generally spend a windfall, did.

This paper adds to the literature by responding to Epley and Gneezy's (2007) call for "a broader sample of participants, varying amounts of payment, and alternative frames" to "identify important and interesting moderators of windfall framing effects," with implications for behavioral economic theory, compensation theory, and financial planning.

Introduction

The difference that existed between what taxpayers did with tax rebates that are paid out monthly versus tax rebates of a similar amount that are paid out in a lump-sum is now well documented (Chambers and Spencer, 2008; Sahm, et al., 2012). But is this difference related to the source of the payment, specifically tax rebates, or does this effect extend to other sources of transitory payments, e.g. lottery winnings, when the timing of a fixed amount is altered? That is, do people's mental accounts (Thaler, 1999) depend not only on timing and use of money, but also on the source? To answer this question, this study tests whether people spend a distribution from a hypothetical tax rebate as they would if the distribution came from another source, such as a bonus from work, a game show winning, an inheritance or a lottery winning.

This study extends research into what determines whether people earmark income from different sources in making allocations by asking how an income recipient might consider some of these sources as similar and others as different. How might the recipient consider some of these sources as similar and others as different? Lottery winnings are similar to tax rebates in the United States, in that both lottery systems and tax systems are run by a government or its appointed agency. Both types of payment amounts are largely outside the respondent's control. To what extent the money is "earned" is debatable in both cases, but bonuses and game show winnings - and sometimes inheritances - require some personal effort. Tax rebates sometimes differ from the other four sources of payment because the tax rebate is a refund or return of withholdings the taxpayer has previously paid in. That is, outside of refundable credits tied to specific performance, respondents generally cannot materially profit from a tax rebate because it is a refund of money already paid in, but can profit from a lottery, game show or bonus. An inheritance is not a profit, per se, but is generally not a return of one's own capital. Inheritances might be property or money that carries with it memories of the decedent, and those emotions might carry over to how the respondent intends to use the inheritance. Further, some political rhetoric frames taxes as money belonging fundamentally to taxpayers, not the government, whereas lottery winnings come with no similar sense of entitlement. Bonuses are likely to be closely tied to an individual's performance, however. Game show winnings might be as well, if the winner attributes success to having a higher skill level than fellow contestants.

If significant differences are found, such results would imply that money is not as fungible as commonly thought, and people's mental accounts are not just a function of use and timing, but also of source, consistent with mental accounting theory and representing a contribution to literature. Conversely, if no significant differences are found, the results would imply the fungibility of money, consistent with neoclassical economic theory, and may represent a limitation of mental accounting theory, representing a different contribution to literature.

Literature Review

Overview

According to mental accounting theory, people create different mental accounts (e.g. long-term savings), and have different marginal propensities to consume from each account. Numerous studies support mental accounting from a regular income flow or from an irregular, lump-sum windfall (Johnson, et al., 2006; O'Curry, 1999; and Souleles, 2002). Informally, people periodically reconcile their mental accounts for income and expense (Camerer et al., 1997; Heath and Soll, 1996; Read, et al., 1999; and Rizzo and Zeckhauser, 2003).

Karlsson, et al. (1999) reported that cash spending on a durable good depended on compatible reasons for saving. Abeler and Marklein (2008) found that high school math grades seemed to matter in mental budgeting, and Benjamin (2006) found a relationship between low high school test scores and non-rational behavior in general. Cheema and Soman (2006) and Wertenbroch (2001) concluded that mental budgeting is a matter of self-control. Frederick (2005) reported a negative relationship between non-rational behavior and cognitive reflection.

Source Literature

Some evidence suggests that the source of one's income does affect the use of those funds. In 1992, Henderson and Peterson reported that an individual would be more likely to spend $2,000 on a vacation if the funds were a gift rather than a work bonus. Arkes, et al. (1994) found that a greater percentage of a small amount of unexpected income (windfalls) was spent than the percentage of the same amount of anticipated income. Dobbelsteen and Kooreman found in 1997 that individuals were more sensitive to changes in a child's allowance than to other income sources for the decision to spend on their child's clothing. Winkelmann, et al. (2010) used evidence from German lottery winners and a theoretical model to show that different sources of income spent did confer different marginal utilities. Thus the purchase of an item with one source of funds provided a different marginal utility than another, and that it takes about two years before lottery winners feel that they "deserve" their good fortune. Bradford (2008) found that individuals allocate gifted and inherited assets in support of relational goals. Similarly, Trump, et al. (2015) found that individuals would make riskier choices with a stranger's money than with a friend's money.

Still, the framing of payments seems to matter: Baker, et al. (2007) found that more money was spent from likely recurring income (dividends) than less regular capital gain income. Epley, et al. (2006) found that people spent more from an income source labeled "bonus" than they did of a "rebate" of the same amount and timing. Similarly, Shefrin and Thaler (1988) found that more of a lump sum bonus is saved than if the same amount increases regular income, even when the bonus is fully anticipated.

Effort

The amount of effort required may affect responses. Boylan (2010) found that compliance with the tax system is influenced by whether taxable income is earned or endowed. Epley and Gneezy (2007) reported that a windfall that positively deviates from the status quo, like a bonus, is more likely spent than a windfall that restores the status quo. Zagorsky (2013), studied consumption of inherited money and found that over 40% of those who inherited less than $1,000 spent their bequest. Only 18.7% of those receiving $100,000 or more spent it all. In all, this research indicates that only about one half of inherited money was retained, the remainder was reduced by capital losses or is spent. Agarwal and Qian (2013) studied how consumers responded to an exogenous income shock, and found that consumption rose significantly at the rate of $0.80 per $1 received. Spending began with the announcement of the income shock. Low-liquidity consumers and low-credit consumers consumed more.

Frequency of Distribution

Neoclassical economics assumes that the decision to spend, and how to spend one's income would not depend on the way in which it is distributed. Yet the difference in spending patterns from a limited number of monthly payments and a lump-sum tax rebate of the same amount is well documented. Rucker (1984) studied the retroactive payment of a raise approved by a university, reversed by the Federal Pay Board but reinstated by the U.S. Supreme Court. The size of the windfall was found to be the most important factor for deciding how the funds were spent, with smaller checks more likely to be consumed. In addition, the length of time that the recipient had to anticipate the receipt of the funds also influenced the use of the money. The shorter the time before the receipt of the money was anticipated, the more likely that the money was consumed. Karlsson,

et al. (1999) noted that individuals considered the future consequences of spending in their mental budgeting, which may indicate a contemplation of permanent income.

Shapiro and Slemrod (1995) found that almost half the respondents surveyed would spend the 1992 decreased tax withholding refunded to them, even though the total yearly tax liability remained unchanged, resulting in a lower end-of year tax refund. However in 2001, when a tax cut took the form of either a $300 or $600 lump-sum rebate, only about one-fourth of those surveyed expected to spend the payment (Shapiro and Slemrod, 2003). Slemrod and Bakija (2004) attributed the change in behavior of taxpayers between the differently distributed rebates to changes in economic conditions, however applying Thaler's (1999) mental accounting theory, Chambers and Spencer (2008) found that the timing of payments (whether paid as a lump-sum, or spread out in equal monthly installments for a year) matters. This was confirmed by Sahm, et al. (2012).

Permanence of Distribution

Neoclassical economics tells us that neither the marginal cost nor the marginal benefit of a purchase is dependent on the source of the income from which it is spent. The permanence of payments may also be a factor in how much people choose to save. Blinder (1981) posited that a permanent tax decrease would elicit more spending than a temporary tax rebate, which he surmised would be treated as one half from a normal income tax change and the other half from a windfall. Parker (1999) studied tax cuts, finding that a temporary, end-of-year reduction in social security tax for high-income wage earners was spent when received, not averaged evenly over the fiscal year. Friedman's (1957) permanent income hypothesis says that people will spend money consistent with what they believe to be their permanent income level, but stopped short of examining the source of the income or testing the spending on amounts of limited duration.

Studies of unique, one-time payments are rare. However, Bodkin (1959) estimated the marginal propensity to consume to be between 0.72 and 0.97 of a one-time dividend paid in 1950 to World War II veterans by the National Service Life Insurance. The payments averaged $175, roughly $1,723.39 in 2015 dollars (BLS.gov, 2016). Similarly, Kreinin (1961) analyzed the spending of a sample of Israeli citizens receiving restitution payments from Germany in 1957 and 1958 and estimated that 35% was spent while 65% of the restitution payment was saved, with 45% saved in liquid assets and 20% in real estate (Kreinen, 389).

Materiality of Amount

Chambers, et al. (2009) studied responses to small hypothetical tax rebates, of the size distributed in 2008, $300 and $600, as well as larger amounts, $1,500 and $3,000. They found that at some amount over $600, materiality mattered greatly in how the money would be used. Under the $600 amount, individuals were likely to spend a rebate if that was the government's intent for distributing it, but at or above $600, the government's wishes were ignored (Chambers, et al., 2009).

Research on large, regular bonuses includes Hsieh (2003) who studied consumption associated with receipt of the Alaska Permanent Fund. The annual receipt was fully anticipated and no spike in consumption was found. However, consumption by the same households was very responsive to income tax refunds. Hsieh wrote, "This evidence suggests that households will take anticipated income changes into account in their consumption decisions when the income changes are large, regular and easy to predict, but will not do so when they are small and irregular" (Hsieh, 2003, 397). Another situation with large, regular and predictable bonuses was documented by Browning and Collado (2001). They studied Spanish panel data to measure the effect of the bonus payments customary in that market. Workers in this bonus paying scheme usually received payments of $1/14^{th}$ of their annual wage per month for 10 months. However, in two months, usually December and June or July, they received $2/14^{th}$s of their salary. They did not "find any effect of anticipated changes in income on expenditure patterns over the year for those who receive the bonus payments are indistinguishable from the patterns of those who do not receive a bonus," (Browning and Collado, 2001, 682).

Research Questions

In light of this literature, does a different source of a payment change a consumer's amount saved, controlling for the amount of the payment and the distribution frequency? Or is the timing more general and stubbornly entrenched enough to be the same no matter the source? Stated as research questions in Table 1 below, how does the saving from bonuses differ from either a tax rebate, work, game show winnings, lottery winnings or inheritance? And, how does the saving from other pairs of sources differ, if at all? These research questions below attempt to capture whether people earmark income from different sources, and from different timing, in making that allocation.

Table 1. Research Questions

RQ1	Do people intend to save the same amount of a hypothetical lump sum (monthly) bonus payment as they would a hypothetical lump sum (monthly) game show winning?
RQ2	Do people intend to save the same amount of a hypothetical lump sum (monthly) bonus payment as they would a hypothetical lump sum (monthly) inheritance?
RQ3	Do people intend to save the same amount of a hypothetical lump sum (monthly) bonus payment as they would a hypothetical lump sum (monthly) lottery winning?
RQ4	Do people intend to save the same amount of a hypothetical lump sum (monthly) bonus payment as they would a hypothetical lump sum (monthly) tax rebate?
RQ5	Do people intend to save the same amount of a hypothetical lump sum (monthly) from game show winnings as they would a hypothetical lump sum (monthly) inheritance?
RQ6	Do people intend to save the same amount of hypothetical lump sum (monthly) game show winnings as they would a hypothetical lump sum (monthly) lottery winning?
RQ7	Do people intend to save the same amount of hypothetical lump sum (monthly) game show winnings as they would a hypothetical lump sum (monthly) tax rebate?
RQ8	Do people intend to save the same amount of a hypothetical lump sum (monthly) inheritance as they would a hypothetical lump sum (monthly) lottery winning?
RQ9	Do people intend to save the same amount of a hypothetical lump sum (monthly) inheritance as they would a hypothetical lump sum (monthly) tax rebate?
RQ10	Do people intend to save the same amount of a hypothetical lump sum (monthly) lottery winning as they would a hypothetical lump sum (monthly) tax rebate?

In answering these questions, the amount of the income was controlled for, as were the order of presentation and the demographic characteristics of the respondents. This study extends that research to determine whether people earmark income from different sources in making that allocation.

Methodology

Sheppard, et al.'s (1988) meta-analysis of 86 theory-of-reasoned-action studies found a 0.53 correlation between intention and behavior, indicating that intent is a good predictor of action. For this study, 80 different instruments were developed to test the intended spending/saving patterns of respondents. Participants were given one of these 80 instruments at random and asked how they would use the funds, both if they were to receive a lump-sum and if they were to receive the same amount spread out over 12 equal monthly payments (within-subject design), from two of these five sources: bonus, game show winnings, inheritances, lottery winnings and tax rebates (between-subjects design). Each instrument hypothesized one of these four different amounts: $300, $600, $1,500, $3,000. Some instruments presented the periodic amounts first and some presented the lump-sum amounts first to test for the order effect.

The instruments asked how much of a lump sum refund would be used for: (1) investing, (2) paying off credit card debt, (3) paying off notes, (4) regular monthly expenses, (5) buying a durable asset, (6) saving for an infrequent expense, and/or (7) used for fun. Hershfield, et al. (2015) found that consumers' tendency to place savings and debt into separate mental accounts makes them insensitive to the significant differences between the interest rates on these accounts. The instrument also asked how much of a monthly payment (equal to 1/12 of the lump sum amount) would be used for each of these seven purposes, consistent with Chambers and Spencer (2008). Similarly, the flip side of each instrument asked these same questions, changing only the source of the payment from one source to another – such as from a tax rebate to a lottery, work bonus, inheritance or game show payment. Experimental questionnaires were distributed to university students at these universities: Coastal Carolina University, Francis Marion University, Longwood University, Metropolitan State University of Denver, Texas A & M

University - Corpus Christi, University of Alabama – Birmingham, and University of Houston-Clear Lake. Students were considered provisionally acceptable respondents per Walters-York and Curatola (1998) and Ashton and Kramer (1980).

All research questions were analyzed with descriptive statistics, converted to percentages, and then then analyzed using four sets of OLS regressions, where the choices were (1) investing, (2) paying off credit card debt, (3) paying off notes, and (6) saving for an infrequent expense were coded as savings, and choices (4) regular monthly expenses, (5) buying a durable asset, and (7) used for fun were coded as spending. Two of the sets of regressions (one where the monthly distribution was shown first, the other where the lump sum was shown first) used "longer-term savings" as its dependent variable, which excluded savings from item (6), saving for an infrequent expense. The other two sets of regressions (one where the monthly distribution was shown first) used dependent variables dubbed "total savings" included in item (6). The four dependent variables were each regressed against the source of the windfall income, and demographic variables were included to control for income, gender, age, importance to the budget, business experience level and education level.

The regression models were of the form:

> Percent Saved = F(income, zero income, amount, education, gender, age, importance, seatbelt use, smoker, spend1 (default for spender), experience level, dummy variables for the source of the payment (lottery, tax rebate, inheritance, game show, or bonus), and a dummy for the order of presentation (monthly payment first, or lump sum payment first)).

"Income" is the log of the respondent's income plus one. "Amount" is the hypothetical amount of the distribution, in dollars. As four discrete values were possible for the amount, dummy variables were created for each amount rather than treat this variable as continuous. Education is divided into four categories: high school, associate degree, undergraduate degree, and graduate degree. "Gender" is a categorical male/female variable, where female was coded as "1." "Age" is the participant's age in years. As there may be some nonlinearity in the age variable, the square of age, "AgeSq" was added to the model to measure the non-linear contribution to the dependent variable that occurs as the reported age increases. "Importance" was defined to be the payment divided by the income of the survey participant. The "Seatbelt" and "Smoker" dummy variables were included as proxies for respondents' risk preference; seatbelt wearers and smokers were coded as "1." For the variable "Spend1" the participants were asked "When you get 'extra money,' do you spend it or save it?" The dummy was set to 1 for those that answered "spend." Experience was a categorical, self-reported measure coded as "0" for "none", "1" for responses of "low" and progressing upward to "5" for "high," This categorical variable was then transformed to dummy variables for use in the regression as described below. Various formulations of the credit card debt variable were also introduced to observe whether debt in dollars or as some proportion would affect the results; the results were not affected, however.

Results

Table 7. Significance of the Source of Payment

A. Longer Term Savings – Lump Sum Payment			
Sources	**Beta**	**P-value**	**Research Question**
At the 1% significance level			
Game Show Vs. Tax Rebate	-.110	.001	7
At the 5% significance Level			
Game Show Vs. Lottery	-.067	.047	6
Game Show Vs. Inheritance	-.065	.045	5
At the 10% Significance Level			
Bonus Vs. Tax Rebate	-.066	.056	4

An economically and statistically significant lower amount was saved from game show winnings than were saved from lottery winnings, inheritance and tax rebates. Less of a bonus was saved than a tax rebate, but this was significant at slightly more than the 5% level.

B. Total Savings – Lump Sum Payment			
Sources	**Beta**	**P-value**	**Research Question**
At the 10 % significance level			
Game Show Vs. Tax Rebate	-.056	.063	7

There was weak support indicating that 5.6% less was saved from game show winnings than from tax rebates.

C. Longer Term Savings – Monthly Payments			
Sources	**Beta**	**P-value**	**Research Question**
At the 1% significance level			
Game Show Vs. Tax Rebate	-.1058	.005	7
At the 5% significance Level			
Game Show Vs. Lottery	-.079	.036	6
At the 10% Significance Level			
Game Show Vs. Bonus	-.068	.065	1
Inheritance Vs. Tax Rebate	-0.69	.066	9

Strong economic and statistically significant results showed that savings from game show winnings were less than those from lottery and tax rebates. Weak statistical support indicated that savings from game show winnings were lower than that from bonus payments; and that savings from inheritances were lower than that from tax rebates.

D. Total Savings – Monthly Payments			
Sources	**Beta**	**P-value**	**Research Question**
At the 1% significance level			
Game Show Vs. Lottery	-.096	.008	6
Game Show Vs. Bonus	-.108	.002	1
At the 5% significance Level			
Game Show Vs. Tax Rebates	-.076	.033	7
Inheritance Vs. Bonus	-.077	.033	2
At the 10% Significance Level			
Inheritance Vs. Lottery	-.065	.072	8

Strong economic and statistically significant results showed that savings from game show winnings were less than those from lottery, bonus or tax rebates. In addition, less of inheritance payments were saved than from bonus payments. Weak evidence suggested that less of inheritance payments were saved than from monthly lottery payments.

The results summarized in Table 7 show the noteworthy results of the tests for Research Questions 1-10 in Table 1. In at least one set of regressions, the results indicated that savings from game show winnings were significantly lower than from bonuses, inheritances, lottery winnings or tax rebates.

Discussion

There can be little doubt that the uses of "windfall" income depend on the income source – especially given that regressions indicated that game show winnings would be used differently from those of tax rebate income (Research Question 7). This is a clear exception to neoclassical economic theory but consistent with mental accounting theory and behavioral economics. Except for game show winnings, a greater percentage of money from an inheritance was spent than was money from a bonus, tax rebate or lottery winning. As the amounts used in this study ranged from $300 to $3000, these results are consistent with the findings of Zagorsky (2013), who found that over 40% of those who inherited less than $1,000 spent their entire bequests.

However, there were no significant differences for several of the other Research Questions, consistent with neoclassical economic theory and possibly indicating a boundary for behavioral economic theory. It is as if, generally, people treat money as fungible when it comes in, but then use mental accounting "buckets" to determine where it will go out, combining elements of both neoclassical economic theory and behavioral economics.

These findings are important, first, because most of the previous evidence on savings in favor of the behavioral economics and behavioral accounting approach is in the area of expenditures, not source of revenue. More importantly, neoclassical economic theory and behavioral economics are seen as somewhat competing theories when in fact there may be a place for both. People could decide that money from one source will go to one set of bills, and money from another will go to savings or another set of bills. It appears that they don't. It appears that they mix their revenues together, then decide how to allocate the mixed pool of money. That is, people tend to be neoclassical, but not rigidly neoclassical when making revenue decisions, and follow behavioral economics when making expenditure decisions. Finding a balance between competing theories materially adds to current literature.

Conclusion

Earlier research showed that tax rebate spending and saving choices are affected by whether the rebate is distributed as a lump sum or as a series of partial (e.g. monthly) rebates totaling the same amount, and that people allocate their income into different mental accounts for purposes of spending and saving. This study extends that research to determine whether people earmark income from different sources in making that allocation, or whether income from several types of hypothetical windfalls was allocated equivalently across spending vs. savings.

Although economic theory would say the source of the money is irrelevant, this study finds that the source sometimes influenced the amount spent or saved, consistent with mental accounting theory. A greater percentage of game show winnings would be spent by respondents than a windfall from any of the other sources - lottery, bonus, tax rebate, or inheritance. Except for game show winnings, a greater percentage of an inheritance receipt would be spent than from a bonus or tax rebate. However, for several of the other Research Questions, there were no significant findings. It appears that people mix their revenues together, then decide how to allocate the mixed pool of money. That is, people tend to be flexibly balanced between neoclassical and behavioral economics. They are predominantly neoclassical when making revenue decisions, but predominantly behavioral economic beings when making expenditure decisions.

This project extends previous behavioral economics literature to study the effects of the sources of revenue windfalls and could have practical implications for behavioral economic theory and financial planning practitioners. Theoretically, finding an applied balance between competing theories adds to the current behavioral economic and mental accounting literature.

References

Abeler, J., & Marklein, F. (2008). Fungibility, labels, and consumption. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1139870

Agarwal, S., & Qian, W. (2013). Consumption and debt response to unanticipated income shocks: Evidence from a natural experiment in Singapore. *Unpublished working paper. National University of Singapore*.

Arkes, H. R., Joyner,C.A., Pezzo, M.V., Nash, J.G., Siegel-Jacobs, K. & Stone, E. (1994). The psychology of windfall gains. *Organizational Behavior and Human Decision Processes, 59*(3), 331–347.

Ashton, R. H., & Kramer, S. S. (1980). Students as surrogates in behavior accounting research. *Journal of Accounting Research, 18*(1), 1-15.

Baker, M., Nagel, S. & Wurgler, J. (2007). The effect of dividends on consumption. *Brookings Papers on Economic Activity, 1*(2007), 231-276.

Benjamin, D. J. (2006). *Psychology, preferences, and markets. ProQuest Dissertations and Theses*, 231-231 Retrieved from: http://search.proquest.com/docview/305344915?accountid=7084 (305344915).

Blinder, A. S. (1981). Temporary income taxes and consumer spending. *Journal of Political Economy, 89*(1), 26-53.

Browning, M., & Collado, M.D. (2001). The response of expenditures to anticipated income changes: Panel data estimates. *The American Economic Review, 91*(3), 681-692.

Bodkin, R. (1959). Windfall income and consumption. *The American Economic Review, 49*(4), 602-614.

Boylan, S. (2010). Prior audits and taxpayer compliance: Experimental evidence on the effect of earned versus endowed income. *Journal of the American Taxation Association, 32*(2), 73-88.

Bradford, T.W. (2008). Intergenerationally gifted asset dispositions. *Journal of Consumer Research, 36*(1), 93-111.

Databases, Tables & Calculators by Subject – CPI Inflation Calculator. Bureau of Labor Statistics. Retreived from: http://bls.gov/data/inflation_calculator.htm

Camerer, C., Babcock, L., Loewenstein, G. & Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *Quarterly Journal of Economics, CXIII* (May), 407-442.

Chambers, V., & Spencer, M. (2008). Does changing the timing of a yearly individual tax refund change the amount spent vs. saved? *Journal of Economic Psychology, 29*(6), 856-862.

Chambers, V., Spencer, M., & Mollick, J. (2009). Goldilocks rebates: Complying with government wishes only when the rebate amount is "just right." *Journal of Economics and Economic Research, 10*(1), 101-119.

Cheema, A., & Soman, D. (2006). Malleable mental accounting: The effect of flexibility on the justification of attractive spending and consumption decisions. *Journal of Consumer Psychology, 16*(1), 33-44.

Dobbelsteen, S., & Kooreman, P. (1997). Financial management, bargaining and efficiency within the household; an empirical analysis. *De Economist, 145*(3), 345-366.

Epley, N., Mak, D., & Idson, L. (2006). Bonus or rebate?: The impact of income framing on spending and saving. *Journal of Behavioral Decision Making, 19*(3), 213-227.

Epley, N., & Gneezy, A. (2007). The framing of financial windfalls and implications for public policy. *The Journal of Socio-Economics*, *36*(1), 36-47.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25-42.

Friedman, M. (1957). *A theory of the consumption function.* Princeton: National Bureau of Economic Research.

Heath, C., & Soll, J.B. (1996). Mental accounting and consumer decisions. *Journal of Consumer Research, 23*(1), 40-52.

Henderson, P. W., & Peterson, R. A. (1992). Mental accounting and categorization. *Organizational Behavior and Human Decision Processes*, *51*(1), 92-117.

Hershfield, H. E., Sussman, A.B., Rourke, L.O., & Bryan, C. J. (2015). Leveraging psychological insights to encourage the responsible use of consumer debt. *Perspectives on Psychological Science, 10*(6), 749-752.

Hsieh, C. (2003). "Do consumers react to anticipated income changes? Evidence from the Alaska permanent fund." *The American Economic Review*, *93*(1), 397-405.

Internal Revenue Service (2012). Retrieved from: http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Returns

Johnson, D. S., Parker, J.A. & Souleles, N. S. (2006). Household expenditure and the income tax rebates of 2001. *American Economic Review*, *96*(5(December)), 1589-1610.

Karlsson, N., Gärling, T., & Selart, M. (1999). Explanations of effects of prior income changes on buying decisions. *Journal of Economic Psychology*, *20*(4), 449-463.

Kreinin, M. E. (1961). Windfall income and consumption: additional evidences. *The American Economic Review*, *51*(3), 388-390.

O'Curry, S. (1999). Income source effects. *Unpublished working paper, DePaul University*.

Parker, J. (1999). The reaction of household comsumption to predictable changes in Social Security taxes. *American Economic Review*, *89*(4), 959-973.

Read, D., Loewenstein, G. & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, *19*, 1-3 and 171-197.

Rizzo, J. A., & Zeckhauser, R. J. (2003). Reference incomes, loss aversion, and physician behavior. *The Review of Economics and Statistics, 85*(4), 909-922.

Rucker, M. H. (1984). Allocation of windfall income: a case study of a retroactive pay increase to university employees. *The Journal of Consumer Affairs, 18*(1), 101-118.

Sahm, C. R., Shapiro, M. D. & Slemrod, J. (2012). Check in the mail or more in the paycheck: Does the effectiveness of fiscal stimulus depend on how it is delivered? *American Economic Journal-Economic Policy, 4*(3), 216-250.

Shapiro, M. D., & Slemrod, J. (1995). Consumer response to the timing of income: Evidence from a change in tax withholding. *American Economic Review, 85*(1), 274-283.

Shaprio, M. D., & Slemrod, J. (2003). Did the 2001 tax rebate stimulate spending? Evidence from taxpayer surveys. In J. M. Poterba (Ed.), *Tax policy and the economy,* (p. 17). Cambridge, MA: *National Bureau of Economic Research and MIT Press*, 83-109.

Shefrin, H. M., & Thaler, R. H. (1988). The behavioral life-cycle hypothesis. *Economic Inquiry, 26* (October), 609-643.

Sheppard, B., Hartwick, J., & Warshaw, P. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research, 15*(3), 325-343.

Slemrod, J., & Bakija, J. (2004). *Taxing ourselves: A citizen's guide to the debate over taxes (3rd Ed.).* Cambridge, MA: The MIT Press.

Souleles, N. S. (2002). Consumer response to the Reagan tax cuts. *Journal of Public Economics, 85*(1), 99-120.

Spencer, M., & Chambers, V. (2012). National heuristic shift toward saving any form of tax rebate. *Accounting and the Public Interest, 12*(1), 1-31.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making, 12*(3), 183-206.

Trump, R. K., Finkelstein, S. R. & Connell, P. M. (2015). I will risk a stranger's money, but not my own or my friend's money: Effect of proximity of the money source to the self on financial risk-taking. *Marketing Letters, 26*(4), 501-512.

Walters-York, L. M., & Curatola, A. P. (1998). Recent evidence on the use of students as surrogates subjects. *Advances in Accounting Behavioral Research, 1*(1998), 123-143.

Wertenbroch, K. (2001). Self-rationing: Self-control in consumer choice.

Winkelmann, R., Oswald, A.J., & Powdthavee, N. (2011, August). What happens to people after winning the lottery? In *European Economic Association & Econometric Society Parallel Meetings* (pp. 25-29).

Zagorsky, J. L. (2013). Do people save or spend their inheritances? Understanding what happens to inherited wealth. Journal of Family Economic Issues, 34*(1), 64-76.*

Capacity Management in Higher Education: An Exploration of Emerging Policies and a Call for Further Research

Karen A. Loveland, Texas A&M University, Corpus Christi
Eugene Bland, Texas A&M University, Corpus Christi

Abstract

The term capacity management refers to a broad set of strategies/tactics used to adjust an organization's ability to produce its product to the level of market demand. Capacity management is especially well-studied in the field of services marketing due to the inseparable and perishable nature of service products. Surprisingly, academic researchers haven't devoted much attention to the issue of capacity management in higher education. Recent growth in enrolment combined with reductions in higher education funding related to the great recession has placed increased emphasis on capacity in public institutions in the U.S.

This paper begins with a discussion of capacity management principles and then applies selected principles to higher education institutions. We also identify a trend toward stricter space utilization policies affecting traditional course scheduling practices. Then, using raw schedule data, we explore the relationship between student learning outcomes and specific space utilization policies. Specifically, the space utilization policy studied increased the number of early morning classes, increased class sizes and classroom crowding during prime time, and reduced the number of classes that met two days per week. We found a negative association between student grades and each of the capacity management strategies. We also found a smaller, but significant, difference in perceived learning reported by students on the course evaluation. "One-size-fits-all" space utilization policies may have a differential impact on student outcomes in certain levels of courses or in specific colleges or disciplines. The paper concludes with a call for future research on the outcomes of capacity management strategies in higher education.

Introduction

Colleges and universities in the United States face significant pressure to "do more with less". Growth in enrolment combined with intensified budget scrutiny by legislators and other governing bodies makes efficient use of resources a growing concern for higher education administrators at many institutions. While we acknowledge and even applaud efforts to use resources more efficiently, recent policies at our institution targeting classroom utilization rates prompted concerns about the impact of such policies on educational quality. As we investigated those concerns, we came to realize that we were dealing with a basic capacity management issue.

The term capacity management refers to a broad set of strategies/tactics used to adjust an organization's ability to produce its product(s) to the level of market demand. While some authors distinguish between the management of supply (e.g., adding capacity with new plant/equipment or improving efficiencies in the use of human and capital resources) and the management of demand (e.g., using price or other variables to spread out demand), we follow the example of Klassen and Rohleder (2002) and use the term capacity management to refer to the collection of strategies an organization might use to adjust supply and/or influence demand.

Capacity management is especially well-studied in the field of services marketing due to the unique nature of service products. In particular, the commonly referenced characteristics of inseparability (services produced and consumed simultaneously) and perishability (services can't be stored) elevate the importance of strategies to balance supply and demand in service industries. Familiar examples of capacity management strategies range from restaurants that require reservations (to spread out demand and minimize waiting times) or offer lunch specials (to encourage off-peak patronage) to airlines adding flights (or new routes) to popular destinations to expand capacity to meet demand. Regardless of the specific strategy, savvy service marketers acknowledge the importance of assessing the impact of capacity management strategies on overall service quality.

When we adopt a capacity management perspective, we can see many examples of capacity management principles at work in higher education institutions faced with the twin challenges of increasing enrolment and shrinking funding in the past 20+ years. One of the most obvious, and widely studied, capacity-related strategies is the development of hybrid and online courses. Using technology to deliver some or all of the learning outside of a physical classroom allows institutions to satisfy additional demand without additional classroom space. Academic researchers have devoted considerable attention to the assessment of online and hybrid courses to ensure that educational quality doesn't suffer from this popular capacity management strategy.

Hiring part-time (adjunct) instructors to staff additional class sections and developing new courses or programs to relieve enrolment pressure on existing courses/programs are also strategies that can increase capacity. While not as popular as the more radical move to online education, researchers have also devoted considerable attention to assessing the impact of such strategies to satisfy increased demand.

Emerging Trends in Capacity Management

In the not-so-distant past, higher education governing bodies often based funding for new buildings mostly (or even solely) on enrolment. When student enrolment began to approach the limits of available classroom space during the most popular class times/days, administrators could request and often receive funding for additional capacity (i.e., new classroom space via new building construction). In recent years, however, expectations for increased accountability and efficiency in resource utilization appear to be motivating decision-makers to consider additional variables when assessing the need for new buildings/facilities.

One variable that seems to be receiving significant attention is the utilization rate for existing classroom space. A severe shortage of classroom space during "prime time" (generally defined as classes with a start time from 10:00 AM to 2:00 PM) is no longer sufficient to justify new buildings if the institution has a significant amount of unused capacity (i.e., open classrooms outside prime time and/or empty seats in classrooms during prime time). For example, in 2010, decision-makers in Texas increased the weight of the "classroom utilization rate" in the formula used for allocating funds for capital improvements (e.g., new classroom space). Public institutions in Texas responded by instituting new course scheduling policies. We found dozens of examples of similar institutional policies implemented across the country during "the great recession" of the last 5 to 10 years.

Of particular concern for many institutions is the use of classroom space during early morning, later afternoon, and evening hours. Numerous institutions have instituted formal policies to shift courses out of prime time, for example:

- Columbia University established a policy that requires each department to schedule no more than 10% of its courses during any one time/day block (Gaubatz 2003).
- The University of Arizona implemented a rule that allows each college to schedule no more than 70% of its courses during "prime time" (Gaubatz 2003).
- Texas A&M University, Corpus Christi (TAMUCC), requires that 10% of all three-credit undergraduate classes in each department must start at or before 9:30 AM (TAMUCC 2010).
- University of North Carolina, Charlotte (UNCC) requires that a minimum of 40% of all classes on Monday through Thursday be in early morning (defined as a start time of 8:00 AM) or late afternoon (defined as a start time of 3:30 PM or later) (UNCC 2012).
- The University of Michigan (UM) scheduling policy allows up to 35% of classes (and events) to be scheduled during prime time and requires at least 30% of classes (and events) to be scheduled before 10 AM or after 4 PM (UM 2013).

Classroom utilization on Friday is also an issue at some institutions. The TAMUCC scheduling policy requires each department to schedule at least 20% of its undergraduate lecture classes on MWF. In addition, the university eliminated all 75-minute classes on Monday and Wednesday between the hours of 8:00 AM and 2:00 PM to eliminate scheduling conflicts and underutilization of space on Fridays (TAMUCC 2010). The UNCC policy requires all colleges to schedule a minimum of 20% of their classes to include a meeting time on Friday (UNCC 2012). The goal at the University of Michigan is to schedule at least 15% of all classes (and events) on Fridays (UM 2012).

While some administrators suggest that faculty preferences are the primary reason institutions offer so many classes during prime time, it is possible instead that student preferences play a major role in prime time scheduling decisions. The implementation of space utilization policies such as those noted above gives rise to faculty concerns that such policies may negatively affect student learning outcomes and student satisfaction with their courses. If evolving space utilization policies force students into class periods that are less optimal, then student performance may suffer. Furthermore faculty may receive lower evaluations in these suboptimal schedule slots if students perceive lower levels of learning or lower course quality during the suboptimal schedule times/days. In the next section, we briefly review the literature studying the impact of class size, time of day, and days per week on student learning.

Prior Research

Our research crosses several threads in the body of knowledge. To begin, there is a vast literature discussing the impact of class size on student-related outcomes. In one of the earliest works on class size, Edmonson and Muldek (1924) compared grades and student perceptions of large and small classes. They found no significant difference in grades between the large and small class; however, they did find that students preferred the smaller class environment and perceived a higher quality of

learning in the smaller class environment. They also noted that, "…differences in the size of the rooms combined with the differences in the size of the classes had some influence on the procedure of the instructor," (Edmonson and Muldek 1924, 5). The larger class was more formal, had little discussion, more lectures and drill, and an impersonal attitude. They also noted a difference in behaviour including students being "mentally sluggish" and unlikely to take an active part in the discussion in larger classes (Edmonson and Muldek 1924). It seems that part of the problem with class size is that larger classes function differently from smaller classes.

Studies examining the relationship between class size and student performance (measures of learning) include Belante (1972), Raimondo, Esposito and Gershenberg (1990), Gibbs, Lucas and Simonite (1996), Kennedy and Siegfried, (1997), Bandiera, Larcinese and Rasul (2010), and Monks and Schmidt (2011). Two of these studies, Belante (1972) and Kennedy and Siegfried (1997), reported no relationship between student performance and class size. Belante (1972) studied three smaller size classes and a mass lecture course and measured the difference in student grades and class size. Kennedy and Siegfried (1997) used standardized test results to measure performance and found that class size did not affect student performance especially after student ability (SAT score for example) was considered. Interestingly, both studies were limited to economics classes.

A larger number of studies reported a negative relationship between class size and student performance. Gibbs, Lucas and Simonite (1996) studied 250,000 student grades from all 1st, 2nd and 3rd year classes at a UK university from 1984-1994. Their analysis found that "students in larger classes stood significantly lower chances of getting good grades" (Gibbs, Lucas and Simonite 1996, 261). Bandiera, Larcinese and Rasul (2010) studied full time students enrolled in one-year M.Sc. degree programs at a UK university and found, "…robust evidence of a negative class size effect – on average, larger classes reduce students' academic achievement as measured by test scores" (1395).

Raimondo, Esposito and Gershenberg (1990) added depth to our understanding of class size effects by measuring student performance in one class in relation to the size of the prerequisite class the student completed. While noting that the results of their article should be considered extremely preliminary, they suggested "…that students who took a large lecture introductory macroeconomics class have, ceteris paribus, lower grades in intermediate macroeconomics theory course than students who take a small lecture section introductory macroeconomics class" (Raimondo, Esposito and Gershenberg 1990, 379).

Monks and Schmidt (2011) provided a unique perspective on student performance by using information derived from student evaluations of teaching (SET). They found that class size had a significant negative effect on students' perceptions of the amount learned in the class. This finding prompted us to expand our literature review to include studies that examined the relationship between class size and SET ratings. While SET typically measures a wide variety of teaching behaviours and other factors beyond the scope of this paper, many SET instruments include at least one question measuring students' perceived learning in the class. All the studies reviewed revealed a significant negative relationship between class size and SET. We also learned that, while most of the studies using direct measures of student learning did not control for important exogenous variables, many of the SET-related studies of class size controlled for course level (e.g., upper level versus lower level) with significant results.

The discussion of class size above clearly establishes a relationship between one of the schedule-related variables targeted by emerging capacity management policies in higher education and student performance/learning (actual and/or perceived). Many institutions willingly acknowledge that class size is a course characteristic beyond the instructor's control and consider class sizes in evaluations of teaching performance. However, few studies have examined the other two schedule-related course characteristics we identified.

Each of the policies we reviewed included at least one provision designed to increase the number of courses offered early in the morning (start times before 10 AM). Most of the policies also contained at least one provision designed to increase the number of classes the meet three days per week (to increase classroom utilization on Fridays).

The only study we found examining both variables discussed above was Dills and Hernandez-Julian (2008). They collected grade information from 12,886 students (a total of 105,428 grades) to determine the extent that learning depends on class scheduling. They found that, "…students perform slightly better in courses offered later in the day" (Dills and Hernandez-Julian 2008, 647). They also found that, "students perform slightly better in courses offered…more days per week" (Dills and Hernández-Julián 2008, 647). However this finding was not universal, but dependent on the time of day that the class met. They indicated that, "the results follow a pattern of increasing grades throughout the day…the increase is most dramatic among TTh classes, but the pattern is consistent throughout," (Dills and Hernández-Julián 2008, 650). They went on to explain that, "grades are lower in MWF afternoon classes than in MW or TTh afternoon classes; grades are higher in MWF morning classes than in MW or TTh morning classes" (Dills and Hernández-Julián 2008, 651) .

Two additional studies included time of day effects. Smith and Stephens (2010) found that learning outcomes were significantly higher in accounting classes that met later in the day (10:00 AM start time) than in classes that met earlier in the day (8:00 – 9:00 AM start time). However, with the sample limited to two marketing classes and three accounting classes, the generalizability of these findings is questionable. Carrell, Maghakian and West (2011) was a larger-scale study that provided additional support for the contention that students earn lower grades in classes that meet earlier in the day. Studying freshman-

level classes at the US Air Force Academy, they examined student performance as a function of the start of the students' first class of the day. They found that, "Although students perform worse in first period classes compared to other periods, those with first period classes also perform worse in their subsequent classes on that schedule day" (Carrell, Maghakian and West 2011). The authors acknowledged that the unusually early start times (before 8 AM) at the subject institution and the restriction to freshman students limited the generalizability of their findings.

A larger number of studies examined the impact of the number of class meetings per week, but none of them specifically examined the impact of including a Friday class meeting. Most of these studies focused on a single course. Brookes (1985) studied student outcomes (grades) in 50 sections of a basic writing class at the community college level that offered classes that met one to five times per week. He found no significant impact on student grades across the range of class meeting days (Brooks 1985). Henebry (1997) studied average grades and non-passing rate in a financial management class taught in one day, two days and three days per week formats over the course of several years. She found that students were more likely to earn a passing grade in classes that met more than once a week. The study did not reveal a clear advantage for the two-day or three-day format. Gallo and Odu (2009) studied college algebra students and measured performance and the number of days that the class met. They found that students earned the lowest grades on unit and final exams in classes that meet one day per week. Their findings prompted them to note that, "although many students may prefer intensive courses or compressed schedules that minimize the time they spend on campus, these scheduling options may not be optimal for learning, at least not in mathematics" (Gallo and Odu 2009, 299). Most recently, Carrington studied the grades of 2,012 student grades in intermediate accounting and found a statistically significant relationship between course schedule and student grades. She found that the, "three day per week classes had the lowest success rate, followed by the one day per week schedule" (Carrington 2010, 56).

Reardon, Payan, Miller and Alexander (2008) expanded their study to include all undergraduate courses in a single discipline (marketing). They found that twice-a-week format resulted in higher student grades. They also noted that students perceived that the once-a-week format would result in lower levels of learning.

In the studies above, the authors appeared to attribute the observed days-per-week effect on student performance to the length of each class session (e.g., two-day-per week classes meeting for one hour and 15 minutes compared to three-day-per-week classes meeting for 50 minutes). This perspective makes sense in light of studies related to attention span and learning. However, anecdotal evidence (e.g., derived from student evaluation comments at the subject institution) suggests that some students prefer classes that do not meet on Friday. In addition, some instructors at the subject institution perceive that student attendance rates are lower on Friday. The combination of absenteeism and student preference may contribute to lower SET ratings and student performance in classes that include a Friday meeting.

The mixed results of the studies examining two common capacity management strategies emerging in higher education prompted us to conduct a broad exploratory study to identify suggestions to future research.

Exploratory Research

The data for our exploratory study came from two consecutive semesters immediately following implementation of a new classroom utilization policy across all colleges and departments at a university in the south-western region of the United States. The Office of Institutional Research compiled the data set by taking the publicly-available schedule data and adding the average GPA earned for each course and the average score for selected questions from the SET instrument.

We elected to exclude graduate-level classes from the compiled data set because such courses were not subject to the institution's space utilization policy. Furthermore, most graduate courses were offered in the late afternoon or evening time slots in a one- or two-day-per week format with a significantly smaller range of class sizes. We also eliminated courses offered in the college of nursing as most of the undergraduate courses in nursing were small online or clinical courses. Finally, we decided to limit the courses to "lecture-type" courses of three or more credit hours. The resulting data set included 1,071 courses with complete SET ratings, GPA information, and schedule-related variables.

To represent the broad concept of student learning, we used the average GPA earned in the course (SECT_GPA). The decision to use average grades is consistent with many of the larger studies investigating student performance across multiple fields/disciplines. Following the example of Monks and Schmidt (2011), we also included a measure of perceived learning from the SET ratings. Specifically, question 20 (Q20) on the SET instrument asked students to indicate their level of agreement with the following statement using a five-point Likert scale with 1 being "Strongly Disagree" and 5 being "Strongly Agree": "I perceive that my knowledge/skills in this content field have improved as a result of this course".

Course sizes ranged from 5 to 259 students with a mean enrolment of 36.1 students. We created a categorical measure that identified the cut points in class size for the 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentile to divide courses into Very Small (less than 19 students), Small (19 to 25 students), Medium (26 to 34 students), Large (35 to 61 students) and Very Large (62 or more students) classes. To explore the impact of increased classroom crowding related the space utilization policy, we created a dummy variable (CROWD) to represent courses in which enrolment was at least 90% of the classroom capacity.

To measure time of day effects, we created a categorical measure to divide courses into morning (MORNING – start times before 10:00 AM), primetime (PRIMETIME – start times between 10:00 AM and 1:45 PM), afternoon/evening (AFTERNOON/EVENING – course times from 2:00 PM to the end of the day. To measure the days per week effects, we created dummy variables for courses that met one (DPW1), two (DPW2), and three (DPW3) days per week. We also created a dummy variable (NO FRIDAY) to isolate the impact of courses that include a Friday class meeting.

We began with an exploratory analysis of the relationships between our independent variables and each of our two dependent measures: actual learning (SECT_GPA) and perceived learning (Q20). Table 1 shows selected results of this analysis.

We conducted t-tests within each set of variables to assess the significance of the differences in learning among the each level. For both measures of student learning, upper-division courses were higher than lower-division courses. Both differences were statistically significant (SECT_GPA mean difference = .416, t = 2.779, p = .003; Q20 mean difference = .121, t = 2.363, p = .018).

As expected from prior research, students earned the highest grades and perceived the highest levels of learning in very small classes (18 or fewer students) and the lowest grades and perceived learning in very large classes (62 or more students). For the measure of actual learning, the mean difference between the smallest class size group and the next higher group was not statistically significant nor was the difference between the second and third smallest groups. The difference between very small and medium classes was statistically significant as was the difference between medium and large classes and the difference between large and very large classes. Students in more crowded classes earned lower grades than students in less crowded classes; but, the difference was not statistically significant. Contrary to our expectations, students in crowded classes perceived a significantly higher level of learning (t = -2.658, df = 1,069, p = .008) than students in less crowded classes.

Students earned the lowest grades and perceived the lowest level of learning in early morning classes; however, none of the t-tests of the perceived learning measure were statistically significant indicating that students did not perceive a difference in learning across the three time-of-day categories. While students also earned the lowest grades in classes that met three days per week, those same students perceived higher levels of learning in the class. The same pattern applied to the Friday class meeting variable with students earning significantly lower grades in classes that include a Friday class meeting but not perceiving a significant difference in learning. However, for the perceived learning measure, none of the t-tests was statistically significant indicating that students don't perceive the measured effect of time-of-day or the inclusion of a Friday class meeting on their learning.

For our final set of exploratory analyses, we identified 65 instructors who taught at least one section of a course in an early morning time slot (start times before 10 AM) and at least one section of the same course during primetime (start times from 10 AM to 2 PM). We calculated average SECT_GPA and Q20 for each instructor's morning section(s) and primetime section(s) and then computed the mean difference. Both measures of learning were significantly lower in morning classes (SECT_GPA mean difference = -.423, t = 2.743, p = .002; Q20 mean difference = .237, t = 2.341, p = .016). We also identified 27 instructors who taught one section of a course in a MW or TH format and another section of the same course in a MWF format and calculated the mean difference between the two- and three-day formats for each instructor. For this group of instructors, the SECT_GPA was significantly lower for MWF classes (mean difference = -.135, t = 2.183, p = .045) but we found no significant difference in perceived learning (Q20).

Limitations & Call for Further Research

Effective capacity management is an important consideration for most service businesses; and, higher education institutions may find capacity management even more critical in the face of increased demand (enrolment) and tighter funding. Common sense demands that service providers constantly assess the impact of capacity management strategies on the quality of the service they provide. If a particular strategy employed in higher education, such as a space utilization policy or rule, has a negative impact on student learning and/or student satisfaction with their courses/instructors, administrators can and should modify the strategy to minimize those unintended consequences.

Our findings demonstrate that space utilization policies that increase class sizes, the number of early morning classes, the number of classes that meet three days per week and/or the number of classes that include a class meeting on Friday may have a negative impact on student learning. Perhaps the most important implication of our paper is that "one-size-fits-all" space utilization policies that require all departments to schedule a fixed percentage of courses outside of prime time or in a MWF time format (such as those cited in the introduction to this paper) may have a differential impact on student learning in certain levels of courses or in specific colleges or disciplines. To illustrate this problem, in the College of Business at the subject institution, 100% of the marketing courses studied were upper-level (junior or senior) courses, while only 50% of the economics classes in the same college were upper-level courses. If student grades are lowest in upper-level courses taught in early morning,

Table 1: Descriptive Statistics

		SECT_GPA		Q20	
Variables	**n**	**mean**	**SD**	**mean**	**SD**
COLLEGE:					
Business (BUS)	157	2.716	0.502	4.311	0.445
Education (EDU)	107	3.263	0.470	4.493	0.570
Liberal Arts (LAS)	485	2.827	0.551	4.491	0.447
Science & Technology (SCT)	322	2.596	0.631	4.248	0.607
COURSE LEVEL:					
Lower Division	453	2.531	0.643	4.335	0.524
Upper Division	618	2.947	0.544	4.456	0.505
CLASS SIZE:					
Extra Small (less than 19)	245	2.958	0.633	4.451	0.603
Small (19 to 25)	217	2.843	0.535	4.426	0.573
Medium (26 to 34)	198	2.865	0.607	4.344	0.524
Large (35 to 61)	200	2.698	0.587	4.412	0.435
Extra Large (62 and over)	209	2.527	0.485	4.313	0.435
CROWDING:					
Yes	222	2.753	0.538	4.458	0.374
No	849	2.793	0.605	4.375	0.556
TIME OF DAY:					
Morning (MORN)	182	2.569	0.575	4.339	0.528
Primetime (PRIME)	326	2.752	0.587	4.406	0.507
After Primetime (AFTERPRIME)	563	2.874	0.581	4.401	0.533
DAYS PER WEEK:					
1 day per week (DPW1)	220	3.046	0.552	4.426	0.534
2 days per week (DPW2)	641	2.762	0.568	4.366	0.541
3 days per week (DPW3)	210	2.580	0.610	4.434	0.455
FRIDAY CLASS MEETING:					
Yes	217	2.602	0.622	4.420	0.474
No	854	2.831	0.575	4.385	0.536
TOTALS	**1,071**	**2.785**	**0.592**	**4.392**	**0.534**

then a policy that requires each department to schedule 20% of its courses before 9:30 AM would have a bigger impact on marketing students than on economics students.

Our use of data from the SET instrument further suggests that class scheduling variables may have a negative impact on student satisfaction; and, in turn, have a negative impact on faculty promotion, tenure, and merit pay decisions. A quick examination of the correlations between course size and SET questions in our larger data set revealed significant negative correlations associated with ratings of instructor knowledge, instructor encouragement of participation, use of real-world examples, reasonableness of workload and more in larger classes, classes that meet in the early morning, and classes that meet three days per week. Clearly, faculty may also benefit greatly if administrators devote increased attention to analysing the impact of various capacity management strategies on student satisfaction with their courses.

Making space utilization policies more flexible could minimize the negative impact on important measures of educational quality. Using data that is readily available at most institutions, administrators could identify course levels, disciplines/departments, course types (e.g., lab v. lecture v. seminar), individual courses, or even individual instructors or groups of students that are least likely to experience negative effects of size (or other capacity management variables) on student learning. For example, at the subject institution, we found that the negative correlation between time-of-day and student learning was smallest in freshman-level courses. Adjusting the space utilization policy to increase the number of freshmen-level courses scheduled in the early morning would minimize the impact of time-of-day on student learning at the subject institution.

Our findings also have implications for higher education researchers, especially those studying SET or student learning. Many SET studies, especially the most recent studies we reviewed (e.g., Johnson, Narayanan and Sawaya 2013), control for the effects of course characteristics such as course size, course level, and discipline. Our finding of significant relationships between student learning and course crowding, the time of day the course meets, and the number of class meetings per week indicates the need to include additional course characteristics in future studies on educational outcomes (e.g., test results, grades, SET).

If nothing else, our study should motivate higher education researchers to take a closer look at the wealth of secondary data available to them in the records of their own institutions. Our study merged selected data from the university's schedule system (Banner) with summary student evaluation ratings for each course from the faculty database system (Digital Measures) and summary grade data from the Registrar. We limited this purely exploratory study to aggregate (course-level) data, but

consider the fact that the 1,071 courses we studied represented a total enrolment of 39,765 students; and, student-level data is also readily available in university records!

Future research could focus on isolating the impact of space utilization policies using time-series data to track changes in learning or SET before, during and after implementation of new capacity management policies. Researchers could also use hierarchical linear modelling to combine individual and course-level data to provide greater insight into the interactions between course characteristics (such as time of day), instructor characteristics (e.g., gender, experience), and student characteristics (SAT score, age, course grade). Researchers could apply the same methods to analyse the impact of other capacity management strategies such as reducing demand for classroom space by increasing the number of online classes or hiring more adjunct instructors to teach courses outside of prime time. Finally, merging data sets from different institutions may allow researchers to identify a set of "best practices" for capacity management in higher education.

References

Bandiera, O., V. Larcinese, and I. Rasul. 2010. Heterogeneous Class Size Effects: New Evidence From A Panel Of University Students*. *The Economic Journal* 120(549): 1365-1398.

Bedard, K., and P. Kuhn. 2008. Where Class Size Really Matters: Class Size And Student Ratings Of Instructor Effectiveness. *Economics of Education Review* 27: 253-265.

Bellante, D. M. 1972. A Summary Report Of Student Performance In Mass Lecture Classes Of Economics. *Journal of Economic Education (pre-1986)* 4(1): 53.

Brookes, M. 1985. Time In Learning: Does Distribution Make A Difference? *Community /Junior College Quarterly* 291-296.

Carrell, S. E., T. Maghakian, and J. E. West. 2011. A's From Zzzz's? The Causal Effect Of School Start Time On The Academic Achievement Of Adolescents. *American Economic Journal: Economic Policy* 3(3): 62-81.

Carrington, L. G. 2010. The Impact Of Course Scheduling On Student Success In Intermediate Accounting. *American Journal of Business Education* 3(4): 51.

Crittenden, K. S., J. L. Norr, and R. K. LeBailly. 1975. Size Of University Classes And Student Evaluation Of Teaching. *The Journal of Higher Education* 46(4): 461-470.

Dills, A. K., and R. Hernández-Julián. 2008. Course Scheduling And Academic Performance. *Economics of Education Review* 27(6): 646-654.

Edmonson, J. B., and F. J. Muldek. 1924. Size Of Class As A Factor In University Instruction. *Journal of Educational Reserch* 9(1): 1-12.

Feldman, K. A. 1984. Class Size And College Students' Evaluaion Of Teachers And Courses: A Closer Look. *Research in Higher Education* 21(1): 45-116.

Gallo, M. A., and M. Odu. 2009. Examining The Relationship Between Class Scheduling And Student Achievement In College Algebra. *Community College Review* 36(4): 299-325.

Gaubatz, N. 2003. Course Scheduling Formats And Their Impact On Student Learning. In *The National Teaching and Learning Forum*, (Vol. 12, No. 1).

Gibbs, G., L. Lucas, and V. Simonite. 1996. Class Size And Student Performance: 1984-94. *Studies in Higher Education* 21(3): 261-273.

Henebry, K. 1997. The Impact Of Class Schedule On Student Performance In A Financial Management Course. *The Journal of Education for Business* 73(2): 114-120.

Johnson, I. Y. 2010. Class Size And Student Performance At A Public Research University: A Cross-Classified Model. *Research in Higher Education* 51(8): 701-723.

Johnson, M. D., A. Narayanan., and W. J. Sawaya. 2013. Effects Of Course And Instructor Characteristics On Student Evaluation Of Teaching Across A College Of Engineering. *Journal of Engineering Education* 102(2): 289-318.

Kennedy, P. E., and J. J. Siegfried. 1997. Class Size And Achievement In Introductory Economics: Evidence From The TUCE III Data. *Economics of Education Review* 16(4), 385-394.

Klassen, K. J. and T. R. Rohleder. 2002. Demand And Capacity Management Decisions In Services: How They Impact On One Another. *International Journal of Operations & Production Management* 22(5): 527-548.

Kokkelenberg, E. C., M. Dillon, and S. M. Christy. 2008. The Effects Of Class Size On Student Grades At A Public University. *Economics of Education Review* 27(2): 221-233.

Ludlow, L. H. 1996. Instructor Evaluation Ratings: A Longitudinal Analysis. *Journal of Personnel Evaluation in Education* 10, 83-92.

Mateo, M. A., and J. Fernandez. 1996. Incidence Of Class Size On The Evaluation Of University Teaching Quality. *Educational and Psychological Measurement* 56: 771-778.

Monks, J., and R. M. Schmidt. 2011. The Impact Of Class Size On Outcomes In Higher Education. *The BE Journal of Economic Analysis & Policy* 11(1).

Raimondo, H. J., L. Esposito, and I. Gershenberg. 1990. Introductory Class Size And Student Performance In Intermediate Theory Courses. *The Journal of Economic Education* 21(4): 369-382.

Reardon, J., J. Payan, C. Miller, and J. Alexander. 2008. Optimal Class Length In Marketing Undergraduate Classes: An Examination Of Preference, Instructor Evaluations And Student Performance. *Journal of Marketing Education* 30(1): 12-20.

Smith, D. F., and B. K. Stephens. 2010. *Student outcomes and satisfaction: Does time of day matter?* Accessed February 15, 2013 http://dadsrolltopdesk.com/uploads/Smith_Stephens_Time_of_Day.doc.

TAMUCC 2010. Class Schedule Distribution Rules. Resource Document Texas A&M University – Corpus Christi. http://registrar.tamucc.edu/staff/assets/Fall-Spring%2520Standard%2520time%2520block%2520schedule.pdf&sa=U&ved=0ahUKEwih4a722KXMAhUFr4MKHUV-CLcQFggEMAA&client=internal-uds-cse&usg=AFQjCNH3RKN0yiSspBks1pqaW20MBTWy7Q.

UM 2012. Class And Classroom Scheduling Policy. Resource Document University of Michigan. http://www.provost.umich.edu/space/instruct/ClassClassroomSchedulingPolicy.pdf.

UNCC 2012. Class Scheduling Guidelines And Classroom Utilization. UNC Charlotte Academic Procedure. http://provost.uncc.edu/policies/class-scheduling-guidelines-classroom-utilization.

Westerlund, J. 2008. Class Size And Student Evaluation In Sweden. *Education Economics* 16(1): 19-28.

Wulff, D. H., J. D. Nyquist, and R. D. Abbott. 1987. Students' Perception Of Large Classes. *New Directions for Teaching and Learning* 1987(32): 17-30.

Rating Sport Business and Sport Human Development Journals: Transforming Standard Ratings to Capture Two-Dimensionality

Bhavneet Walia, Syracuse University
Shankar Ghimire, Western Illinois University

Abstract

There is a growing set of journals catering to the field of sport business, sport human development, or both (see. e.g., *NASSM* 2014). Among this set of journals, objective ratings are difficult to obtain given the paucity of *Journal Citation Reports* (*JCR*) impact factors available among this set of journals. We propose two transformations of the journal-level h5-index measure, each of which allows for a cardinal comparison as to relative journal impacts among a set of journals. We then compute and list journal ratings for the aforementioned field according to these transformed measures for the time period between 2011 and 2015. Given the paucity of *Journal Citation Reports* impact factor data for sport journals, there persists a need for information upon which to rate and rank journals in the field. However, the transformed impact measures presented herein can be applied to journals within any field.

Introduction

There is a growing set of journals catering to the field of sport business, sport human development, or both (see. e.g., *NASSM* 2014). Among this set of journals, objective ratings are difficult to obtain given the paucity of *Journal Citation Reports* (*JCR*) impact factors available among this set of journals (see, e.g., Shilbury and Rentschler 2007; *NASSM* 2014). As Shilbury and Rentschler (2007) note, sport business and related fields are quickly evolving. As such, several sport journals have emerged and grown in recent years. Given this evolution, the field would benefit from a journal rating methodology that, while empirically-valid and unbiased, has open-source flexibility and updatability.

Shilbury and Rentschler (2007) address this issue by constructing perceptual survey-based ratings of four leading sport journals, and Woratschek et al. (2009) follow a similar approach. Shilbury (2011) adopts a bibliometric (citation-count) analysis of four sport journals upon which the present article builds.[1]

These studies have been very valuable to the field in past years. To build upon this momentum in a sustainable manner, it may be further beneficial to develop easily-updateable relative and absolute impact factors from open-source data. Such work will allow researchers and editors to maintain current and objective indicators of journal impact within the field. Even if *JCR* impact factors become available for a larger set of sport journals, the annual release of these factors features a considerable time lag. Journal editors and researchers seeking a journal venue are typically responding to *JCR* citation data that was generated one or two years prior.

Given issues of unbiasedness, information availability, immediacy, and sustainability in rating a set of journals, the journal-level h5-index (Hirsch 2005; Braun et al. 2006) is potentially important in ranking journals. In their primary set of disciplinary journal listings, the *North American Society for Sport Management* (*NASSM*) lists and orders sport journals by h-5 index value (*NASSM*, 2014; NASSM 2015). Across academic disciplines in general, this index has generated a great deal of interest and support within the academic community (for a discussion of early use of the index in other fields, see, e.g., Harzing and Van Der Wal 2009, Braun et al. 2006, Vanclay 2006, Ball 2005). The journal-level h5-index is defined as follows:

> *A journal's h5-index value, h5, is the highest integer such that the following statement is true: The journal has published at least h articles (in the past five years) each of which subsequently garnered at least h citations.*

At the journal level, h5 values and h5 median values are listed and frequently updated within the Google Scholar Metrics database. In the event that a journal is not listed in this database, its h5 value can be obtained through time-corresponding article-level and citation-level searches within the broader Google Scholar database. As will be made clear in the following sections, however, the h5 and h5-median metrics are not complete measures in the sense that they do not directly allow for valid, cardinal (relative impact) comparisons of journal impact. With appropriate transformations, however, these metrics represent credible approaches to the absolute *and* relative rating of a set of journals.

Herein, we develop two new, straightforward methodologies by which to update some of the results and comparisons within Shilbury (2011). The present article seeks to obtain an understanding as to the present relative impacts of sport journals.

The study also seeks to develop a valid, open-source, and updatable metric for the rating and ranking of sport journals going forward.

Transforming h5 Values: A Baseline Approach to Rating Sport Journals

Let us formally define the h5-index of journal impact. Let $U_i = \{u_{i,1}, u_{i,2}, \dots, u_{i,n}\}$ be the set of all articles in journal i (published in the previous five years), and let $A_i(\subseteq U_i) = \{a_{i,1}, a_{i,2}, \dots, a_{i,h}\}$ be the set of articles in journal i (published in the previous five years) that constitute the h5 index of a given journal i. Then, it is tautological that $h5_i = |A_i|$. Moreover, let $B_i(\subseteq U_i) = \{b_{i,1}, b_{i,2}, \dots, b_{i,n-h}\}$ be the set of articles in journal i (published in the previous five years) that do not constitute the h5 index of a given journal i. Then, $c(a_{i,j})$ represents the set of articles citing $a_{i,j}$ and $|c(a_{i,j})| \geq |A_i| \; \forall \; a_{i,j} \in A_i$. As such, we can define $h5_i$ as follows.

Definition of h5-index: $h5_i = |A_i|$, where the elements of A_i are those that form the largest subset of U_i (maximize $|A_i|$) for which $|c(a_{i,j})| \geq |A_i| \; \forall \; a_{i,j} \in A_i$.

Unlike simple impact factors, the h5-index is a right-tailed measure of journal productivity. Rather than measuring a journal's average impact, it (by definition) measures the quality of a journal's most impactful set of recently published articles. Also in contrast to simple impact factors, the h5-index is a two-dimensional measure. Therefore, one cannot simply take the ratio of two journal-level h5-index values to assess the relative impact of two journals (as in the case of simple impact factors). To examine this point, let us consider two statements. The first statement represents a valid interpretation of simple impact factors:

> *The average impact of an article in journal x is 1.5 times that in journal y. According to average impact factor, therefore, journal x has 1.5 times the impact of journal y.*

However, the following similar statement regarding journal-level h5-index values would represent an invalid interpretation:

> *The h5-index value of journal x is 15, and the h5-index value of journal y is 10. According to the h5-index metric, therefore, journal x has 1.5 times the impact of journal y.*

The latter statement understates the relative impact of journal x. A simple ratio of h5-index values recognizes that the minimum number of citations in the best set of journal x is 1.5 times that of journal y. However, it does not recognize that there are more articles in journal x meeting this higher standard. That is, a simple ratio of h5-index values reflects only the threshold for h5-index inclusion without reflecting the larger minimum set of articles included. From the perspective of journal x, then, such a simple ratio is downward biased.

There are at least $15^2 = 225$ citations in journal x's h5-index set of articles, whereas there are at least 10^2=100 citations in journal y's h5-index set of articles. Therefore, the minimum (aggregate) citation count of journal x's h5-index set of articles is 2.25 times the minimum citation count of journal y's h5-index set of articles. If we are interested in measuring the aggregate (citation count based) impact of articles that meet the standard for h5-index inclusion, then it is important to transform the h5-index (e.g., by comparing the minimum number of citations in each journal's h5-index set of articles). Let us define the $(h5)^2$ index value.

Definition and De-composition of $(h5)^2$ index: $(h5_i)^2 = |A_i| \cdot |A_i| = c(a_{i,j})_{fl} \cdot |A_i|$, where $c(a_{i,j})_{fl}$ represents the article citation count floor for an article in U_i to also be in A_i. Hence, our baseline transformation of the h5-index metric simply squares each journal's h5-index value to obtain the minimum number of citations in a given journal's best set of articles.

The relative impact of two journals can then be compared cardinally by taking the ratio of $(h5)^2$ values for the respective journals. This ratio measures impact in terms of relative minimum number of citations of articles within each journal's (respective) h5 set of articles. In Table 1, we construct an $(h5)^2$ relative impact comparison of sport journals listed as comprehensive within NASSM (2015). The h5 values were obtained from Google Scholar (2011-2015) through Google Scholar Metrics journal searches. In the event that a journal was not searchable within the Google Scholar Metrics database, we collected citation data manually (i.e., on an article-by-article basis) via Google Scholar article searches. These searches were time-restricted to coincide with the methodology of the official Google Scholar h5 measure.

Table 1: $(h5)^2$ Relative Impact Ratings of Comprehensive Sport Journals

Journal/Rank	**h5**	**$(h5)^2$**	**$(h5)^2$ impact rel. to *JSM***
Journal of Sport Management (JSM), 1	29	841	1
Sport Management Review (SMR), 2	28	784	0.93
European Sport Management Quarterly, 3	21	441	0.52
International Journal of Sport Mgmt and Mkting, 4	12	144	0.17
Sport, Business, & Management: An Int'l Journal, 5	11	121	0.14
Journal of Applied Sport Management, 6	7	49	0.06
International Journal of Sport Management, 7	6	36	0.04
Sport Management International Journal, 8	5	25	0.03
International J of Developmental Sport Mgmt, 9	2	4	0.005

According to the $(h5)^2$ metric, the *Journal of Sport Management* is most impactful in the set, followed somewhat closely by *Sport Management Review*. In terms of $(h5)^2$ impact, the *SMR* is at an impact level that is roughly fourteen-fifteenths that of the *JSM*. That is to say, there are roughly fourteen-fifteenths the minimum number of journal citations within the h5 article set of *SMR* as compared to that of the *JSM*. The *European Sport Management Quarterly* is the only other comprehensive sport management journal possessing at least half the $(h5)^2$ value of either the *JSM* or the *SMR*. Relative to the remaining seven journals, the impact of each of the top two (or even three) journals is staggeringly high according to the $(h5)^2$ metric. If we sum the $(h5)^2$ values for the remaining seven journals (i.e., those ranked three through nine), this sum is less than the $(h5)^2$ value for the *Journal of Sport Management*. This indicates that the *Journal of Sport Management* possesses a higher minimum number of citations in its h5 best article set than do those seven journals ranked three through nine *combined*. We obtain the same result when comparing *Sport Management Review* to those journals ranked four through nine (or three through seven) and also for *European Sport Management Quarterly* in relation to those ranked four through nine. In other words, the $(h5)^2$ metric indicates that it is difficult to piece together impact within the fields of sport business and human development by publishing several articles in lower-ranked journals.

This relative impact dominance iterates as we move down the list. The fourth-ranked journal on the list dominates the combination of journals six through nine in terms of $(h5)^2$ value, as does the fifth-ranked journal. From these results, we conclude that there are several clear points of impact stratification among the nine journals studied.

Considering h5-median values: An Alternative Rating Approach

A problem with the $(h5)^2$ metric is that it provides only a lower bound on the number of citations within a given journal's h5-index article set. It does not count or even approximate the number of citations within said set. Assuming that the distribution of citation counts within each journal's h5-index article set is roughly symmetric (i.e., median near mean), one can take the product of each journal's h5 value and its h5-median value, ((h5-median) · h5), to approximate the number of citations in the journal's h5 article set.

Definition and De-composition of (h5-median) · h5 index:

(h5-median) · h5= $c(a_{i,j})\text{med} \cdot |A_i| \left[\approx \bar{c}(a_{i,j}) \cdot |A_i| = \frac{\sum_{i=1}^{h} c(a_{i,j})}{h} \cdot h = \sum_{i=1}^{h} c(a_{i,j})\right]$, where $c(a_{i,j})\text{med}$ represents the median article citation count among articles in A_i, $\bar{c}(a_{i,j})$ is the mean article citation count among articles in A_i, and the bracketed approximate equality holds under the assumption that the distribution of $c(a_{i,j})'s$ is roughly symmetric such that $c(a_{i,j})\text{med}$ is near $\bar{c}(a_{i,j})$.

Within this metric, the h5 value represents the number of articles in the best article set, and the h5-median value represents the median number of citations across articles within the set (where this median approximates the mean in the case that the distribution of citations by article is roughly symmetric). This measure allows for an approximation of best set citation counts without requiring the individual to perform exhaustive citation counts. Table 2 provides ((h5-median) · h5) index values for the same set of comprehensive sport journals. The h5 values were obtained from Google Scholar (2011-2015) via Google Scholar Metrics journal searches. In the event that a journal was not searchable within the Google Scholar Metrics database, we again collected (h5-median) citation data manually (i.e., on an article-by-article basis) via time-corresponding, Google Scholar article searches during June 2016.

Table 2: (h5-Median) · h5 Relative Impact Ratings Of Comprehensive Sport Journals

Journal/Rank	h5-median (2011-2015)	(h5-median) · h5 (2011-2015)	(h5-median) · h5 impact rel. to *JSM* (2011-2015)
Journal of Sport Management (JSM), 1	38	1102	1
Sport Management Review (SMR), 2	35	980	0.89
European Sport Management Quarterly, 3	30	630	0.57
International J of Sport Mgmt and Mkting, 4	19	228	0.21
Sport, Business, and Management, 5	19	209	0.19
Journal of Applied Sport Management, 6	9	63	0.06
International J of Sport Management, 7	7	42	0.04
Sport Management International Journal, 8	10	50	0.05
Int'l J of Developmental Sport Mgmt, 9	6	12	0.01

In the case of comprehensive sport journals, the ((h5-median) · h5) metric provides a similar account of relative journal impacts as does the $(h5)^2$ metric. The Pearson correlation coefficient between paired observations of the two measures is 0.998. The Spearman rank correlation coefficient between paired observations of the two measures is 0.983. For both metrics, the *Journal of Sport Management* and *Sport Management Review* lead all other journals substantially, and *European Sport Management Quarterly* uniquely resides in middle-ground. Moreover, two additional journals—*the International Journal of Sport Management and Marketing* and *Sport, Business, and Management: An International Journal*—rate substantially higher than the remaining four journals for both metrics. Among rated journals, we observe a very similar pattern of impact stratification in the ((h5-median) · h5) metric as was observed within the $(h5)^2$ results.

Conclusion

By transforming the h5 index metric of journal citation impact, we are able to provide a set of open-source metrics for the cardinal (relative) rating of journals within a given field. These alternative metrics are constructed, defined, and computed for the set of sport business and human development journals. Given the dearth of sport business and human development journals from which one can obtain a *JCR* impact factor, valid ratings of journal impact from open-source metrics are potentially quite valuable to the field. This approach can be applied to rate journals in any other field.

Notes

1 Ryan (2005), Hall (2011), and McKercher et al. (2006) each examine journal impact in the intersecting field of tourism and hospitality.

References

Ball, Philip. 2005. "Index Aims for Fair Ranking of Scientists." *Nature*, *436*(7053), 900-900.

Braun, Tibor, Wolfgang Glänzel, and Andras Schubert. 2006. "A Hirsch-Type Index for Journals." *Scientometrics*, *69*(1), 169-173.

Hall, Colin. M. (2011). "Publish and Perish? Bibliometric Analysis, Journal Ranking and the Assessment of Research Quality in Tourism." *Tourism Management*, *32*(1), 16-27.

Harzing, Anne-Wil and Ron van der Wal. 2009. "A Google Scholar h- Index for Journals: An Alternative Metric to Measure Journal Impact in Economics and Business." *Journal of the American Society for Information Science and Technology, 60(1),* 41-46.

Hirsch, Jorge. E. 2005. "An Index to Quantify an Individual's Scientific Research Output." *Proceedings of the National Academy of Sciences of the United States of America*, 16569-16572.

Mckercher, Bob, Rob Law, and Terry Lam. 2006. "Rating Tourism and Hospitality Journals." *Tourism Management*, *27*(6), 1235-1252.

North American Society for Sport Management (NASSM). (2015). Journals Serving the Community of Sport Management Scholarship. Available online at: https://goo.gl/Cux1jN

North American Society for Sport Management (NASSM). (2014). Journals Serving the Community of Sport Management Scholarship. Available online at: https://goo.gl/qWCZtn

Ryan, Chris. 2005. "The Ranking and Rating of Academics and Journals in Tourism Research." *Tourism Management*, *26*(5), 657-662.

Shilbury, David. 2011. "A Bibliometric Study of Citations to Sport Management and Marketing Journals." *Journal of Sport Management*, *25*(5), 423-444.

Shilbury, David, and Ruth Rentschler. 2007. "Assessing Sport Management Journals: A Multi-Dimensional Examination." *Sport Management Review*, *10*(1), 31-44.

Vanclay, Jerome K. 2008. "Ranking Forestry Journals Using the h-Index." *Journal of Informetrics*, *2*(4), 326-334.

Woratschek, Herbert, Guido Schafmeister, and Florian Schymetzki. 2009. "International Ranking of Sport Management Journals." *Wirtschaftswissenschaftliche Diskussionspapiere der Rechts-- und Wirtschaftswissenschaftlichen Fakultät an der Universität Bayreuth, Diskussionspapier*, *4*(09).

The Role of Service Industry in Development among African Countries

Shahdad Naghshpour, Alabama A&M University

Abstract

This study explores the possibility that international tourism could act as an engine of growth. The idea is based on export-oriented growth models utilizing the positive externality of export goods and their impact on the economy.
Keywords: International tourism, economic growth, export-oriented growth, positive externalities

Introduction

It could be said that the modern economic development era began by Harold (1939) and lasted till Solow (1956). For decades afterward no major theory was developed in the field while the theories of this golden age were tested and the ideas were tested and implemented as economic policies in many of the third world countries leaning towards the West. The failure of attempts to induce economic growth in Ghana after its independence in 1957 ended the hope of rapid and rather simple path to development. The early models used the economic theory of production and pointed out that production, and hence, the growth, depends on labor and capital. Implicit in this argument is a given area of land and a particular level of technology. The term technology in economics is used to define the way the factors of production are combined using the best available knowhow, which implicitly indicates a particular level of education and entrepreneurship.

During most of the Twentieth Century the growth bottleneck for underdeveloped countries was not labor; in fact, improvements to mechanized agriculture and healthcare created a labor surplus, especially in urban areas, albeit mostly lacking appropriate production skills. Therefore, shortage of capital was considered the hindrance to growth. Consequently, when Ghana gained its independence in 1957 many economists stepped in to bring economic growth to the country. The assistance, both financial and intellectual, poured in directly from Western countries as well as world organizations such as IMF and the World Bank. The efforts of the economists were backed by the United States, and international organizations and created a model, which was repeated by numerous African, Asian, and Latin American countries. Efforts to achieve economic growth by increasing investment failed. Import substitution was offered as an alternative to specialization in mineral extraction (Prebisch, 1964), which spread throughout the Latin American countries and numerous Asian countries (most notably India), and some African countries. Import substitution also depended on investment to achieve growth but the focus of the investment is on industries that result in consumer goods that were previously imported. To assure domestic companies could survive and prosper, tariffs and other means of import restrictions were implemented. However, instead of industrialization and growth the outcome was the enrichment of the few investors who did maximize their profits by producing low quality goods with no product support services. Once again economic resources were wasted without any noticeable economic growth. The next attempt was growth through export-oriented industrialization. The argument was that in order to succeed in exports the producers must be able to compete with manufacturers in other countries in quality and consumer appeal. True as this may be, there were no suggestions on how that competitiveness was supposed to be achieved. A half-hearted argument was the concept of learning-by-doing, which turned out to be the foundation of neocolonialism. In order to be competitive, domestic producers have to be innovative; in order to be innovative, producers must be rewarded, which is achieved through increased revenue from exports (Grossman and Helpman, 1991). As in all previous growth theory model; the question of how technology and innovation would occur was left unanswered. The not so implicit assumption was that having the correct political and economic system would somehow lead to innovation and new technology. The problem with growth theories of the mid nineteenth century was not that they identified capital shortage and lack of investment as the major bottleneck in the path of economic growth, rather, it was the inability to determine where the limited available capital should be invested and how to link a specific factory or investment project with other sectors and areas both vertically and horizontally. This concept has been identified in the growth pole theories.

Export-Oriented Growth

The beneficiary effects of trade have been known since Ricardo (1821). The possibility of exports leading to economic growth was considered by Chenery (1961), Chenery and Strout (1966), and Maizels (1968). There is no doubt that economic growth would not occur when there is not sufficient investment to exceed depreciation of the existing stock. There is equally no doubt that investment alone does not lead to economic growth as demonstrated by numerous countries in all continents. The

question seems to hinge on where and to what extent a country should invest to manage to increase output to initiate and finally to achieve economic growth. What sets investment in export-oriented industries is the claim that there are positive externalities in those industries (Naghshpour 2012).

Traditionally, implicit in any growth plans was the increase in manufacturing. Since the industrial revolution every developed country has achieved economic growth through substantial increase in manufacturing goods. One advantage of manufacturing is its compatibility with automation that can increase productivity, speed, and accuracy of production. The advent of computers has increased the speed of production substantially. Massive production of manufacturing goods requires a large enough market to absorb the output, which is not possible in most of small economies, thus, requiring the ability to export the surplus.

While automation and technology increases output and productivity it is not without its problems. One problem is pollution during production while the other is pollution due to obsolescence. Every country has experienced increased pollution in the process of industrialization. Although most of the industrialized countries have been taking steps to reduce pollution, currently, there seems little evidence of combating obsolescence in any country. It is ironic that some of the growth in GDP is due to forcing products out of the market deliberately, adding to pollution. For example, every generation of cellular phones requires a different jack for its charger effectively making the previous versions useless. Furthermore, many manufactured products such as TVs, rechargeable batteries and numerous other products contain toxins that end up in the dumps after the product malfunction as well as the product becomes obsolete as explained above.

Another byproduct of modern manufacturing is the decreased need for labor per unit of output, again caused by automation. Since 1776, the birth of the industrial age, the manufacturing output has increased while the need for labor per unit of output has decreased. During the same time the population of industrial countries have declined to the point that the birthrates of some of the industrial countries are not sufficient to maintain the existing population. On the other hand, the population of the world has increased substantially. In the early years of industrialization, the world population grew because of the increase in the population of the industrializing country, but by the middle of the nineteen century the increase has been accruing in the less developed countries. In both cases the population increase is attributed to better access to food and health care. In light of these two facts the path to economic growth may or may not be through manufacturing.

There is no doubt that mineral extraction will continue to be a main, if not necessarily the prime source of revenue for many of the currently underdeveloped countries for the foreseeable future. Equally evident is the fact that the share of manufacturing sector will remain high for the majority of currently underdeveloped countries for several reasons. One reason is the easily expandable output of the manufacturing sector using modern capital goods and relatively easy and inexpensive education. Another reason is the currently developed nations are gaining comparative advantages in developing new technologies, new products in the service industry in general, thus, forcing the currently less developed countries to experience comparative advantages in manufacturing and while maintaining comparative, if not necessarily absolute advantage in extraction of minerals. Unfortunately, the terms of trade have been gradually eroding for minerals as well as products of the agriculture sector; as noted by Prebisch (1964).

The export oriented growth of the 1970s was based on two bases; one was the fact that import-substitution was failing to materialize. The second was the positive externalities of export investment.

Service Oriented Growth

Equally obvious is that industrialization is not necessarily the path to economic growth for all countries. Some countries do not have the necessary comparative advantage in manufacturing to use exports as the engine of growth. This is especially the case in industries where there is economy of scale. In such industries, there is a substantial advantage for early arrivals who only need to continue growing rather than starting from scratch facing the taunting task of a large-scale production. Industries with increasing return to scale also are more practical in larger economies where the market is sufficiently large enough to assist with development and growth of the industry in its early stages without having to compete with other countries. The final reason is that some countries will be more suited to adapt new technologies in other areas as well as having comparative advantages in service industries.

The majority of service industry, such as retail sale and tourism, requires large number of low-skilled labor. However, most of the activities in the service industries have low demand and high income elasticities. In the case of the former, the demand will increase substantially in response to a decline in price and not only the prices for such services are low but the impact on growth are lacking. For example, although the demand for transportation is elastic it does not create new or additional products it just causes a redistribution of products, which may or may not be beneficial. When refrigerated trucks allow shipment of the fresh produce and fruits the standard of living of the poor in India declined because they could not have leftovers at the end of the day as was customary; such redistributions are counter-developmental. A major criticism of expanding service industry is that while it increases the utility of consumers it does not produce a tangible good. Agriculture products are sufficient for sustaining life. Manufacturing goods improve the production of agricultural products and facilitate their distribution, as well as

produce products that are directly consumable or increase productivity. However, service goods seldom are capable of doing either.

The consequence of high income elasticity for service goods is that their demand increases only when income increases, other things equal. The majority of the people in the third world countries are poor, which means their demands for many service goods are low. Usually, the ability of the small number of affluent people in the country to demand service goods is resented by the poor. In addition, many of the services and products that the rich people demand are satisfied by imported goods.

International Tourism

International tourism is a service good with interesting characteristics. It is an export good because its consumers are from abroad, although, the location of consumption is domestic. It is a source of foreign exchange, thus, functions as an export good. It has to compete with similar goods from other countries; therefore, it benefits from the same innovating and competitive requirements of other export goods. It has certain advantages as compared to other export oriented goods. The human capital requirement for the tourism industry is much lower than that of manufacturing goods. It is labor intensive and requires much smaller capital than manufacturing goods, especially the goods with economy of scale. Since international tourism is an export good it could possibly have positive externalities.

Literature Review

Obviously, any production would increase the gross domestic product (GDP), which potentially could yield development. Realistically, however, it is not the mere creation of a product, and hence income, that would result in economic development. The literature on the importance of tourism and its contribution to the economy abound, especially in tourism journals. The obvious arguments include the generation of tax revenues and employment. The contribution of international tourism toward foreign exchange is also a point often made. Many tourism studies focus on economic growth as perceived by residents (Belisle and Hoy, 1980), economic impact (Uysal and Giltelson, 1994), or as a measurement of the magnitude and extent of the revenue or other metrics; another form of economic impact study is at the level of a single activity, a city, or a region (Khan, Seng, and Cheong, 1990, West, 1993) as well as the economic impact of tourism for a country. Studies based on demand for tourism studies explain the demand for international tourism in the origin country (Archer, 1995; Bryden, 1973; Heng and Low, 1990). There is nothing a destination (country) can do to influence the demand from the origin country, thus, there is no policy value for economic growth. A notable exception is advertisement in the targeted origins. Many of the demand for tourism studies are survey-based and depend on a single equation using (estimated) expenditures by tourists or the number of arrivals as an explanatory variable. The latter is a poor indicator of expenditures, while the former suffers from exaggeration by tourists who would like to make their activity seem more valuable to improve and increase the availability of the attractions. Often, the main purpose of such studies is to assess the economic impact of tourism through the multiplier effect. The primary methodology of these studies is based on the recommendations of the Tourism Satellite Account. While an acceptable approach is not available among tourism-based studies there are ample studieson effectiveness of export-led growth that provide both theoretical and conceptual foundation for the problem at hand. Studies that focus on the supply rather than demand for exports are especially useful. International tourism is a special type of export from service industry.

Export-Oriented Growth

Chenery (1961) claims that investment in the export sector would provide the greatest boost to economic growth. Maizels (1968) extends the idea; Little, Scitovsky, and Scott (1970) and Balassa (1971) strengthened the theory by questioning the effectiveness of the import substation, which was the dominant policy recommendations of the time. Mizaels (1968) acknowledges that shortage of skilled labor and domestic savings as well as imports constraints and net borrowing hinder economic growth. However, he claims that savings shortcoming is a more immediate hindrance than that of the skilled labor. The importance of this observation is that instead of having a long list of factors the attention is focused on the main obstacle to growth; which makes it practical to provide policy recommendations. In other words, trade provides an extra dividend over and beyond any other investment; more technically, trade provides positive externality, which makes it an "engine of growth." Grossman and Helpman (1990) attribute this to the reward for innovation.

Krueger (1978) and Balassa (1978) add that trade not only increases the output, but also efficiency. Edwards (1993) provides a substantial list of studies supporting a causal link between trade and economic growth. Naghshpour and Sergi (2010) indicate that "[t]he claim is that not only trade causes a one-time gain in GDP, but it also changes the competitiveness of the exporting sectors that continues to increase growth." Chenery (1961) and Maizels (1968) state that successful trade requires

competitiveness that result in a dynamic advantage for trade as compared to the static benefits outlined in the comparative advantage theory of Ricardo (1821). The possibility that some investments could yield positive externalities is explicitly stated in the endogenous growth theory (Romer, 1986). An advantage of dynamic analysis is that it indicates causality (Michaely, 1977; Feder, 1983; Kormendi and Meguire, 1985; Fischer, 1991; and Harrison, 1996). Frankel and Romer (1999) demonstrate that trade has a low impact on income. Economic growth theories have evolved from acknowledging the importance of investment in growth to focusing on exports as the main policy instrument because of its positive externalities. Some of the studies linking trade and growth demonstrate low impact (Michaely, 1977; Feder, 1983; Kormendi and Meguire, 1985; Fischer 1991), and Harrison (1996) Frankel and Romer (1999); while others indicate the share of trade share from GDP is an endogenous variable (Helpman, 1988; Bradford and Chakwin, 1993; and Rodrik 1995).

Conclusion

At least there exist a possibility that a service good such as international tourism could act as an engine of growth similar to export-oriented manufacturing goods. The logical procedure would be to utilize a model that is based on supply of international tourism, accounts for externality, and is an extension of literature on export-oriented manufacturing models of economic growth.

References

Archer, Brian. 1995. "The Impact of International Tourism on the Economy of Bermuda. 1994." *Journal of Travel Research.* 34(2): 27-30.

Balassa, Bela. 1978. "Exports and Economic Growth, Further Evidence." *Journal of Development Economics* 5: 181-189.

Balassa, Bela. 1971. *The Structure of Protection in Developing Countries.* Baltimore: Johns Hopkins University Press.

Belisle, Francois, J. and Don R. Hoy. 1980. "The perceived impact of tourism by residents." *Annals of Tourism Research*. 8: 83–97.

Bradford, Collin I. and Naomi Chakwin. 1993. "Alternative Explanations of the Trade-Output Correlation in the East Asian Economics." *OECD Development Center Technical Paper No. 87*.

Bryden, John, M. 1973. *Tourism and Development: A Case Study of the Commonwealth Caribbean.* Cambridge: Cambridge University Press.

Chenery, H.B. 1961. "Comparative Advantage and Development Policy." *The American Economic Review* 51(1): 18-51.

Chenery, H. B., and A. Strout. 1966. "Foreign Assistance and Economic Development." *The American Economic Review* 4(1): 679-733.

Edwards, Sebastian. 1993. "Openness, Trade Liberalization, and Growth in Developing Countries." *Journal of Economic Literature* 31(3): 1358-1393.

Feder, Gershon. 1983. "On Exports and Economic Growth." *Journal of Development Economics* 12 (1/2): 59-73.

Fischer Stanley. 1991. "Growth, Macroeconomics, and Development." In *NBER Macroeconomics Annual 1991*, edited by Olivier Jean Blanchard and Stanley Fischer, 329-364. Cambridge, MA: MIT Press.

Frankel Jeffrey A, and David Romer. 1999. "Does Trade Cause Growth?" *The American Economic Review* 89(3): 379-399.

Harrod, Roy F. 1939. "An Essay in Dynamic Theory". The Economic Journal. 49 (193) 14–33.

Harrison, Ann. 1996. "Openness and Growth: A Time Series, Cross-Country Analysis for Developing Countries." *Journal of Development Economics*, 48(2): 419-427.

Helpman, Elhanan. 1988. *Growth, Technological Progress, and Trade*. Cambridge, MA: National Bureau of Economic Research. Reprint No. 1145. Empirica 15(1): 5-25.

Heng Toh Mun and Linda Low. 1990. "Economic Impact of Tourism in Singapore." Annals of Tourism Research. 17(2): 246-269.

Khan, Habibullah, Chou Fee Seng, and Wong Kwei Cheong. 1990. "Tourism multipliers effects on Singapore." *Annals of Tourism Research.* 17: 408–418.

Krueger, Anne O. 1978. *Foreign Trade Regimes and Economic Development: Liberalization Attempts and Consequences*. Cambridge, MA: Ballinger Press for the National Bureau of Economic Research.

Little, Ian, Tibor Scitovsky, and Maurice Scott. 1970. *Industry and Trade in Some Developing Countries.* London: Oxford University Press for Organization for Economic Cooperation and Development.

Maizels, Alfred. 1968. *Exports and Economic Growth of Developing Countries.* Cambridge: Cambridge University Press.

Michaely, Michael. 1977. "Exports and Growth: An Empirical Investigation." *Journal of Development Economics* 4(1): 49-53.

Naghshpour, Shahdad. 2012. "Are Exports an Engine of Growth?" *International Journal of Trade and Global Markets,* 5: 2, 153-166.

Prebisch, Raul. 1964. *Ttowards a New Trade Policy for Development*. New York: United Nations.

Ricardo, David. 2010, [1821]. *The Principles of Political Economy and Taxation,* CreateSpace.

Rodrik, Dani. 1995. "Getting Interventions Right: How South Korea and Taiwan Grew Rich." *Economic Policy* 10(20): 55-107.

Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94(5): 1002-1037.

Solow Robert, M. 1956. "A contribution to the theory of economic growth". *Quarterly Journal of Economics*. Oxford Journals. 70 (1): 65–94.

Uysal Muzaffer and Richard Giltelson 1994. "Assessment of Economic Impact: Festivals and Special Events." *Festival Management and Event Tourism*. 2(1): 3-10.