

FUSIO

The Bentley Undergraduate
Research Journal

Volume: I

Fall 2016

Issue: I



Integrative and multidisciplinary undergraduate research
at the leading edge of the Arts and Sciences and Business
Published by the Honors Program at Bentley University



BENTLEY
UNIVERSITY



The Bentley Undergraduate
Research Journal

EDITOR

AARON L. JACKSON – ECONOMICS

ASSOCIATE EDITORS

OTGONTSETSEG ERHEMJAMTS – FINANCE

JEFF GULATI – GLOBAL STUDIES

CHRISTIAN RUBIO – MODERN LANGUAGES

DAVID SZYMANSKI – NATURAL AND APPLIED SCIENCES

EXECUTIVE BOARD

MIRIAM BOERI – SOCIOLOGY

DEBORAH GREGORY – FINANCE

DANIELLE HARTIGAN – NATURAL AND APPLIED SCIENCES

FRED LEDLEY – CENTER FOR INTEGRATION OF SCIENCE AND INDUSTRY

RACHEL MADDEN '18 – ACTUARIAL SCIENCES

KAREN OSTERHELD – ACCOUNTING

EUTHEMIA STAVRULAKI – MANAGEMENT

KATIE VADAKIN '17 – FINANCE

MATEUSZ ZEGLÉN '17 – MATHEMATICAL SCIENCES

PRODUCTION AND EDITORIAL ASSISTANCE

MELISSA RUSH, MATTHEW DIMOND

BUSINESS AND EDITORIAL CORRESPONDENCE:

BENTLEY UNIVERSITY

HONORS PROGRAM

175 FOREST STREET

WALTHAM, MA 02452

FUSIO@BENTLEY.EDU

WWW.FUSIOJOURNAL.ORG

VOLUME I FALL 2016 ISSUE I

ARTICLES

EDITOR'S INTRODUCTION AARON L. JACKSON	3
DOING WELL AND DOING GOOD: PERFORMANCE OF ESG INTEGRATION APPROACHES GEORGE ILIOPOULOS AND OTGONTSETSEG ERHEMJAMTS	7
SOUTHWEST AIRLINES AND THE IMPACT OF LOW-COST CARRIERS ON AIRLINE TICKET PRICES SIDNEY FIELD	36
THE IMPACT OF APPEARANCE IN EVALUATING POLITICAL CANDIDATES JULIA T. PARADIS	58
THE SHORT-TERM EFFECT OF MONETARY POLICY ON FINANCIAL MARKETS BRIAN LEVINE	86
PATIENT-PROVIDER DISCUSSION ABOUT NUTRITION DURING ROUTINE VISITS: FREQUENCY, QUALITY, AND OUTCOMES KERRIANN FITZGERALD AND DANIELLE BLANCH-HARTIGAN	107

AIMS AND SCOPE

Established following Bentley University's mission of creating impactful knowledge within and across business and the arts and sciences, *Fusio* is a multidisciplinary undergraduate journal committed to the dissemination of original, high-quality undergraduate research. The journal is published by Bentley University's Honors Program and edited by both students and faculty across disciplines. *Fusio* encourages submissions from undergraduate students, with an emphasis on articles that span both business and arts and sciences topics as well as multidisciplinary topics. The journal is currently open only to undergraduate students at Bentley, and will consider original research by students as well as student/faculty joint work. All submissions undergo a blind peer review process.

Fusio is published twice yearly, in the fall and spring. For subscription information, please contact the Honors Program at Bentley University, or the journal office at fusio@bentley.edu. Online open access to these articles may also be found at www.fusiojournal.org.

Copyright © 2016 by Bentley University. Printed by Quad Graphics. Permission is granted to reproduce articles published in this journal, so long as reproduced copies are used for non-profit, educational, or research purposes. No responsibility for the views expressed by the authors in *Fusio* is assumed by the editors, by the publisher, or by Bentley University.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of our reviewers in supporting *Fusio* and the authors in ensuring the highest standards of academic quality. The authors have benefitted tremendously from their time and insightful comments, so we thank you.

EDITOR'S INTRODUCTION

By Aaron L. Jackson*

I. Introduction

I am pleased and delighted to usher in the inaugural edition of *Fusio: The Bentley Undergraduate Research Journal*. This journal has emerged out of growing recognition that student research has become a much more important part of the undergraduate student experience. Moreover, we have seen tremendous growth over the last few years in the quality, as well as quantity, of student research opportunities at Bentley University and expect this growth will continue.

Up until now, many projects by students are completed, and then shelved for eternity with little recourse for others to learn from the original discoveries of our bright and motivated students. As the Director of our Honors Program, I have witnessed this first hand. All students in the Program are required to complete a semester-long capstone research project under the supervision of a faculty member, yet most do not do anything with their research beyond completing the paper for graduation.

Over the last few years in particular, we have seen a sharp increase in the quality of those projects that our Honors Program students produce (commensurate with our ability to draw ever brighter students into Bentley each year)—to the point that many student research projects have exceptional originality and impactful results worthy of and necessitating dissemination to a wider audience in the finest traditions of the academy. This high quality threshold extends to our broader undergraduate student body and is not the exclusive purview of students in the Program. As a result, we have developed this journal as a permanent channel for all undergraduates to showcase these projects to the world, and broaden the knowledge base.

We have established *Fusio* following Bentley University's mission of creating impactful knowledge within and across business and the arts and sciences. *Fusio* is a multidisciplinary undergraduate journal committed to the dissemination of original, high-quality undergraduate research. The journal is published by Bentley University's Honors Program and edited by both students and faculty across disciplines; however as mentioned above it is not exclusive to students in the Program. We seek out high quality research from our undergraduates, whether it is sole-authored, co-authored with other students, or co-authored with faculty. *Fusio* encourages submissions with an emphasis on articles that span both business and arts and sciences topics, as well as multidisciplinary topics. It is the fusion of both of these elements which provides the Latin inspired name of the journal.

* Professor of economics, and Director, Honors Program, Bentley University. Email: ajackson@bentley.edu

WHAT IS FUSION?

Although Bentley is a business focused university, it is somewhat unique among its cohorts in stressing the integration of arts and sciences across the curriculum of mainly business focused degrees that its students predominantly seek out. But fusion has potentially more complex and subtle meanings in the context of higher education. For instance, Dan Everett, Dean of Arts and Sciences at Bentley University asserts that fusion “*is not in the [curriculum] per se; [fusion] is only effective if it occurs in the brains of the students,*” which is a function of what the institution, its programs, faculty, and students do as a whole. More specifically, as Everett suggests, “*fusion allows for students who can reason more effectively across a variety of intellectual and professional traditions to find solutions and identify problems that people without a ‘fusion experience’ would struggle much more to discover.*”¹

This perspective is indicative of the best practices of how our students learn, and we think, ultimately become better students and citizens of the world. It is for this reason that we have created the journal with the theme of fusion in mind: to encourage, and be reflective of, students thinking across a variety of intellectual and professional platforms in order to seek out new ways to think about the world around them – and by extension, make new discoveries. With this, as Everett suggests, “*fusion people are bilingual people: they can act, think, and speak effortlessly across and within the cultures and languages of businesses and the liberal arts.*”

Although the journal encourages the types of multidisciplinary, integrative research that fusion suggests, we still believe there is value in more traditional, disciplinary based research too, so do not wish to exclude exceptional, more traditional topical research. After all, the common denominator to *all* research—whether integrative and multidisciplinary, or largely within a field of study—is the invaluable self-discovery that comes to the student in the process, and the crucial skills that are developed that will last a lifetime.

WHY UNDERGRADUATE RESEARCH?

Student research has traditionally been relegated to the sciences, and predominantly the student’s role has been as a lab assistant, research assistant, or similar secondary roles. More universities, including Bentley, have come to understand that student research does not just happen in a lab, and students do have the capabilities to be principal investigators. With motivated students and faculty and a bit of preparation, undergraduates have the ability to do cutting-edge research in (and across) the spectrum of fields.

The benefits of student research are numerous and well documented: greater engagement by students; better critical thinking and problem solving skills; and of course the ability to think more creatively to name a few. More broadly, the process of research allows students to be discerning consumers of information, which can be crucially important in an increasingly data and technology driven society. In addition, having students involved

¹ Memorandum, August 13, 2015.

in their own research allows for the state of the art to be brought into the classroom, and therefore enhances the learning experience.

Ultimately, this journal will showcase student achievement and perhaps more importantly, set a high bar of achievement and possibility for academically curious students in order to reap the immense benefits that come to student researchers.

ABOUT THE FIRST ISSUE

Going forward, we expect a healthy mix of strong research projects from all corners of campus. However the inaugural issue of *Fusio* features some of our best research papers from our Honors Program seniors from the class of 2016.

Our first paper, “Doing Well and Doing Good: Performance of ESG Integration Approaches” succinctly captures the essence of integration of arts and sciences and business. Iliopoulos and Erhemjamts investigate the returns on stock portfolios of companies that place value on environmental, social, and governance (ESG) issues, and find that these portfolios tend to generate significantly higher returns than standard indexed portfolios. The implication of course is a stark and refreshing one: doing good (ESG), and doing well (shareholder returns) are not mutually exclusive.

The second featured paper also integrates elements of both sides of the academic house, but from a slightly different approach. “Southwest Airlines and the Impact of Low-Cost Carriers on Airline Ticket Prices” provides a valuable framework for thinking about the current structure of airline pricing and competition, but through an historical lens. With a complete historical picture of airline consolidation, and sound econometric treatment applied to updated data, Sidney Field concludes that the ‘Southwest Effect’ of low cost carriers reducing average fares did decline, but has recently re-emerged.

As of the release of this journal, perhaps one of the timeliest papers in the inaugural issue examines the impact of appearance in evaluating political candidates. Previous literature has suggested that, not surprisingly, looks can matter for a candidate’s political viability. The paper by Julia Paradis re-examines this question through the use of a carefully crafted experiment. Participants in the experiment were shown random picture groupings of actual (but perhaps not well known) political candidates, and rated them on attractiveness. Controlling for a number of factors, the author finds that contrary to the current orthodoxy, attractiveness may not matter that much, and in particular there may be less of a bias against minority and women candidates than previously thought. The implication is that appearance may matter much less than substance (although the current presidential election may be challenging this premise)!

Our fourth paper has important implications for how we view the effectiveness of policy, particularly that of monetary policy. Brian Levine examines the impacts of ‘surprise’ monetary policy changes on a variety of financial market indicators. The author uses sound econometric application to updated data, as well as an expanded scope of questions from a previous study. He concludes that not surprisingly, unexpected monetary policy changes do have impacts on a number of financial market indicators over the short-term. More

importantly however, the author finds that these impacts have declined in recent years, in large part due to improvements in communication by the Federal Reserve leading to lower volatility in financial markets.

The final paper in this issue also highlights an important, emerging area of policy concern. "Patient-Provider Discussion About Nutrition During Routine Visits: Frequency, Quality, and Outcomes" assesses the impact of communication in the doctor-patient relationship, and how it impacts the health of patients. Fitzgerald and Blanch-Hartigan use survey methodologies to examine how much and how often doctors discuss diet and nutrition with their patient, and show that more discussion lead to a greater intent to change behaviors by the patient, which has the important implication of potentially leading to improved health outcomes. This suggests that doctors need to spend more time discussing their health with patients.

Although the inaugural issue is highly reflective of the diversity of thought and topics explored by just some of our students, it is by no means exhaustive. There are too many unknowns in our world, and too much energy and academic curiosity of our students and faculty to maintain the intellectual status quo. I am incredibly excited to see what challenges our faculty put in front of our students, what unknown discoveries lie ahead, what depths of creativity will be plumbed, and by extension, what papers will be in future issues. Every issue and every paper will be a unique, engaging, and meaningful addition to our stock of human capital. I hope you share in the excitement and curiosity that *Fusio* provides to our student researchers: opportunity for meaningful engagement and contributions to the knowledge base that informs our citizenry.

DOING WELL AND DOING GOOD: PERFORMANCE OF ESG INTEGRATION APPROACHES

By George Iliopoulos & Otgontsetseg Erhemjamts*

A growing number of investors are interested in investing in companies with higher environmental, social, and governance (ESG) performance to create positive societal impact while making money at the same time. In this paper, we investigate the profitability of ESG investing strategies by incorporating various ESG criteria into conventional stock screens that are based on fundamentals. Starting with the Buffett: Hagstrom, Graham-Enterprising Investor, and Piotroski F-score screens, we create three baseline portfolios at the end of each year from 2003 to 2013, with stocks that meet the value and financial performance criteria specified in those screens. We then impose additional ESG criteria to further narrow down the set of stocks in each portfolio, and record the holding period returns of each portfolio for the following year. After incorporating combinations of various ESG criteria to form ESG integrated portfolios, we find that one could generate portfolio returns that are 2.7%–3.5% higher than the baseline portfolios, and 6.3%–11.7% higher than the S&P 500. Our results suggest that incorporating social ratings in community and employee relations areas improve baseline portfolio returns most consistently. This study implies that it is possible to earn competitive returns while achieving positive societal impact.

Keywords: Socially responsible investing; fundamental analysis; ESG integration.

I. Introduction

A growing number of investors are interested in investing in companies with higher levels of corporate social responsibility (CSR) to create positive societal impact while making money at the same time. Such strategies have a variety of labels such as socially responsible investing (SRI), socially conscious investing, ethical investing, morally responsible investing (MRI), sustainable investing, impact investing, mission investing, etc. Modern SRI gained global traction in the 1980s with the movement to divest investments from South Africa in protest to its system of racial segregation known as Apartheid.¹ Then,

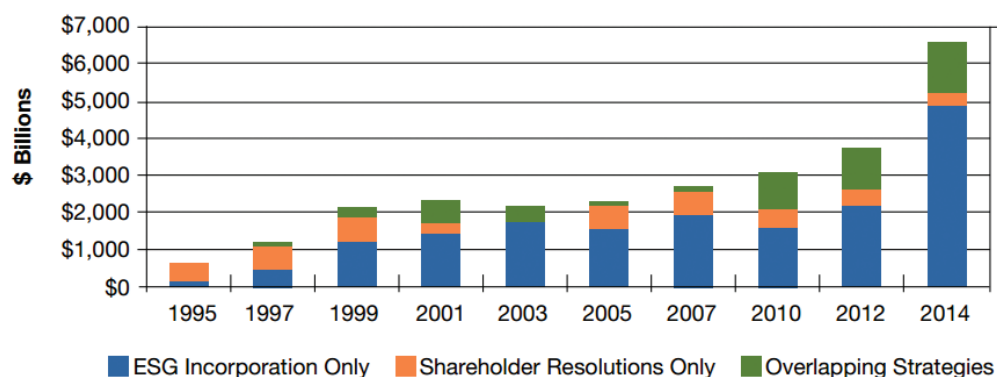
*We thank the Bentley University Honors Program for making this research possible. We are grateful for the useful comments from participants at the 13th Annual Honors Conference at Bentley University, three anonymous referees, and Jeff Gulati (the associate editor). Iliopoulos: Valuation Associate, Grant Thornton. Erhemjamts: Associate Professor of Finance, Bentley University; Email: oerhemjamts@bentley.edu.

¹SRI has ancient origins in Jewish, Christian, and Islamic traditions. The Catholic Church imposed a universal prohibition on usury in 1139, which had not been relaxed until the 19th century. In England, a law called “The Act Against Usury”, which prohibited excessive interests on loans was in effect from 1571 to 1624. In the 17th century, the Quakers refused to profit from the weapons and slaves trade when they settled in North America (Renneboog et al., 2008a). In contrast to ancient ethical investing which was mainly faith-based, modern SRI is more based on the varying personal ethical and social convictions of individual investors.

with the Bhopal, Chernobyl, Exxon Valdez oil spills, as well as the more recent Deepwater Horizon spill in the Gulf of Mexico and Fukushima nuclear disaster in Japan, the environment became the top concern for socially conscious investors. In the 17 years between the first Trends Report in 1995 and the most recent report in 2014 by the US SIF foundation, responsibly managed asset pools have grown tenfold, or 929%, from \$639 billion to over \$6.57 trillion (See Figure 1). These assets now account for more than one out of every six dollars under professional management in the United States. Between 2012 and 2014 alone, responsibly invested assets grew 76%. Joe Keefe, chief executive of Pax World Management, which launched its socially-responsible Pax World Balanced Individual Fund (Ticker: PAXWX) in 1971, was quoted as saying: “I’ve seen more interest in the past two years than I’ve seen in the past 17 years”. He attributed this recent growth to women and millennials. Women and millennials² are more interested in making the world better, and that sentiment is believed to carry through to their investment choices.

In this paper, we investigate the profitability of SRI strategies by incorporating various environmental, social, and governance (ESG) criteria from the MSCI ESG KLD STATS (KLD) database into conventional stock screens that are based on fundamentals.³ Our goal is to examine whether one can achieve competitive financial returns while making a positive impact on society. By incorporating various ESG performance indicators into widely-used value screens for approximately 2,400 largest U.S. companies over 2003-2013 period, we find that one could generate portfolio returns that are 2.7%–3.5% higher than the baseline portfolios, and 6.3%–11.7% higher than the S&P 500. Our results suggest that incorporating social ratings in community and employee relations areas improve baseline portfolio returns most consistently.

FIGURE 1
SRI Investing in the United States, 1995 - 2014



SOURCE: US SIF Foundation.

² <http://www.investmentnews.com/article/20160306/FREE/160309960/socially-responsible-investing-is-coming-of-age>

³ Fundamentals analysis, in accounting and finance, is the analysis of a company’s financial statements. Examples of fundamentals include debt ratio, free cash flow, operating margin, return on equity, earnings, growth, revenue growth, etc.

II. Incorporating ESG Criteria Into Socially Responsible Investments

The US SIF Foundation reports that money managers engage in SRI strategies via the use of (i) negative or exclusionary screening (screening out companies based on specific ESG criteria), (ii) positive or best-in-class screening (screening in companies with positive ESG performance relative to peers), (iii) ESG integration (incorporating ESG criteria into traditional financial analysis), (iv) impact investing (targeted investments, aimed at solving social or environmental problems), or (v) sustainability themed investing (selection of assets specifically related to sustainability). The five categories are not mutually exclusive, as more than one approach can be used with the same investment vehicle. In US SIF Foundation's 2014 survey, more than half of the US money managers reported using a negative/exclusionary strategy. For example, the \$2.3 billion Neuberger Berman Socially Responsive Fund (NRAAX) first eliminates companies that receive significant amounts of revenues from controversial business areas such as tobacco, alcohol, gambling, weapons, and nuclear power. It then continues to integrate ESG criteria such as good community, workplace, and environmental records to form a focused portfolio of 30 to 40 stocks. Similarly, the \$12.6 billion Parnassus Equity Income fund (PRBLX) – the largest socially responsible fund – also eliminates companies that receive a significant amount of revenues from controversial business areas. In addition, it eliminates companies that have direct involvement in Sudan. This screen was added in 2006 when the international community recognized the Darfur region conflict as genocide. Both funds have impressive 10- and 15-year records, beating 97%–98% of their peers.

Funds that employ negative/exclusionary practices also include funds such as the \$1.5 billion Ave Maria Funds (which consists of six mutual funds) that are geared toward Catholic/Christian investors. As such, it uses religious screens to eliminate businesses that are involved in embryonic stem cell research, ruling out most drug companies, biotech firms, hospitals, and health insurers. It also screens out companies that are related to pornography, including media and cable companies, as well as any hotel chains that offer adult films on pay-per-view. Its five stock funds, including the flagship Ave Maria Catholic Values (AVEMX) consistently land in the bottom half or quartile when compared to their non-faith-based peers. There are also mutual funds that abide by Islamic laws (e.g., Amana Trust Growth fund, Azzad Ethical Mid Cap fund, Iman fund, Al-Yusr US Equity fund, etc.). Islamic law calls for an avoidance of companies involved in alcohol, tobacco, and pork products, and broadly excludes investments in companies whose earnings are derived from charging interest (usury).

According to US SIF Foundation, recent developments in negative/exclusionary screening are the emergence of the fossil fuel restriction/divestment movement due to large scale oil spills and nuclear disasters, and the resurgence of policies restricting investments in firearms due to mass shootings in Sandy Hook Elementary School in Newton, CT; Virginia Tech University in Blacksburg, VA; a Century 16 movie theater in Aurora, CO; etc. While some funds eliminate stocks of fossil fuel companies entirely from their portfolios,

others seek to invest in companies that are “best in class” in the industry. For example, the Green Century Balanced Fund (GCBLX) from Trillium Asset Management, one of the largest SRI firms in the US, keeps their investors’ dollars out of fossil fuel companies and instead seeks to invest in sustainable companies and environmental innovators. The Fund also avoids factory farming, genetically modified organisms (GMOs), nuclear energy and tobacco.

ESG integration is the systematic inclusion of ESG indicators into traditional financial analysis. Among all the different SRI strategies, ESG integration approach has experienced the most explosive growth. According to US SIF Foundation, the single most important factor in the growth of SRI funds over the 2012-2014 period was the growth in ESG integration strategy.⁴ Pax World Management (<http://paxworld.com/>) was the first firm to launch a publicly available mutual fund that used social as well as financial criteria to make investment decisions (See Figure 2). Pax World seeks to identify companies that are leaders in their industries, are better managed and are more forward-thinking, are better at anticipating and mitigating risk, meet positive standards of corporate responsibility, and companies that are focused on the long term. The company now has 34 funds that follow an integration strategy and often outperform their benchmarks.

FIGURE 2
PAX World Investments – ESG Integration Strategy



⁴ Fundamental analysis, in accounting and finance, is the analysis of a company’s financial statements. Examples of fundamentals include debt ratio, free cash flow, operating margin, return on equity, earnings growth, revenue growth, etc

III. Performance of SRI Funds and Strategies

There is a large body of literature showing how environmental, social, and governance (ESG) factors impact firm performance. In terms of the corporate environmental performance, Guenster, Bauer, Derwall, and Koedjik (2011) find that eco-efficiency scores produced by Innovest Strategic Value Advisors have a positive relationship with sample firms' operating performance. Using the same Innovest ratings, Aktas, De Bodt, and Cousin (2011) show that the stock market rewards acquirers for making investments in socially and environmentally responsible targets. An increase in the target rating by one unit (over a seven-unit scale), leads to an abnormal gain of 0.9% for acquirer shareholders. For an acquirer worth \$100 million in equity, this represents a dollar gain of \$0.9 million. Dowell, Hart, and Yeung (2000) show that U.S. multinational corporations adopting stringent global environmental standards tend to have higher price-to-book ratios than companies adopting local environmental standards. In particular, average Tobin's Q ratio (ratio of market value of the firm to book value of tangible assets) for firms that adopt stringent global standards is 4.1, which is significantly higher than that for firms that adopt local standards, which is 2.2. Similarly, Russo and Fouts (1997) find that companies with better environmental records have higher return on assets (ROA). Environmental ratings in Russo and Fouts are from Franklin Research and Development Corporation (FRDC), and they range from 1 to 5, higher values reflecting better environmental performance. Their OLS regression results show that 1 unit increase in the environmental rating leads to 1.5% increase in ROA.

As for the social component, Verwijmeren and Derwall (2010) find that firms with highest employee well-being scores have lower debt ratios, and better credit ratings. More specifically, each extra point on the employee well-being scale lowers the debt ratio by 0.015. The median credit rating for firms with positive scores for employee well-being is BBB, while it is BB for firms with non-positive employee well-being scores. Statman and Glushkov (2009) find that stocks of companies with high ratings on social responsibility characteristics outperformed companies with low ratings. They also find that "shunned" or "sin" stocks (i.e., stocks of companies that are related to controversial business issues such as alcohol, tobacco, gambling, firearms, military, and nuclear) outperformed those in other industries. As a result, the two effects balance out, so that socially responsible indexes have returns that are approximately equal to those of conventional indexes. Edmans (2011) reports that a value-weighted portfolio of the Forbes' "100 Best Companies to Work For in America" outperform industry benchmarks by 2.1%, even after accounting for firm characteristics, removal of outliers, and different weighting methodologies.

Finally, the governance aspect is also shown to have a significant impact on firm performance as well. Barber (2007) reports that CalPERS' corporate governance initiatives created around \$3.1 billion in shareholder wealth between 1992 and 2005. These gains to the activism of CalPERS focus-list firms were based on short-run analysis, indicating small, but positive, market reactions of 23 basis points on the date focus-list firms are publicly announced. Gompers, Ishii, and Metrick (2003) construct a governance index to proxy for the shareholder rights at about 1,500 large firms during the 1990s. An investment

strategy that bought firms in the lowest decile of the index (strongest rights) and sold firms in the highest decile of the index (weakest rights) would have earned abnormal returns of 8.5% per year during the sample period. Gompers et al. (2003) also find that firms with stronger shareholder rights had higher firm value, higher profits, higher sales growth, lower capital expenditures, and made fewer corporate acquisitions.

Despite the evidence that firms with positive ESG attributes may be financially rewarded, SRI funds tend to be concentrated and focused, due to their strict, non-financial selection criteria. Therefore, SRI portfolios might be expected to underperform conventional portfolios as the investment opportunity set is constrained and a mean–variance efficient portfolio may not be achievable. According to the *underperformance hypothesis*, SRI funds are expected to generate weaker financial performance than conventional funds for the following reason: SRI funds underinvest in financially attractive investment opportunities as some of these opportunities are excluded from the investment universe due to the fact that they do not sufficiently contribute to the SRI objectives of the funds. In other words, ESG screens that limit the full diversification potential may shift the mean-variance frontier towards less favorable risk-return tradeoffs than those of conventional portfolios. For example, excluding stocks by using negative screening (i.e., excluding firms with ESG concerns, or involvement in controversial business practices such as tobacco, alcohol, gambling, etc.) might result in elimination of risky stocks with high expected returns, which is consistent with the *underperformance hypothesis*.

In contrast, *outperformance hypothesis* suggests that by screening on ESG criteria, SRI funds eliminate poor performing funds from their portfolios, and therefore can outperform non-SRI funds. The *outperformance hypothesis* is clearly at odds with the efficient market hypothesis. A key assumption underlying the *outperformance hypothesis* is that stock markets misprice information on CSR in the short run that SRI funds may outperform conventional funds in the long run. If SRI screening processes do generate value-relevant information (e.g., sound social and environmental performance can signal high managerial quality, which can translate into favorable financial performance), conventional portfolio managers could replicate the screens and the performance edge of SRI over conventional investments should diminish. Edmans (2011) finds that employee satisfaction is not fully valued by the stock market, and SRI screen that incorporates employee satisfaction may improve investment returns.

In 2013, Harvard University rejected student demands to divest fossil fuels from its endowment fund based on an argument consistent with the *underperformance hypothesis*, even though Harvard has divested from tobacco, Sudan, and South Africa in the past. Drew Faust, the university's president, said that it came down to dollars: "Despite some assertions to the contrary, logic and experience indicate that barring investment in a major, integral sector of the global economy would – especially for a large endowment reliant on sophisticated investment techniques, pooled funds and broad diversification – come at a substantial economic cost."⁵ However, others believe that is not necessarily true. Hugh Lawson, the global head of ESG investing at Goldman Sachs, said that return profile doesn't have to be

⁵ <http://www.harvard.edu/president/news/2013/fossil-fuel-divestment-statement>

compromised. Oil and gas investments, for example, can be replaced by other holdings in the energy sector, commodities or real estate that correlate with fossil fuel assets.⁶

While the literature on performance of SRI funds is growing, results are far from conclusive. Many studies report that the returns of SRI funds do not appear to be significantly different from conventional fund returns on a risk-adjusted basis. Hamilton, Jo, and Statman (1993), Goldreyer, Ahmed, and Diltz (1999), Statman (2000), and Bello (2005) compare the returns of US ethical funds with those of US conventional funds. Using Jensen's alpha, Sharpe, and Treynor ratios, they find that returns of SRI funds are not significantly different from those of non-SRI funds. In a sample of German, UK and US ethical mutual funds, Bauer, Koedjik, and Otten (2005) also find no evidence of significant differences in risk-adjusted returns between ethical and conventional funds.

However, some studies report that performances of SRI and non-SRI funds can differ. Some funds are much more stringent in their selection of firms than others. By pooling all SRI funds into the same general category, comparison between SRI and non-SRI fund performances is not straightforward. For example, Nofsinger and Varma (2014) find that socially responsible mutual funds underperform conventional mutual funds during non-crisis periods, but outperform during periods of market crises. This asymmetric return pattern is pronounced in funds that use positive screening techniques. Benson and Humphrey (2008) suggest SRI fund flows are less sensitive to returns than conventional funds. They also show that SRI investors are more likely to reinvest in a fund they already own relative to conventional investors. Given that the non-financial criteria differ significantly across SRI funds, it is difficult for SRI investors to find an alternative fund that will exactly meet their non-financial goals.

Goldreyer et al. (1999) report that within SRI funds, funds that employ inclusion (positive) screens outperform funds that do not employ such screening. Renneboog, Ter Horst, and Zhang (2008b) find that the screening activities and processes of SRI funds have a significant impact on the risk-adjusted returns. Funds adopting a community involvement policy or employing an in-house SRI research team to screen portfolios have better returns than SRI funds without such processes policies. Fund returns decrease with screening intensity on social and corporate governance criteria (proxied by the number of social and governance screens applied). This is also consistent with the *underperformance hypothesis* of SRI funds stating that high SRI screening intensity constrains the risk-return optimization and does not help fund managers to pick underpriced stocks.

Barnett and Salomon (2003) use the total number of social screens used by a mutual fund, as reported in the *Social Investment Forum*, as their main measure of social performance. According to this data source, there are 12 possible social screens associated with socially responsible mutual funds: alcohol, tobacco, gambling, weapons, animal testing, products or services, environment, human rights, labor relations, equal employment, community services, and community relations. Controlling for a variety of factors, Barnett and Salomon find the relationship between social and financial performance to be curvilinear.

⁶ <http://www.theguardian.com/sustainable-business/2015/sep/16/goldman-sachs-morgan-stanley-merrill-lynch-fossil-fuel-divestment>

That is, financial performance declined as the number of social screens increased, until it reached a low point around 6 (of 12) screens, then turned back up, increasing as the number of screens increased. This curvilinear finding suggests that two competing viewpoints (i.e., *underperformance vs. outperformance*) in the SRI literature may be complementary, and that both may be right to varying degrees.

Some of the above mentioned studies also compare performance of SRI funds against benchmark indices. For example, Statman (2000) and Bello (2005) compare the returns of US ethical funds with those of Domini 400 and S&P 500 indices.⁷ Statman's sample includes 64 socially conscious funds from Morningstar as of September 1998, while Bello's sample includes 42 socially responsible funds from Morningstar as of March 2001. Their results show that US ethical funds underperform the Domini 400 Social Index and the S&P 500 (however, no worse than conventional mutual funds). Similarly, Renneboog et al. (2008b) find that SRI funds in the US, the UK, and most continental European and Asia-Pacific countries strongly underperform their Fama-French-Carhart (FFC) benchmarks. In particular, the risk-adjusted returns of the average SRI funds in Belgium, Canada, France, Ireland, Japan, Netherlands, Norway, Singapore, and Sweden are between -4% and -6% per annum. The FFC-adjusted alphas of UK and US SRI funds are -2.2% and -3.4%, respectively. Their data sources for SRI funds are Standard & Poor's Fund Service, CRSP Survivor-bias Free Mutual Fund Database, and Bloomberg. In contrast, Dah, Hoque, and Wang (2015) find that Islamic mutual funds do not under-perform the broader benchmarks (e.g., US Sustainability Index, S&P 500, and value-weighted return on all NYSE, AMEX, and NASDAQ stocks for the US funds; Tadawul All Share TASI Index for Saudi Arabia funds; Bursa Malaysia KLCI Index for Malaysian funds, and Kuwait Stock Exchange Index KWSE for Kuwaiti funds). Their sample includes four Islamic mutual funds in the U.S., five Islamic mutual funds in Kuwait, 13 Islamic funds in Malaysia, and 23 Islamic funds in Saudi Arabia.

Some studies examine performance of SRI strategies more directly by forming SRI portfolios by using ESG criteria. Results are mixed here as well. Anderson and Myers (2007) compare the performance of four groups of portfolios from S&P 500 firms based on ESG social screens with the Russell 3000 stock index for the 1991-2004 time period. The authors create the following four portfolios using combinations of screens: (i) White – all firms with at least 2 strengths and none of the six exclusionary concerns (41-96 firms), (ii) Beige – all with strengths (73-240 firms), (iii) Brown – all with concerns (279-470 firms), and (iv) Black – all with any of the six exclusionary concerns. Through either value or equally weighted portfolios, Anderson and Myers (2007) find no statistically significant difference among SRI screen portfolios and the market benchmark. Kempf (2007) implements a trading strategy where stocks with high CSR ratings are bought and stocks with low CSR ratings are sold. He finds that this strategy leads to abnormal returns of up to 8.7% per year. The maximum abnormal returns are reached when investors employ the best-in-class screening approach, use a combination of several socially responsible screens at the

⁷ Domini 400 is a capitalization-weighted index of stocks of socially responsible companies that was initiated in May 1990 by Kinder, Lydenberg, Domini & Company.

same time, and restrict themselves to stocks with extreme socially responsible ratings.

In this study, we follow the footsteps of Andersen and Myers (2007) and Kempf (2007), and examine whether one can outperform the market (i.e., S&P 500) by creating ESG integrated portfolios. We start with baseline portfolios that follow well-known and proven to be successful stock screens that use financial criteria only, and then add ESG criteria. Due to the importance of ESG ratings for this approach, we use ESG ratings from MSCI ESG KLD STATS database over 2003-2013 period. Financial information is from Compustat North America Fundamentals Annual database, and conventional stock screen information is from the Association of American Individual Investors website (www.aaii.org). Following sections will describe our data sources, key variables, as well as our methodology in forming ESG integrated portfolios.

IV. ESG Ratings From MSCI ESG KLD STATS Database

Morgan Stanley Capital International (MSCI), an independent provider of research tools for institutional investors, releases the MSCI ESG KLD STATS (KLD) ratings, which are an annual set of positive and negative environmental, social, and governance (ESG) performance indicators for approximately 2,400 of the largest U.S. companies.⁸ Companies that score well on the list are included in global indexes including the MSCI KLD 400 (formerly known as Domini 400), which has become widely recognized as the standard of SRI performance measurement for socially responsible investors. The MSCI KLD 400 index has closely tracked the S&P 500 index over the past decade with returns of 10.77% from 2003 to 2013 compared to the S&P 500's 10.97%.

The rating system uses over 60 indicators in 7 categories which all fall under the three pillars: environmental (E), social (S), and governance (G), which is commonly referred to as ESG. Since the number of possible strength and concern indicators varies from year to year, we present 2013 set of indicators in Appendix A. ESG performance indicators in KLD are scored by a simple binary scoring model: If a company meets the assessment criteria established for an indicator, then this is signified with a "1". If a company does not meet the assessment criteria established for an indicator, then this is signified with a "0". If a company has not been researched for a particular ESG indicator, then this is signified with a "NR" (Not Researched). These 60 plus indicators can be used to calculate three aggregate CSR scores to rank each company: *All Strengths* score, *All Concerns* score, and the *Net Strengths Score* which is *All Strengths* minus *All Concerns* (also known as the *ESG index*). *All Strengths* score is a sum of all strengths indicators, and *All Concerns* score is a sum of all concerns indicators. Investors that look to invest in socially responsible companies can look at these measures in order to aid their investment decisions. More specifically, one's investment strategy can be based on picking companies with high *All Strengths* scores, low *All Concerns* scores, or a high *ESG Index*. Downside of using a net score such as *ESG*

⁸ The MSCI ESG KLD STATS database was originally created by Kinder, Lydenberg, and Domini Research & Analytics, Inc. (KLD) in 1991. Founded in 1988, the firm was acquired by RiskMetrics in November 2009, and RiskMetrics itself was acquired by MSCI in March 2010.

Index without examining *All Strengths* and *All Concerns* scores separately is that the heterogeneity between firms is lost (Erhemjamts, Li, and Venkateswaran, 2013). For example, a firm with ten strengths and ten concerns will have an *ESG Index* of zero, as well as a firm with two strengths and two concerns.

In addition to raw *All Strengths* and *All Concerns* scores, we calculate *Adjusted All Strengths* and *Adjusted All Concerns* scores following Cornett, Erhemjamts, and Tehranian (2016). The reason for adjusting the raw scores is that the number of possible strengths and concerns varies from year to year. Every year, some new indicators are introduced and/or other indicators get discontinued. That can result in mechanical increase or decrease in strength and concern scores without any change at the company level CSR activities. Therefore, we scale the score for each category (community, environment, etc.) by the maximum value for that category in a given year. We then add up all adjusted strengths and concerns scores across all categories to calculate *Adjusted All Strengths* and *Adjusted All Concerns*. In order to calculate *Adjusted ESG Index*, we first find adjusted net score for each category, and then add up the adjusted net scores. We then utilize *Adjusted All Strengths*, and *Adjusted All Concerns* scores in addition to the *Adjusted ESG Index* in our portfolio formation process.

V. Conventional Stock Screening Techniques

Screening is an application of quantitative criteria to a broad universe of stocks in order to narrow the list down to a few companies. It allows you to focus your attention on a smaller but more promising group of stocks. It also forces you to use a consistent framework to decide which stocks to add or remove from your portfolio. Over the years a handful of investors have become well known for their skills at consistently identifying portfolios of stocks that “beat the market.” These include such well recognized names as Warren Buffett, Peter Lynch, David Dreman, Martin Zweig, John Neff, and William O’Neal. The interest of individual investors in trying to duplicate the performance of these successful investors is extremely high, which is evidenced by the overwhelming number of books on investing and their popularity.⁹

In order to implement our ESG integration approach, we utilize some of the most well-known stock screens identified and offered by the American Association of Individual Investors (AAII), a nonprofit investment education organization based in Chicago, IL. Through their website (www.aaii.org) and print publications, AAII attempts to educate individual investors and offers a number of tools to help investors build wealth. AAII tracks the performance of over 60 screens by creating an imaginary portfolio that buys and sells stocks that meet the screen’s criteria on a monthly basis. At the end of each month, stocks in the portfolio that no longer meet the criteria are sold and new stocks that meet the criteria

⁹ Examples of best sellers include *The Intelligent Investor* by Benjamin Graham, *How to Make Money in Stocks* by William O’Neil, *One Up on Wall Street* by Peter Lynch, *The Little Book of Common Sense Investing* by John Bogle, *The Little Book that Beats the Market* by Joel Greenblatt, and *Common Stocks and Uncommon Profits* by Philip Fisher.

are purchased. Most of the screens identified on the AAI website are focused mainly on value, growth, or momentum strategies. Value investing consists of buying unappreciated or ignored stocks at attractive prices. Value investors seek stocks that are priced attractively relative to some measure of intrinsic worth. For example, they look for stocks selling at temporarily low price-to-book, price-to-cash-flow, price-to-earnings, or price-to-sales multiples.¹⁰

The *Piotroski F-Score* (Piotroski, 2000; Bajkowski, 2001), one of the most well-known value screens, has enjoyed an average annual return of 49.8% from 2003 to 2013 (Table 1). The Piotroski screen is categorized as a value screen because it first selects companies that fall into the bottom 20% of the market in terms of their price to book value. Price to book is found by dividing the current market value of a firm by the book value of that firm. This primary value screen aims to identify stocks that are neglected by the market or out-of-favor by the investors. The Piotroski screen then narrows down this pool of stocks by using nine metrics that focus on the financial condition of the firm. These include four profitability ratios, three leverage and liquidity ratios, and two operating efficiency ratios. For every condition that is met, the company is given one point and no points otherwise. To identify the healthiest companies among the basket of value stocks, the screen ranks the stocks by their F-score (sum of the nine points mentioned above) and picks the stocks with the highest F-scores. This screen has enjoyed annual returns of 49.8% on average from 2003 to 2013 (Table 1).

The *Buffett: Hagstrom* screen focuses on growth and value and attempts to follow the investment philosophy of Warren Buffett, the “Oracle of Omaha”. Robert Hagstrom has authored three popular books that highlight Buffett’s core investment principles. The most recent is titled “The Essential Buffett: Timeless Principles for the New Economy” (2002). Hagstrom identifies 12 basic principles that a company should possess to be considered for purchase. Not all of Buffett’s purchases displayed all of these tenets, but as a group the principles help to establish a reasonable approach to selecting stocks. The tenets cover both qualitative and quantitative business elements. AAI use these to create a stock screen using AAI’s fundamental stock screening program. This screen has enjoyed annual returns of 17.1% on average from 2003 to 2013 (Table 1). Since Buffett likes to focus on free cash flow, the primary value screen is to look for low price-to-free-cash-flow stocks. This represents a starting point for in-depth analysis, and imposing additional financial criteria.

¹⁰ In finance, the term “multiple” stands for a financial metric that is used to value a company. It is used as part of comparable analysis that identifies comparable companies, and converts these market values into standardized values relative to a key statistic, since absolute prices cannot be compared.

TABLE 1
Performance of Select Stock Screens from AAI Website

In this table, we present performance of three stock screens from the American Association of Individual Investors (AAII) website (www.aaii.com). The percentage amounts represent the “price gain”, the amount that each portfolio has appreciated or lost for each year. It only considers price change and does not include dividends. AAII uses *Stock Investor Pro* to perform the screens, which covers a universe of over 9,000 NYSE, Amex, Nasdaq, and over-the-counter stocks.

Screens	2003	2004	2005	2006	2007	2008
Piotroski: F-Score	131.6%	37.7%	37.8%	8.1%	1.8%	-35.3%
Graham-Enterprising	50.1%	10.2%	48.2%	34.5%	29.4%	-22.6%
Buffett: Hagstrom	35.2%	27.6%	11.4%	11.3%	14.4%	-25.8%
S&P 500	28.4%	10.7%	4.8%	15.6%	5.5%	-36.6%

Screens	2009	2010	2011	2012	2013	Average
Piotroski: F-Score	34.6%	138.8%	-36.4%	91.7%	137.6%	49.8%
Graham-Enterprising	168.4%	32.9%	-4.1%	-4.9%	2.1%	31.3%
Buffett: Hagstrom	30.0%	27.7%	8.2%	13.0%	35.4%	17.1%
S&P 500	25.9%	14.8%	2.1%	15.9%	32.2%	10.9%

The new and revised *Graham-Enterprising Investor* screen was created by AAII in conjunction with Bajkowski (2012) article. This screen follows Benjamin Graham’s investing philosophy, described in two best-selling books: “Security Analysis” (2008), and “Intelligent Investor” (2003).¹¹ Graham is considered the father of value investing, an investment approach he began teaching at Columbia Business School in 1928. Buffett described him as the second most influential person in his life after his own father. In fact, he named his son Howard Graham Buffett after Graham. “The Intelligent Investor” lays out specific sets of rules to follow when selecting stocks for both the conservative/defensive investor and the more aggressive/enterprising investor. The defensive or passive investor is one who does not have or is not willing to spend a great deal of time to analyze or track individual stocks. In contrast, the enterprising investor has greater market experience, as well as additional time to devote to portfolio management (Graham and Zweig, 2003). The *Graham-Enterprising Investor* screen has attained an average annual return of 31.3% from 2003 to 2013, according to AAII (Table 1). Graham’s primary value screen for the enterprising investor was to look for companies trading with price-to-earnings ratios below nine or 10 times trailing earnings. Graham thought about one in 10 stocks would pass such a filter in late 1971. AAII took this cue to establish a filter that required a company’s price-

¹¹ “Security Analysis” by Graham and Dodd was first published in 1934. After the third edition was published in 1951, fourth, fifth, and sixth editions were updated by the likes of Seth Klarman, Joel Greenblatt, James Grant, and others. Similarly, “Intelligent Investor” was first published in 1949, with the fourth revised edition being published in 1973, three years before Graham’s death.

to-earnings ratio to be among the lower 10% of all stocks and that their price-to-book ratio is less than or equal to 1.2. *The Graham-Enterprising Investor (Revised)* screen requires that a company's price-to-earnings ratio to be among the lower 25% of all stocks and that their price-to-book ratio is less than or equal to 1.2. The screen also assesses the financial health of a company in order to screen out obvious red flags by eliminating companies with a current ratio below 1.5, low earnings growth, and a negative bottom line. Appendix B lists all the financial criteria for each of these screens.

VI. Implementing ESG Integration Strategies

DATA

Our starting point is Compustat North America Fundamentals Annual database for 2003-2013 period, where we take firms with nonnegative sales, assets, and common equity.¹² We also require that all variables necessary to perform the *Piotroski F-score*, *Graham-Enterprising Investor*, and *Buffett: Hagstrom* screens are nonmissing. In particular, we collect over 40 pieces of financial information including market capitalization, earnings per share, shares outstanding, free cash flow per share, ROA, ROE, gross profit margin, operating profit margin, net profit margin, dividend yield, and valuation multiples. When valuation multiples such as price-to-earnings and price-to-free-cash-flow are negative, we replace them with missing values, so that firms with negative valuation multiples do not enter our analysis.¹³ See Appendix C for definition of these scores, and financial information that goes into each score calculation. Finally, the holding period returns (*HPR*) for all companies in the sample are calculated annually, which are used to calculate annual portfolio *HPRs*. It is important to note that while portfolio formation is done at $t=0$, portfolio return is evaluated at $t=1$ (which we call *Lead HPR*). Since our sample period is 2003-2013, we use Compustat data for 1998-2014 to ensure that we have lead returns for 2013, and five previous years of financials for 2003. For example, *Graham-Enterprising Investor* screen looks for companies with positive earnings per share for each of the five previous years and *Buffett: Hagstrom* screen looks for companies with positive operating income in each of the five previous years. These steps result in a sample of 53,449 firm-year observations, averaging 4,859 firms per year.

BASELINE PORTFOLIOS

The next step is to create baseline portfolios for *Piotroski F-score*, *Graham-Enterprising Investor*, and *Buffett: Hagstrom* screens using the Compustat sample, following

¹² The sample period is 2003-2013 due to the fact that MSCI ESG KLD STATS coverage of companies was much smaller at 1,100 companies prior to 2003, and 2013 represents the latest year for which MSCI ESG KLD STATS ratings are available to us.

¹³ Firms with positive price-to-book, but negative price-to-equity will enter our analysis when we are screening stocks based on price-to-book multiple (e.g., Piotroski F-score screen), but will not enter our analysis when we are screening based on price-to-earnings multiple (e.g., Graham-Enterprising Investor screen).

the criteria specified by www.aaii.com. Our universe of 4,859 stocks is smaller compared to the AAI universe of over 9,000 stocks. Table 2 presents descriptive statistics for the initial Compustat sample, S&P 500 sample (for which all financial data are available), and three portfolios for *Piotroski F-score*, *Graham-Enterprising Investor*, and *Buffett: Hagstrom* screens. When we implement the AAI criteria to create these portfolios, we get 333 firm-year observations for the *Piotroski F-score* portfolio, which represent approximately 30 firms per year. We get 182 firm-year observations for the *Graham-Enterprising Investor* portfolio, which represent approximately 17 firms per year. Finally, we get 349 firm-year observations for the *Buffett: Hagstrom* portfolio, which represent approximately 32 firms per year.

Table 2 reveals that the average firm in the Compustat sample has Piotroski F-score of 5.213, Buffett Hagstrom score of 3.201, Graham-Enterprising score of 2.235. It also has price-to-book multiple of 1.58, price-to-earnings multiple of 30.293, price-to-free-cash-flow multiple of 32.197, and market capitalization of \$4.83 billion. S&P 500 sample has higher average market capitalization of \$24.49 billion. As for the portfolios presented in Table 2, *Buffett: Hagstrom* portfolio has larger firms (average market capitalization of \$17.90 billion), and the *Piotroski F-score* portfolio has smaller firms (average market capitalization of \$4.85 billion). *Piotroski F-score* portfolio has an average F-score of 8.06 (since we require stocks to have at least 8 out of 9), and price-to-book multiple of 0.206 (since we require stocks to be in the bottom 20% of all stocks). *Graham-Enterprising Investor* portfolio has an average Graham Enterprising score of 5 (since we require stocks to have 5 out of 5), price-to-earnings multiple of 7.865 (since we require stocks to be in the bottom 25% of all stocks), and price-to-book multiple of 0.711 (since we require this multiple to be less than or equal to 1.2). *Buffett: Hagstrom* portfolio has an average Buffett Hagstrom score of 7.103 (since we require stocks to have at least 7 out of 8), and price-to-free-cash-flow multiple of 8.679 (since we require stocks to be in the bottom 30%).

TABLE 2
Descriptive Statistics for the Portfolios based on Compustat Sample

In this table, we present descriptive statistics for our initial Compustat sample, and the S&P500, *Piotroski F-score*, *Graham-Enterprising Investor*, and *Buffett: Hagstrom* portfolios over 2003-2013 period. In contrast to the universe of over 9,000 stocks that AAI uses, our universe is based on a sample of 4,859 firms per year, on average. *Piotroski F-score* portfolio consists of stocks with price-to-book (P/B) ratio in the bottom 20%, and with F-score of at least 8 (out of 9). *Graham-Enterprising Investor* portfolio consists of stocks with price-to-earnings (P/E) ratio in the bottom 15%, price-to-book (P/B) ratio less than or equal to 1.2, and with Graham-Enterprising score of at least 5 (out of 5). *Buffett: Hagstrom* portfolio consists of stocks with price-to-free-cash-flow (P/FCF) ratio in the bottom 30%, and with Buffett-Hagstrom score of at least 7 (out of 8). In addition to differing in sample of firms from AAI portfolios in Table 1, we also differ in our calculation of portfolio performance. In particular, we include not only capital gains yield, but also dividend yield to compute holding period return (AAI shows capital gains yield only). Below, we report lead holding period return (HPR), which is measured at the end of the following fiscal year, one year away from portfolio formation.

Compustat Sample	N	Mean	Median	Std Dev	Min	Max
Piotroski F-score	53,449	5.213	5.000	1.533	0.000	9.000
Buffett Hagstrom score	53,449	3.201	3.000	1.943	0.000	8.000
Graham Enterprising score	53,449	2.235	2.000	1.183	0.000	5.000
P/B	53,449	1.580	0.850	18.777	0.002	2705.250
P/E	37,777	30.293	18.152	44.604	0.062	331.333
P/FCF	35,976	32.197	15.469	64.164	0.031	486.655
MKTCAP (\$ bil)	53,449	4.825	0.479	18.866	0.000	623.622
HPR	49,427	0.126	0.051	0.603	-0.999	3.261
S&P500 Portfolio	N	Mean	Median	Std Dev	Min	Max
Piotroski F-score	5,214	5.543	6.000	1.381	1.000	9.000
Buffett Hagstrom score	5,214	4.923	5.000	1.564	0.000	8.000
Graham Enterprising score	5,214	3.034	3.000	1.147	0.000	5.000
P/B	5,214	1.257	0.937	1.174	0.005	12.847
P/E	4,721	24.349	17.662	31.582	1.331	331.333
P/FCF	4,539	26.502	15.951	48.115	0.215	486.655
MKTCAP (\$ bil)	5,214	24.489	11.232	42.818	0.120	623.622
HPR	5,032	0.113	0.102	0.390	-0.963	3.261
Piotroski F-score Portfolio	N	Mean	Median	Std Dev	Min	Max
Piotroski F-score	333	8.063	8.000	0.243	8.000	9.000
Buffett Hagstrom score	333	2.201	2.000	1.431	0.000	7.000
Graham Enterprising score	333	1.715	2.000	1.027	0.000	5.000
P/B	333	0.206	0.202	0.106	0.019	0.437
P/E	240	21.678	12.287	36.691	0.762	331.333
P/FCF	301	9.134	4.098	34.400	0.234	486.655
MKTCAP (\$ bil)	333	4.849	0.242	19.353	0.002	185.724
HPR	310	0.313	0.179	0.708	-0.870	3.261
Graham-Enterprising Investor Portfolio	N	Mean	Median	Std Dev	Min	Max
Piotroski F-score	182	5.797	6.000	1.409	2.000	9.000
Buffett Hagstrom score	182	4.698	5.000	1.598	1.000	8.000
Graham Enterprising score	182	5.000	5.000	0.000	5.000	5.000
P/B	182	0.711	0.706	0.279	0.066	1.196
P/E	182	7.865	8.156	2.332	2.009	12.008
P/FCF	150	19.300	11.238	42.319	1.426	486.655
MKTCAP (\$ bil)	182	9.548	1.202	31.704	0.011	211.362
HPR	172	0.187	0.136	0.511	-0.817	2.302
Buffett: Hagstrom Portfolio	N	Mean	Median	Std Dev	Min	Max
Piotroski F-score	349	5.805	6.000	1.423	3.000	9.000
Buffett Hagstrom score	349	7.103	7.000	0.305	7.000	8.000

Graham Enterprising score	349	3.415	3.000	0.972	1.000	5.000
P/B	349	1.686	1.202	2.277	0.062	20.610
P/E	349	12.706	12.441	6.913	2.419	82.253
P/FCF	349	8.679	8.705	2.731	1.328	13.899
MKTCAP (\$ bil)	349	17.898	3.523	51.329	0.035	453.113
HPR	332	0.145	0.127	0.406	-0.792	1.981

ESG INTEGRATED PORTFOLIOS

To implement the ESG integration approach, we merge Compustat data with MSCI ESG KLD STATS database by six-digit CUSIP and year. This process results in a sample of 27,598 firm-year observations, averaging 2,508 firms per year. Table 3 shows descriptive statistics on the ESG variables. The average firm in the sample has *All Strengths* score of 1.215, and *All Concerns* score of 1.399. The number of all strengths for a company ranges from 0 to 21, and the number of all concerns ranges from 0 to 15. *Adjusted All Strengths* variable ranges from 0 to 4.800, and *Adjusted All Concerns* ranges from 0 to 4.433. Due to number of concerns being higher than number of strengths on average, both *ESG Index* and *Adjusted ESG Index* numbers are negative (-0.183 and -0.208, respectively). As for controversial business involvement indicators, these are dummy variables, taking a value of one if a company has involvement in particular controversial area, and zero otherwise. Only 1.4% of our sample of firms has gambling involvement, 3.3% has military involvement, 1.5% has nuclear involvement, 0.6% has tobacco involvement, and 0.2% has fire-arms involvement.

TABLE 3
Descriptive Statistics for Merged Compustat and ESG Sample

All Strengths (*All Concerns*) is a sum of all strengths (concerns) indicators across environment, community, human rights, employee relations, diversity, and product categories. In addition to raw *All Strengths* and *All Concerns* scores, we calculate *Adjusted All Strengths* and *Adjusted All Concerns* scores. The reason for adjusting the raw scores is that the number of possible strengths and concerns varies from year to year. Every year, some new indicators are introduced and/or other indicators get discontinued. Therefore, we scale the score for each category (community, environment, etc.) by the maximum value for that category in a given year. We then add up all adjusted strengths and concerns scores across all categories to calculate *Adjusted All Strengths* and *Adjusted All Concerns*. In order to calculate *Adjusted ESG Index*, we first find adjusted net score for each category, and then add up the adjusted net scores.

Variable	Label	N	Mean	Median	Std Dev	Min	Max
All_Str	All Strengths	27,598	1.22	0.00	2.23	0.00	21.00
All_Con	All Concerns	27,598	1.40	1.00	1.50	0.00	15.00

ESG_Index	All Strengths – All Concerns	27,598	-0.18	0.00	2.25	-9.00	18.00
All_Str_Adj	Adjusted All Strengths	27,598	0.25	0.00	0.49	0.00	4.80
All_Con_Adj	Adjusted All Concerns	27,598	0.46	0.50	0.46	0.00	4.43
ESG_Index_Adj	Adjusted ESG Index	27,598	-0.21	-0.25	0.57	-3.00	4.05
Com_Str_Num	Community - # of Strengths	27,598	0.13	0.00	0.44	0.00	5.00
Env_Str_Num	Environment - # of Strengths	27,598	0.21	0.00	0.64	0.00	6.00
Div_Str_Num	Diversity - # of Strengths	27,598	0.48	0.00	0.96	0.00	7.00
Emp_Str_Num	Emp. Relations - # of Strengths	27,598	0.31	0.00	0.81	0.00	9.00
Hum_Str_Num	Human Rights - # of Strengths	27,598	0.02	0.00	0.15	0.00	2.00
Pro_Str_Num	Product - # of Strengths	27,598	0.07	0.00	0.27	0.00	3.00
Com_Con_Num	Community - # of Concerns	27,598	0.07	0.00	0.27	0.00	3.00
Env_Con_Num	Environment - # of Concerns	27,598	0.17	0.00	0.58	0.00	5.00
Div_Con_Num	Diversity - # of Concerns	27,598	0.56	0.00	0.67	0.00	3.00
Emp_Con_Num	Emp. Relations - # of Concerns	27,598	0.36	0.00	0.62	0.00	5.00
Hum_Con_Num	Human Rights - # of Concerns	27,598	0.05	0.00	0.22	0.00	3.00
Pro_Con_Num	Product - # of Concerns	27,598	0.19	0.00	0.52	0.00	4.00
Alc_Con_A	Alcohol Involvement	27,598	0.01	0.00	0.10	0.00	1.00
Gam_Con_A	Gambling Involvement	27,598	0.01	0.00	0.12	0.00	1.00
Mil_Con_A	Military Involvement	27,598	0.03	0.00	0.18	0.00	1.00
Nuc_Con_A	Nuclear Involvement	27,598	0.02	0.00	0.12	0.00	1.00
Tob_Con_A	Tobacco Involvement	27,598	0.01	0.00	0.08	0.00	1.00
Fir_Con_A	Firearms Involvement	27,598	0.00	0.00	0.05	0.00	1.00

Since the baseline portfolios discussed in the previous section contain small number of firms (30 for *Piotroski F-score* portfolio, 17 for *Graham-Enterprising Investor* portfolio, and 32 for the *Buffett: Hagstrom* portfolio), we relax the criteria for each portfolio so that we can impose additional ESG criteria and still end up with a reasonable size portfolio. For *Piotroski F-score* portfolio, we require that F-score to be at least 7 (out of 9), and price-to-book multiple to be in the bottom 30%. For *Graham-Enterprising Investor* portfolio, we require that Graham Enterprising score to be at least 4 (out of 5), price-to-earnings multiple to be in the bottom 30%, and price-to-book multiple to be at least 1.2. For *Buffett: Hagstrom* portfolio, we require that *Buffett Hagstrom* score to be at least 5 (out of 8) and price-to-free-cash-flow multiple to be in the bottom 25%.

We first implemented positive and negative screening techniques separately using the aggregated *Adjusted All Strengths* and *Adjusted All Concerns* scores, and then in combination with each other by using the *Adjusted Net Strengths* score. Using different cutoffs (bottom 10%, 30%, and 50% for *Adjusted All Concerns*, top 10%, 30%, and 50% for *Adjusted All Strengths* and *Adjusted Net Strengths*), we evaluated returns of newly formed ESG integrated portfolios. For example, in one scenario, companies that have a lower than 80th percentile *Adjusted All Strengths* score (ALL_STR_ADJ) or a higher than 30th percentile *Adjusted All Concerns* score (ALL_CON_ADJ) were removed from the baseline portfolio. In a different scenario, companies that have a lower than median *Adjusted All Strengths* score (ALL_STR_ADJ) or an above median *Adjusted All Concerns* score (ALL_CON_ADJ) were removed from the baseline portfolio. For each scenario, the new holding period return for the ESG integrated portfolio was calculated by averaging annual returns of the remaining companies. Unfortunately, none of these integration strategies had a significant positive impact on returns. For example, among scenarios presented in top panel of Table 4, only two scenarios show improvement over the baseline portfolio: removing companies with a higher than 30th percentile *Adjusted All Concerns* score improves *Graham-Enterprising* portfolio return by 1.33%, and removing companies with a lower than 80th percentile *Adjusted Net Strengths* score improves *Graham-Enterprising* portfolio return by 2.33%.

TABLE 4
Applying ESG Criteria to Base Portfolios
Based on Merged Compustat and ESG Sample

Below, we present HPR's of base portfolios, as well as those of *Piotroski High F-score* portfolio, *Graham-Enterprising Investor* portfolio, and *Buffett: Hagstrom* portfolio with various ESG criteria integrated, over 2003-2013 period. HPR for S&P500 over the same period is 11.03%. Con_Adj labels stand for *Adjusted Concerns* scores, and Str_Adj labels stand for *Adjusted Strengths* scores. P20, P30, P70, P80, and P90 stand for 20th percentile, 30th percentile, 70th percentile, 80th percentile, and 90th percentile, respectively. *Piotroski F-score* portfolio consists of stocks with price-to-book (P/B) ratio in the bottom 30%, and with F-score of at least 7 (out of 9). *Graham-Enterprising Investor* portfolio consists of stocks with price-to-earnings (P/E) ratio in the bottom 30%, price-to-book (P/B) ratio less

than or equal to 1.2, and with Graham-Enterprising score of at least 4 (out of 5). *Buffett: Hagstrom* portfolio consists of stocks with price-to-free-cash-flow (P/FCF) ratio in the bottom 25%, and with Buffett-Hagstrom score of at least 5 (out of 8). When the HPR of ESG integrated portfolio is greater (lower) than that of S&P500, we present the portfolio return in green (red). When a certain ESG criteria leads to improvement in all three baseline portfolios, we present the portfolio returns in a bold case.

ESG Criteria Applied	HPR of Piotroski F-score Portfolio	HPR of Graham-Enterprising Portfolio	HPR of Buffett: Hagstrom Portfolio
All_Con_Adj ≤ All_Con_Adj_P30	18.34%	15.10%	14.20%
All_Str_Adj ≥ All_Str_Adj_P70	17.38%	13.45%	14.48%
Net_Str_Adj ≥ Net_Str_Adj_P80	16.26%	16.10%	14.01%
Com_Con_Adj ≤ Com_Con_Adj_P30	20.70%	14.45%	16.42%
Env_Con_Adj ≤ Env_Con_Adj_P30	19.39%	14.94%	14.66%
Div_Con_Adj ≤ Div_Con_Adj_P20	18.34%	13.69%	16.01%
Emp_Con_Adj ≤ Emp_Con_Adj_P30	18.40%	14.99%	13.58%
Hum_Con_Adj ≤ Hum_Con_Adj_P30	19.82%	13.82%	15.37%
Pro_Con_Adj ≤ Pro_Con_Adj_P30	19.60%	14.52%	14.78%
Com_Str_Adj ≥ Com_Str_Adj_P90	21.96%	15.49%	17.37%
Env_Str_Adj ≥ Env_Str_Adj_P80	21.07%	13.14%	16.26%
Div_Str_Adj ≥ Div_Str_Adj_P80	14.08%	12.27%	16.32%
Emp_Str_Adj ≥ Emp_Str_Adj_P70	20.81%	13.83%	15.85%
Hum_Str_Adj ≥ Hum_Str_Adj_P90	19.86%	13.77%	15.68%
Pro_Str_Adj ≥ Pro_Str_Adj_P90	19.79%	14.01%	15.57%
Base Portfolio (without any ESG criteria)	19.86%	13.77%	15.68%

Since the use of aggregated ESG variables in ESG integration strategies was not successful, we examined the effect of each ESG category on portfolio returns separately. To show how we experimented with different category screens for our baseline portfolios, we recorded the returns of ESG integrated portfolio compared to those of the baseline portfolios in Table 4 (bottom two panels). We then implemented an exclusionary strategy by removing companies that have involvement in controversial business issues from the baseline portfolios. The differences between ESG integrated portfolio returns and baseline portfolio returns were extremely minimal, largely due to a small number of companies in our sample with involvement in controversial business issues such as gambling, tobacco, and military. From all the different permutations we tried and presented in Table 4, the following ESG criteria improved the baseline portfolio returns most consistently: To pass the ESG screen, a company in the baseline portfolio should have (i) *Adjusted Community Concerns* score (COM_CON_ADJ) in the bottom 30%, (ii) *Adjusted Community Strengths* score (COM_STR_ADJ) in the top 10%, and (iii) *Adjusted Employee Strengths* score (EMP_STR_ADJ) in the top 30% of all stocks in the baseline portfolio. Thus, we

combined these three ESG criteria and formed our final ESG integrated portfolios, and presented their performance year by year in Table 5.

On average, returns on all three of baseline portfolios were increased by approximately 2.7% to 3.5% by using these three criteria, and all ESG integrated portfolios had higher returns than the S&P 500 (shown in the last column in Table 5). Consistent with AAI's findings, we also find that *Piotroski F-score* screen performs the best out of three value screens we examined: baseline *Piotroski F-score* portfolio has 19.9% return on average, compared to 13.8% for *Graham-Enterprising Investor* portfolio, and 15.7% for *Buffett: Hagstrom Portfolio*. The same pattern holds for ESG integrated portfolios: ESG integrated *Piotroski F-score* portfolio has 22.7% return on average, compared to 17.3% for *Graham-Enterprising Investor* portfolio, and 18.4% for *Buffett: Hagstrom Portfolio*. In addition, the ESG integrated *Piotroski F-score* portfolio outperforms the S&P 500 in 10 out of 11 years, the ESG integrated *Graham-Enterprising Investor* portfolio outperforms the S&P 500 in 7 out of 11 years, and the ESG integrated *Buffett: Hagstrom* portfolio outperforms the S&P 500 in 10 out of 11 years.

We also perform two-sample t-tests to check whether returns of our three portfolios are significantly different from those of the S&P 500 index. The null hypothesis is that average return of the portfolio and the S&P 500 index are equal. When we reject the null hypothesis at 1% or 5% significance level, we present the HPR for that specific portfolio in a bold case. When it is statistically significant outperformance (underperformance), we present the portfolio HPR with an asterisk (underline) (see Table 5). These statistical tests reveal that *Piotroski F-score* and *Buffett: Hagstrom* portfolios outperform S&P 500 over our sample period with ESG integration, while *Graham-Enterprising Investor* portfolio does not. Based on the annual returns presented in Table 5, we plot value of \$1,000 invested in our ESG integrated portfolios over time in Figure 3. Along with the Table 5 results, this figure highlights the strong performance of ESG integrated portfolios based on *Piotroski F-score* and *Buffett: Hagstrom* screens.

TABLE 5
ESG Integrated Portfolio Returns

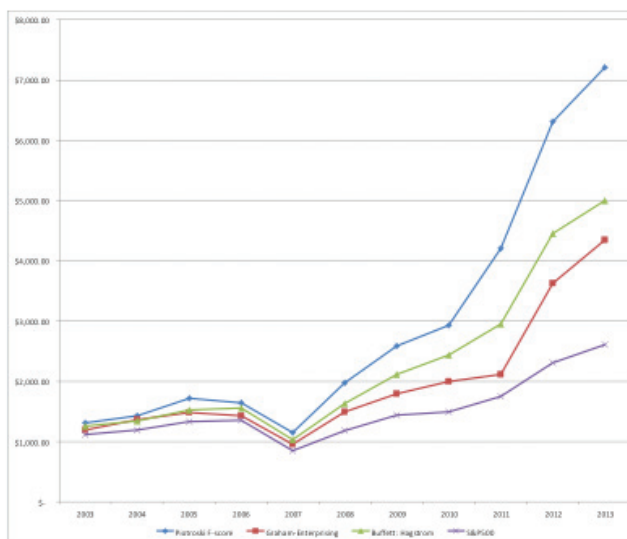
Below we present HPRs of baseline and ESG integrated portfolios, compared with HPRs of S&P500. Baseline *Piotroski F-score* portfolio consists of stocks with price-to-book (P/B) ratio in the bottom 30%, and with F-score of at least 7 (out of 9). *Graham-Enterprising Investor* portfolio consists of stocks with price-to-earnings (P/E) ratio in the bottom 30%, price-to-book (P/B) ratio less than or equal to 1.2, and with Graham-Enterprising score of at least 4 (out of 5). *Buffett: Hagstrom* portfolio consists of stocks with price-to-free-cash-flow (P/FCF) ratio in the bottom 25%, and with Buffett-Hagstrom score of at least 5 (out of 8). Two-sample t-tests were used to test for equality of portfolio average HPRs. When the null hypothesis that the average portfolio HPR is equal to the average S&P500 HPR is rejected at 1% or 5% level, the portfolio HPR's are presented in a **bold** case. When it is statistically significant outperformance (underperformance), we present the portfolio HPR with an asterisk (underline).

Piotroski F-score	2003	2004	2005	2006	2007	2008
N	146	166	95	110	91	68
Base HPR	30.3%*	10.1%	19.0%*	-5.3%	-30.4%	66.2%*
N	131	148	83	95	73	56
ESG HPR	30.4%*	9.5%	20.0%*	-4.5%	-30.1%	72.6%*
Graham-Enterprising	2003	2004	2005	2006	2007	2008
N	118	108	90	93	100	132
Base HPR	19.4%*	12.7%	11.1%	-1.3%	-32.8%	49.7%
N	101	98	74	75	80	95
ESG HPR	19.4%*	13.9%*	9.0%	-3.6%	-33.0%	55.6%*
Buffett: Hagstrom	2003	2004	2005	2006	2007	2008
N	118	179	155	158	126	73
Base HPR	25.3%*	6.2%	12.8%	0.4%	-31.1%*	53.3%
N	105	155	129	135	107	63
ESG HPR	25.5%*	6.8%	13.5%	2.6%	-34.2%	58.6%*
S&P500	2003	2004	2005	2006	2007	2008
N	461	453	381	377	375	399
HPR	11.8%	6.3%	11.4%	1.7%	-37.2%	39.5%
Piotroski F-score	2009	2010	2011	2012	2013	Ave.
N	130	201	121	103	117	123
Base HPR	28.9%*	6.4%	32.8%*	56.1%*	4.5%	19.9%*
N	108	20	9	8	13	68
ESG HPR	30.9%*	13.1%	43.7%*	50.0%	14.2%	22.7%*
Graham-Enterprising	2009	2010	2011	2012	2013	Ave.
N	96	94	97	91	98	102
Base HPR	17.3%	5.4%	14.6%	42.0%*	13.4%	13.8%
N	74	14	22	16	23	61
ESG HPR	20.2%	11.5%	6.1%	71.3%*	19.9%	17.3%
Buffett: Hagstrom	2009	2010	2011	2012	2013	Ave.
N	83	98	105	97	134	121
Base HPR	27.5%	3.9%	21.1%	43.8%*	9.3%	15.7%*
N	68	16	17	16	23	76
ESG HPR	29.8%	15.2%	21.1%	51.0%*	12.5%	18.4%*
S&P500	2009	2010	2011	2012	2013	Ave.
N	409	409	408	400	432	409
HPR	21.8%	3.4%	17.3%	32.1%	13.1%	11.0%

VII. Conclusion

By incorporating specific ESG criteria into a fundamental analysis, competitive returns can be earned; however, not all ESG criteria have the same impact on a portfolio. In fact, using stringent criteria that take into account all strengths and concerns indicators in MSCI ESG KLD STATS database can have a detrimental effect on portfolio performance, consistent with the *underperformance hypothesis*. Our results also suggest that incorporating social ratings in community and employee relations areas improve baseline portfolio returns most consistently. These results are supportive of the *outperformance hypothesis* for SRI investments in that with diligent analysis on company fundamentals as well as their ESG performance, one can outperform the market. While past performance is no guarantee of future results, stock screens are only the first step. Investors can use them to generate an initial list of potential investments that merit further research.

FIGURE 3
Value of \$1,000 Invested in ESG Integrated Portfolios vs. S&P500



With the trend of investing in socially responsible firms growing so rapidly and the increased popularity of ESG ranking systems such as the Bloomberg, Sustainalytics, and MSCI, it is likely that companies with strong ESG performance will become more attractive to investors. “We have the opportunity to marry making money with a purpose,” Mr. Feiner said, who is head of ESG research at Arabesque. “That is a very rare thing in finance.”¹⁴ Arabesque prides itself as the first firm exclusively focused on ESG investing through quantitative analysis — that is, with algorithms, rather than humans, picking the stocks. It has two funds, with about \$50 million under management.

¹⁴ http://www.nytimes.com/2016/04/07/business/dealbook/investing-with-a-conscience-but-done-by-a-robot.html?_r=0

References

- Aktas, N., De Bodt, E., Cousin, J.,** (2011). "Do financial markets care about SRI? Evidence from mergers and Acquisitions." *Journal of Banking & Finance* 35(7), 1753-761.
- Anderson, A., Myers, D. H.** (2007). "Performance and predictability of social screens." Working Paper, Lehigh University.
- Bajkowski, J.,** (2001). "Finding the winners among low price-to-book-value stocks." *AII Journal* August, 13-16.
- Bajkowski, J.,** (2002). "Adjusting the Benjamin Graham Enterprising Investor screen." *AII Journal* October, 25-29.
- Barber, B. M.,** (2007), "Monitoring the monitor: Evaluating CalPERS' activism." *The Journal of Investing* Winter, 66-80.
- Barnett, M. L., Salomon, R. M.,** (2003). "Throwing a curve at socially responsible investing research." *Organization & Environment* 16(3), 381-389.
- Bauer, R., Koedjik, K., Otten, R.,** (2005). "International evidence on ethical mutual fund performance and investment style." *Journal of Banking and Finance* 29, 1751-1767.
- Bello, Z. Y.** (2005). "Socially responsible investing and portfolio diversification." *Journal of Financial Research* 28(1), 41-57.
- Benson, K., L., Humphrey, J. E.,** (2008). "Socially responsible investment funds: Investor reaction to current and past returns." *Journal of Banking and Finance* 32, 1850-1859.
- Cornett, M. M., Erhemjamts, O., Tehranian, H.,** (2016). "Greed or good deeds: An examination of the relation between corporate social responsibility and the financial performance of US commercial banks around the financial crisis." *Journal of Banking and Finance* 70, 137-159.
- Dah, M., Hoque, M., Wang, S.,** (2015). "Constrained investments and opportunity cost: Evidence from Islamic funds." *Managerial Finance* 41(4), 348-367.
- Dowell, G., Hart, S., Yeung, B.,** (2000). "Do corporate global environmental standards create or destroy market value?" *Management Science* 46(8), 1059-1074.
- Edmans, A.,** (2011). "Does the stock market fully value intangibles? Employee satisfaction and equity prices." *Journal of Financial Economics* 101, 621-640.
- Erhemjamts, O., Li, Q., Venkateswaran, A.,** (2013). "Corporate social responsibility and its impact on firm's investment policy, organizational strategy, and performance." *Journal of Business Ethics* 118, 395-412.
- Goldreyer, E.F., Ahmed, P., Diltz, J.D.,** (1999). "The performance of socially responsible mutual funds: Incorporating sociopolitical information in portfolio selection." *Managerial Finance* 25 (1), 23-36.

- Gompers, P., Ishii, J., Metrick, A.,** (2003). “Corporate governance and equity prices.” *Quarterly Journal of Economics* 118(1), 107-155.
- Graham, B., Dodd, D.,** (2008). *Security Analysis*, 6th edition. McGraw-Hill Education.
- Graham, B., Zweig, J.,** (2003). *The Intelligent Investor*, Revised 1973 edition. New York: Harper Business Essentials.
- Guenster, N., Bauer, R., Derwall, J., Koedjik, K.,** (2011). “The economic value of corporate eco-efficiency.” *European Financial Management* 17(4), 679-701.
- Hagstrom, R. G.,** (2002). *The Essential Buffett: Timeless Principles for the New Economy*. New York: Wiley.
- Hamilton, S., Jo, H., Statman, M.,** (1993). “Doing well while doing good? The investment performance of socially responsible mutual funds.” *Financial Analysts Journal* 49 (6), 62–66.
- Kempf, Z. Y.,** (2005). “Socially responsible investing and portfolio diversification.” *Journal of Financial Research* 28(1), 41-57.
- Nofsinger, J., Varma, A.,** (2014). “Socially responsible funds and market crises.” *Journal of Banking and Finance* 48, 180-193.
- Piotroski, J. D.,** (2000). “Value investing: The use of historical financial statement information to separate winners from losers.” *Journal of Accounting Research* 38, 1-41.
- Renneboog, L., Ter Horst, J., Zhang, C.,** (2008a). “Socially responsible investments: Institutional aspects, performance, and investor behavior.” *Journal of Banking & Finance* 32(9), 1723-742.
- Renneboog, L., Ter Horst, J., Zhang, C.,** (2008b). “The price of ethics and stakeholder governance: The Performance of socially responsible mutual funds.” *Journal of Corporate Finance* 14(3), 302-22.
- Russo, M. V., Fouts, P. A.,** (1997). “A resource-based perspective on corporate environmental performance and profitability.” *Academy of Management Journal* 40(3), 534-559.
- Statman, M.,** (2000). “Socially responsible mutual funds.” *Financial Analysts Journal* 56(3), 30-39.
- Statman, M., Glushkov, D.,** (2009). “The wages of social responsibility.” *Financial Analysts Journal* 65(4), 33-46.
- US SIF Foundation,** (2014). *US Sustainable, Responsible and Impact Investing Trends, 10th Edition*. Washington, DC.
- Verwijmeren, P., Derwall, J.,** (2010). “Employee well-being, firm leverage, and bankruptcy risk.” *Journal of Banking and Finance* 34, 956-964.

Appendix A

MSCI ESG Performance Indicators

MSCI ESG KLD STATS (KLD) is an annual data set of positive and negative environmental, social, and governance (ESG) performance indicators applied to a universe of publicly traded companies. ESG performance indicators in STATS are scored by a simple binary scoring model: If a company meets the assessment criteria established for an indicator, then this is signified with a “1”. If a company does NOT meet the assessment criteria established for an indicator, then this is signified with a “0”. If a company has NOT been researched for a particular ESG indicator, then this is signified with a “NR” (Not Researched). The following indicators are from 2013 Methodology Guide provided by MSCI ESG Research Inc. Note that the number of indicators varies from year to year, as new indicators get introduced and/or some of the existing indicators get discontinued each year. For further definitions on each of the indicators, visit MSCI ESG website <https://www.msci.com/www/research-paper/esg-ratings-methodology/0175943017>.

Positive Environment Performance Indicators	Data Set Column Headers	Negative Environment Performance Indicators	Data Set Column Headers
Environmental Opportunities - Clean Tech	ENV-str-A	Regulatory Compliance	ENV-con-B
Waste Management - Toxic Emissions and Waste	ENV-str-B	Toxic Emissions and Waste	ENV-con-D
Waste Management - Packaging Materials & Waste	ENV-str-C	Energy & Climate Change	ENV-con-F
Climate Change - Carbon Emissions	ENV-str-D	Impact of Products and Services	ENV-con-G
Environmental Management Systems	ENV-str-G	Biodiversity & Land Use	ENV-con-H
Natural Resource Use - Water Stress	ENV-str-H	Operational Waste	ENV-con-I
Natural Resource Use - Biodiversity & Land Use	ENV-str-I	Supply Chain Management	ENV-con-J
Natural Resource Use - Raw Material Sourcing	ENV-str-J	Water Stress	ENV-con-K
Environment - Other Strengths	ENV-str-X	Environment - Other Concerns	ENV-con-X
Positive Social Performance Indicators	Data Set Column Headers	Negative Social Performance Indicators	Data Set Column Headers
Innovative Giving	COM-str-B	Community Impact	COM-con-B
Community Engagement	COM-str-H		
Indigenous Peoples Relations	HUM-str-D	Support for Controversial Regimes	HUM-con-C

Human Rights Policies & Initiatives	HUM-str-X	Freedom of Expression and Censorship	HUM-con-J
		Human Rights Violations	HUM-con-K
		Human Rights - Other Concerns	HUM-con-X
Union Relations	EMP-str-A	Union Relations Concern	EMP-con-A
Cash Profit Sharing	EMP-str-C	Health and Safety Concern	EMP-con-B
Involvement	EMP-str-D	Supply Chain Controversie	EMP-con-F
Health & Safety Policies & Initiatives	EMP-str-G	Supply Chain -Child Labor	EMP-con-G
Supply Chain Labor Standards	EMP-str-H	Labor Rights & Supply Chain - Other Concerns	EMP-con-X
Compensation and Benefits	EMP-str-I		
Employee Relations	EMP-str-J		
Professional Development	EMP-str-K		
Human Capital Management	EMP-str-L		
Board of Directors - Gender	DIV-str-C	Workforce Diversity	DIV-con-A
Women & Minority Contracting	DIV-str-E	Board Diversity - Gender	DIV-con-C
Employment of Underrepresented Groups	DIV-str-H	Board Diversity - Minorities	DIV-con-D
Product Safety and Quality	PRO-str-A	Product Quality & Safety	PRO-con-A
Social Opportunities	PRO-str-C	Marketing & Advertising	PRO-con-D
Access to Finance	PRO-str-D	Anticompetitive Practices	PRO-con-E
		Customer Relations	PRO-con-F
		Other Concerns	PRO-con-X
Positive Governance Performance Indicators	Data Set Column Headers	Negative Governance Performance Indicators	Data Set Column Headers
Reporting Quality	CGOV-str-D	Reporting Quality	CGOV-con-H
Corruption & Instability	CGOV-str-G	Governance Structures	CGOV-con-K
		Controversies	
		Controversial Investments	CGOV-con-L
Financial System Risk	CGOV-str-H	Bribery & Fraud	CGOV-con-M
		Governance - Other Concerns	CGOV-con-X
		Controversial Business Issues	Data Set Column Headers
		Alcohol	ALC-con-A
		Gambling	GAM-con-A
		Tobacco	TOB-con-A
		Firearms	FIR-con-A
		Military	MIL-con-A
		Nuclear Power	NUC-con-A

Appendix B

Criteria for Select Stock Screens

Screening criteria for *Graham-Enterprising Investor (Revised)*, *Buffett: Hagstrom*, and *Piotroski F-score* screens are retrieved from www.aaii.com.

Graham - Enterprising Investor (Revised) Screen

Price-to-earnings ratio must be in the lower 25% of all stocks

Price-to-book ratio must be less than or equal to 1.2

1. Current ratio greater than 1.5
2. Long term debt to working capital ratio is less than 110%
3. Earnings per share (EPS) for each of the last 5 fiscal years have been greater than 0
4. Dividend yield greater than 0
5. Earnings per share (EPS) for the last fiscal year must be greater than EPS from 5 years ago

Buffett: Hagstrom Screen

Price-to-free-cash-flow ratio must be in the lower 30% of all stocks

1. Market cap above \$1 billion
2. No operating losses over the last 5 years
3. Current return on equity is greater than 15%
4. Return on equity for each of the last five years was greater than 15%
5. Debt to asset ratio below industry median
6. Operating margin above industry median
7. Net profit margin above industry median
8. Growth in stock price over the last 5 years is greater than growth in retained earnings

Piotroski F-Score Screen

Price-to-book ratio must be in the lower 20% of all stocks

1. Positive return on assets
2. Positive cash flow from operations/total assets
3. Higher return on assets in the current year than the past
4. Cash flows from operations/total assets greater than return on assets
5. Lower long term debt/total assets in current year compared to past year
6. Higher current ratio in current year than previous year
7. No new shares issued in the past year
8. Higher gross margin compared to previous year
9. Higher asset turnover ratio compared to previous year

Appendix C

Description of Financial Variables

Source of these variables is Compustat North America Fundamentals Annual database for 2003-2013 period.

Piotroski F-score Screen	Variable Description
SALE	Sales
PB	Price to Book = Stock price at the end of the fiscal year (PRCC_F) divided by book value per share (AT/CSHO)
OIBDP	Operating Income Before Depreciation
ROA	Return on Assets = Operating income before depreciation (OIBDP) divided by total assets (AT)
ROA_POS	1 if ROA is positive; 0 otherwise
ROA_INC	1 if ROA increased from prior year; 0 otherwise
CFO_ASSETS	Cash flows from operations (OANCF) to total assets (AT)
CFO_POS	1 if CFO_ASSETS is positive; 0 otherwise
CFO_GT_ROA	1 if CFO_ASSETS is greater than ROA; 0 otherwise
LTD_ASSETS	Long term debt (DLTT) to total assets (AT)
LTD_DEC	1 if LTD_ASSETS is decreased from prior year; 0 otherwise
CR	Current ratio = Current assets (ACT) divided by debt in current liabilities (DLC)
CR_INC	1 if current ratio increased from prior year; 0 otherwise
NEW_ISSUES	1 if common shares issued (CSHI) last year is non-zero; 0 otherwise
GPM	Gross profit margin = Gross Profits (GP) divided by sales (SALE)
GPM_INC	1 if GPM increased from prior year; 0 otherwise
TAT	Total asset turnover = Sales (SALE) divided by total assets (AT)
TAT_INC	1 if TAT has increased from prior year; 0 otherwise
PIOTROSKI_F_SCORE	Sum of 9 dummy variables: ROA_POS + ROA_INC + CFO_POS + CFO_GT_ROA + LTD_DEC + CR_INC + NEW_ISSUES + GPM_INC + TAT_INC
Graham-Enterprising Screen	Variable Description
CR_GT_1PT5	1 of current ratio less than 1.5; 0 otherwise
LTD_LT_WCAP	1 if long-term debt (DLTT) is less than working capital (WCAP); 0 otherwise
DIV_YLD	Dividend yield = Dividend per share (DVT/CSHO) divided by stock price (PRCC_F)
DIV_POS	1 if dividend yield is positive; 0 otherwise
PE	Price to earnings = Stock price (PRCC_F) divided by earnings per share (EPSPX)
PB	Price to Book = Stock price at the end of the fiscal year (PRCC_F) divided by book value per share (AT/CSHO)

EPS_ALL5_POS	1 if EPS is positive in each of the 5 previous years; 0 otherwise
G_EPS5YR_POS	1 if current EPS is greater than EPS 5 years ago; 0 otherwise
GRAHAM_ENTERPRISING_SCORE	Sum of 5 dummy variables: CR_GT_1PT5 + LTD_LT_WCAP + DIV_POS + EPS_ALL5_POS + G_EPS5YR_POS
Buffett: Hagstrom Screen	Variable Description
MKTCAP	Market Cap = Stock price (PRCC_F) times common shares outstanding (CSHPRI)
MCAP_GT_1B	1 if MKTCAP is greater than \$1 billion; 0 otherwise
OIADP	Operating Income After Depreciation
OIADP_ALL5_POS	1 if OIADP is positive in each of the 5 previous years; 0 otherwise
ROE	Return on Equity = Net income (NI) divided by book value of common equity (CEQ)
ROE_GT_15	1 if ROE is greater than 15%; 0 otherwise
ROE5_GT_15	1 if ROE is greater than 15% for each of the past 5 years; 0 otherwise
DEBT_ASSETS	Debt to asset ratio = Sum of debt in current liabilities (DLC) and long-term debt (DLTT) divided by total assets (AT)
DEBT_LT_P50	1 if DEBT_ASSETS ratio is less than 50th percentile (median); 0 otherwise
FCF_SH	Free cash flows per share = Cash flow from operations (OANCF) minus capital spending (CAPX), divided by common shares outstanding (CSHO)
P_FCF	Price to free cash flow = Stock price (PRCC_F) divided by free cash flow per share (FCF_SH)
OPM	Operating Profit Margin = Operating income after depreciation (OIADP) divided by sales (SALE)
OPM_GT_P50	1 if OPM is greater than 50th percentile; 0 otherwise
NPM	Net Profit Margin = Net income (NI) divided by sales (SALE)
NPM_GT_P50	1 if NPM is greater than 50th percentile; 0 otherwise
BUFFETT_HAGSTROM_SCORE	Sum of 7 dummy variables: MCAP_GT_1B + OIADP_ALL5_POS + ROE_GT_15 + ROE5_GT_15 + DEBT_LT_P50 + OPM_GT_P50 + NPM_GT_P50
Portfolio Performance	Variable Description
HPR	Lead Holding Period Return = Sum of stock price and dividends at the end of the following fiscal year divided by current fiscal year's stock price = Capital gains yield + dividend yield

SOUTHWEST AIRLINES AND THE IMPACT OF LOW-COST CARRIERS ON AIRLINE TICKET PRICES

By Sidney Field

Southwest Airlines is the indisputable leader of the Low-Cost Airline industry in the United States. A product of deregulation, the expansion of its unique business model has been responsible for the substantial real price declines in consumer airfares since the 1980s. However, recent fuel price declines have had a significant impact on the industry and their impact on the low-cost sector today, as the driver of price competition, is important to understand. This paper includes a multivariate linear regression of 2015 data to examine the impact of various US LCCs on average airport, airport-route, and city-route fares, and compares the impacts to the results of ticket data from 2012 and 2007, with a particular interest in the effect of Southwest. It concludes that despite industry concentration, the association of Southwest with lower average route, airport, and market prices had actually strengthened from 2012 to 2015.

Keywords: Airlines; Low-cost airlines; LCCs; Airline deregulation; Southwest effect; Airline pricing; Legacy carriers.

I. Introduction

The growth of Southwest airlines in the era of deregulated transportation markets has arguably been the most influential development in air transport within the US. This study investigates the effect that Southwest and its Low-Cost Carrier (LCC) contemporaries have had on average US ticket prices in the face of significant structural changes in the industry. This paper presents a brief history of US airline deregulation and the subsequent growth of Southwest Airlines resulting from its unique business model and its impact on airline competition. An econometric model follows examining the fare impacts of Southwest's growth from several perspectives: specific airports, airport-pair routes, and city-pair routes for 2007, 2012, and 2015.

Major merger activity has reduced the eight largest US airlines (including Southwest, which acquired AirTran) to four since 2008, and such a shift makes it important to understand the nature of price competition in the industry's current state. Legacy carriers, which today are only American, Delta, and United, are the "full fare" airlines that the LCC model aspires to compete with on price. These carriers comprise the status quo against which the presence of particular LCCs at specific airports, on routes between specific airports, and on routes between specific cities, is measured in this study, through the impact on average airfares. With the recent attrition of half of the industry players through consolidation, the

* Thank you to my advisor, Professor Joel Deichmann, and my fellow classmate Christopher DeMeo.

expectation is that price competition by Southwest, as measured against the full fare legacy carriers, has decreased. Since only four firms today as opposed to eight in 2008, account for nearly 90% of domestic traffic, it is necessary to track any weakening of the average consumer's pricing power.

The study's multivariate linear regression of airfares, however, indicates that the negative pressure that Southwest and other LCCs have exerted on prices has actually increased since 2012, despite the industry mergers. The most compelling reason for this apparent violation of classic firm-size theory is found to lie in the recent large decline in jet fuel prices that saw every player make enormous windfall profits. Competing was easier.

The development of the LCC concept and the outside emphasis on Southwest Airlines is only understood by examining the fundamentals of both this airline's history and business model in the context of airline deregulation that began in 1978. The necessity of Southwest, as the largest LCC by far, to keep the deregulated environment working for the greatest consumer interest, motivates this study.

II. Background

REGULATION AND DEREGULATION OF THE U.S. AIRLINE INDUSTRY

In 2015, nearly 700 million domestic passengers were carried by U.S. commercial air carriers. The industry is a far cry from the 1925 Kelly Air Mail Act that first established the government's authority to regulate airfares and allocate distinct air mail routes to private carriers. The modern, post-World War II era started when the Civil Aeronautics Board (CAB) separated the industry into sixteen mainline "trunk" carriers that provided intercity service, and a multitude of supporting regional "feeder" carriers that fed service to the main trunk route cities. This initial model distinctly mirrored the layout of the railroad industry (Davies, 2011).

The CAB also possessed the power to approve fares, which it did so that revenues would be "sufficient to ensure the performance of the service" to the point where the airline system was "of a character and quality required for the commerce of the United States" (Civil Aeronautics Act of 1938, as cited in Bornstein, 2007). The economic result was that the independent airlines competed for customers through frills such as meals and increased capacity and more flights.

By the 1970s, jet travel had become normal and acceptable. The Boeing 747 ushered in the wide-body era, which enabled mass-scale air travel. Newer airports of the time such as Washington Dulles, JFK, O'Hare, Houston Intercontinental, and DFW were equipped to handle massive volumes of new passengers that the regulated era had been discouraging with higher prices. With the technology essentially ready and the trunk airlines having stagnated with inefficient service, the swift passage of the Airline Deregulation Act in October 1978 brought the network and price controls of the CAB to an end. The U.S. swiftly became the first large-scale, free market for air transport in the world.

When Braniff Airways, one of the original trunk carriers, started service on 300 new

routes within the first two years of deregulation, its high-cost business model sent the airline arcing deep into the red, resulting in its full demise in 1982. It was the first casualty of deregulation. Likewise, numerous upstart airlines launching in the 1980s did the same. Midway, People Express, and Presidential Airways all tried to take on a low-cost model and rapidly expand, but they were confronted with unsustainable corporate leverage and an oil price shock related to the first Gulf War. These factors saw some liquidated and some acquired by the trunk carriers, known by this time as “legacy” carriers (Davies, 2011). These relatively unsuccessful Low-Cost Carriers (LCCs) that ceased around 1990 comprised the first wave of what would become a disruptive business model in the industry.

In the early 1990s, as deregulation caught up to and extinguished Eastern Airlines and then Pan-American, the domestic market was briefly without the pressure of any formidable low-cost carrier. It had been nearly 15 years since the beginning of deregulation, and most of the upstarts of the post-1978 era had all folded by the beginning of the decade (Davies, 2011). A 180% spike in the price of oil due to the Gulf War in 1990-1991 was also occupying the rest of the industry,¹ and the second generation LCC’s such as ValuJet (to become AirTran), JetBlue, and Spirit had yet to be conceived. For a short period the competitive landscape had almost stagnated, with the remaining seven legacy airlines dominating the domestic market. But such an arrangement was not to last.

THE HISTORY OF SOUTHWEST AIRLINES

Southwest Airlines was incorporated in 1971 by lawyer Herb Kelleher and businessman Rollin King, with a strategic model closely mirroring the successful west-coast discounter Pacific Southwest Airlines (PSA) (Davies, 2011). 1971 was still seven years before deregulation, but in the pre-1978 regulated era the CAB’s authority to set fares applied only to airlines flying inter-state routes. Intra-state routes were not fare-regulated. Operating solely within California, PSA essentially created the first LCC in the “deregulated” airline market within the large state, and Kelleher and King saw the opportunity to do the same in Texas and named their airline Southwest. Southwest started intra-Texas service with flights between Dallas, Houston, and San Antonio, using three new Boeing 737s. These are the same and only type of aircraft the airline uses today.²

By virtue of a quirk of timing, Southwest was able to establish itself early as the main carrier at Dallas Love Field, the older but significantly more convenient main airport in the Dallas area. In 1964, the neighboring cities of Dallas and Fort Worth had come together to plan a joint commercial hub after the FAA stated it would not fund two major airports so close together geographically, which led to the creation and opening of DFW international ten years later in 1974 (Speaker, 2016). In the agreement, all carriers originally at the local airports such as Dallas Love had signed their intent to transfer their service to the new DFW airport upon its completion, and when that occurred, legacy carriers such as American and

¹ See historical jet fuel prices at <https://www.eia.gov/dnav/pet/hist/LeafHandler>

² Company-sponsored history of Southwest is available at <http://www.swamedia.com/channels/By-Date/pages/history-by-date>

Continental moved. This left Love nearly vacant except for a small airline that had been created in 1971 after the others had signed their mutual intent to move to DFW (Speaker, 2016). This small airline was Southwest Airlines; since that era Southwest has expanded greatly from its base at Love over the objections of the legacy carriers at DFW. Legislative prohibitions of flights from Love to states beyond those bordering Texas, originally intended to protect DFW, were slowly chipped away by Southwest until they fully expired in 2014, allowing fully liberalized service from the airport (Speaker, 2016). Southwest today is the major carrier at Dallas's convenient Love Field, a unique service advantage.

With its business model initialized and an advantageous location free of space constraints as its home base, Southwest embarked on the incremental expansion that it has relied on to this day. By the time of the Airline Deregulation Act they held a strong platform on which to move into the US-wide deregulated era.³ Crucially, throughout its formative years Southwest expanded through targeted and modest acquisitions of Muse air in 1985 and Morris air in 1992. These engagements helped build the airline's fleet and employees, and especially its route network in the central and southwestern swaths of the country, moving into the broader regional market (Davies, 2011).

The young Southwest Airlines also differentiated itself from many of its peers by maintaining a good line of credit with its bank financiers. This occurred at a time when airlines of all different types of business strategies from Pan American to low-cost startups were declaring bankruptcy and parking their aircraft. Through the 1980s and 1990s Southwest was actively engaged in stable debt financing and several rounds of equity offerings, stock splits, and buybacks that kept it on a stable growth path. The airline went public in 1971 and has not made a public share offering since 1992.⁴ Southwest was deliberate and cautious with its potential, building up carefully.

THE SOUTHWEST AIRLINES BUSINESS MODEL

The pioneering Pacific Southwest Airlines (PSA) model of investing in a positive, helpful, and even happy, corporate culture to drive customer loyalty alongside low fares was distinctively copied by Kelleher and King, who took the niche idea and brought it mainstream with Southwest. The two recognized from the very beginning that they needed to invest in their employees' well-being in order to build positive rapport with passengers and likewise to keep labor productivity high per block hour (Davies, 2011). The methods included paying pilots and gate workers industry-competitive wages and cultivating a culture of strong team effort. Pilots often helped cleaning the aircraft and loading bags during turnarounds, for example (Vasigh et al., 2013). Flight attendants were dressed in flashy uniforms that the flying public would remember, and they were encouraged by all corporate levels to be humorous, helpful, and happy (Davies, 2011). Creating a positive culture was

³ Southwest city service start dates are available at <http://www.swamedia.com/channels/city-start-up-dates/pages/when-did-we-arrive>

⁴ Southwest's recent corporate history is available at <http://investors.southwest.com/investor-resources/investor-faqs/southwest-airlines-faq>

essential in creating a differentiated air travel product, as PSA had already demonstrated that happy employees will lead to happy passengers, which alongside low fares will lead to a loyal customer base that captures market share from the competition. The key was to gain repeat customers. An additional competitive advantage of this new, positive corporate culture was that it was extremely hard to replicate, especially by entrenched legacy carriers which had long-developed their own corporate stigmas that were not based squarely around personality and friendly service.

Southwest's culture was only a start for a new airline business paradigm. Fundamentally, the low-cost airline needed to seek a "happiness advantage" to retain customers because the revenue stream of low fares could not support extra service or tangible enhancements, normally known as frills. Thus, Southwest developed a no-frills service model. At the time, Southwest's no frills meant peanuts and a soda as opposed to the full hot meal and beverage service served on legacy carriers. Through the 1990's no frills was a main element of holding the cost base low. Ultra-Low Cost Carriers (ULCC) of the present, such as Spirit, Frontier, and Allegiant, take no frills to the next level with zero free services or special benefits (Vasigh et al., 2013).

Beyond no-frills, Southwest was simultaneously the pioneer of more efficient fleet planning and aircraft utilization through two simple tactics: a standardized type of aircraft and minimum turnaround times. Except for a brief early lease of two 727s, Southwest has operated only Boeing 737 aircraft. The benefits of using a common fleet are lower maintenance costs as they do not need to invest in anything besides 737 parts and support, lower network management costs as every Southwest aircraft can fly every Southwest route, and lower pilot and flight crew costs as every Southwest pilot and flight attendant can fly on every Southwest airplane (Vasigh et al., 2013). This allows for a unified and flexible operation throughout the airline's operations; there is no need for sub-dividing the firm's operations by aircraft type and route as legacy airlines must do. This model has been copied by successful LCCs around the world, and today the largest LCC's outside the US: Ryanair, EasyJet, and AirAsia, fly exclusively 737s, A320s (the popular Airbus competitor in the short-haul market), and A320s, respectively.

Southwest's fleet commonality was combined with the active pursuit of higher aircraft utilization in order to ensure the 737s all generated the most revenue. Sitting on the ground, an airliner is essentially a dead asset. Southwest accordingly created a carefully scripted turnaround procedure to minimize time at the gate. They relied on open seating for quick boarding and de-planeing and on not carrying underbelly cargo (apart from checked luggage), which requires time to sort and load. With a simple equation of quickly offloading passengers, picking up the trash, and loading passengers, the airline developed an advantage over legacy competitors by squeezing more daily short-haul flights from each 737 (Vasigh et al., 2013). Flying Southwest today, the time efficiency is still unmistakable.

Southwest's second technical breakthrough for its airline business model was the creation and support of a non-hub-and-spoke route network often based on service to secondary airports. Legacy carriers of the deregulated era built up massive hub-and-spoke operations around focus airports in a drive to maintain wide service while seeking economies of

scale, with the result that a traveler could move from any US city to any other US city on any carrier with one change of aircraft (Vasigh et al., 2013). Today, Atlanta, Chicago, and Dallas are massive hubs for Delta, United, and American respectively. Southwest adopted a different approach, based on a “point-to-point” network in which roughly 8-10 independent, nonstop Southwest routes originate from each airport served, catering to local demand (Lordan, 2014). While such an arrangement does not seek economies of scale by reducing overhead at large hubs, spacing the flights throughout the day frequently avoids the idle worktime between the timed rushes that must be scheduled at the legacies’ hubs to connect passengers efficiently across a hub-and-spoke network. Though the point-to-point system of Southwest does not preclude its use of “focus cities” at several airports to connect passengers (such as Dallas Love and Chicago Midway), Southwest’s reliance on hubs is significantly less than the reliance of legacy carriers on their much larger hubs and is centered on secondary, smaller airports.⁵ Again, a low cost-base drove part of this model as these secondary airports often have lower costs (gate and landing fees, service rentals, etc.) and less congestion than primary airports, allowing Southwest to conduct its trademark fast aircraft turnarounds (Vasigh et al., 2013). The short-haul operations of Southwest also eliminate any need to connect travelers to long-haul flights on aircraft that often require the longer runways of primary, congested airports. Airports such as Love, Midway for Chicago, Islip for New York, Providence or Manchester for Boston, Oakland or San Jose for the Bay Area, and Ontario or Burbank for Los Angeles are all secondary airports that are focus airports for Southwest. The point-to-point network configuration relying on secondary airports allowed Southwest to couple lower service costs with still high-yielding routes, a key part of their improved business model.

By being transparent about the main pillars of its business strategy to customers, Southwest attracted revenue and built a dominant position in the US short-haul market. Attempts by legacy carriers to create their own Southwest-style subsidiaries have all failed. Newer US LCCs have found their own relative success in alternative niches that do not directly compete in Southwest’s particular short-haul role, and accordingly they have not grown larger than one-quarter the size of Southwest.⁶ The second-largest US LCC, JetBlue, operates predominantly longer-haul coast-to-coast and leisure routes to and from Florida and the Caribbean, with significantly less presence than Southwest in the center of the country. Likewise, JetBlue also operates significant hubs in JFK, Boston Logan, and Fort Lauderdale. Virgin America, a financially weak relative newcomer, likewise focuses on long-haul domestic routes.⁷ Spirit, Frontier, and Allegiant Airways all operate with a separate ultra-low-cost business model, “unbundling,” that targets extremely price sensitive leisure customers by eliminating any costly frills and raising significant revenue through ancillary items such as baggage and amenity fees (Rosenstein, 2013).

While the current state of the domestic US LCC market is characterized by several

⁵ Southwest’s Route Map is available at https://www.southwest.com/flight/routemap_dyn.html

⁶ See footnote 10

⁷ JetBlue and Virgin America’s route maps can be viewed at <http://www.jetblue.com/WhereWeJet/> and <https://www.virginamerica.com/cms/airport-destinations>, respectively.

not-inconsequential carriers, there are none that compete directly with Southwest's short-haul, point-to-point domestic turf using all of Southwest's fundamental low-cost principles including a positive corporate culture. One additional carrier, often overlooked as a stand-alone "specialty" carrier, is Alaska Airlines. Though Alaska has a legacy history, their quiet expansion in the lower 48 since the mid-2000s and their strong position on the west coast indicate that they should be included in any domestic airline analysis.⁸ Alaska is not a traditional LCC; however, their small size (just behind JetBlue) increases their agility, their single-type fleet lowers their costs, and their geographic branding is distinctive. Despite these distinct LCC characteristics, Alaska is historically not seen as a disruptor. The April 2016 purchase of small Virgin America by Alaska has thrust Alaska to the fore of the competitive debate, even though the combination remains only marginally larger than JetBlue.

An additional element underscoring the novelty of Southwest's business model has been that dominant legacy carriers have stopped trying to emulate it themselves by virtue of their inability to reduce their higher costs bases and build higher employee productivity. Over the course of the 1990s and early 2000s, the majority of the legacy carriers opened and closed LCC subsidiaries, each try never operating more than four years: Continental with Continental Lite, Delta (with Delta Express and Song), US Airways (with MetroJet), and United (with United Shuttle and Ted). While the legacies could easily remove frills and amenities and use a single type of aircraft, their mainline, union crews did not work in the same upbeat and hardworking corporate culture as Southwest (Vasigh et al., 2013). Labor productivity was weaker and aircraft turnarounds less efficient. When the legacies entered bankruptcy from high costs they often folded their LCC's into their main brand, leaving the market for Southwest.

IMPACT ON AIRFARES BY SOUTHWEST: PREVIOUS RESEARCH

One primary impact of Southwest's radically different business model, built on the foundation of a low cost-base supported by a strong and positive corporate culture, was its ability to influence the fares of any market that it entered. It could cut its prices and pull in price-conscious consumers while remaining profitable and retaining the productivity of its labor force. By the turn of the millennium, Southwest was boarding more than 60 million passengers annually, more than any other non-legacy US carrier has carried through the present day.⁹ When Southwest entered any given market, its low prices placed market pressure on the legacy carriers to lower their fares lest lose customers. The result was a broad decline in fares. In 2001, Steven Morrison published an influential paper that calculated the fare reductions caused by Southwest's presence, or potential presence, in any given market in 1998. Potential competition from an airline in a market is present when the airline serves the two endpoints of the route. With the infrastructure on both ends, input

⁸ Alaska's route map and company-sponsored history are available at <https://www.alaskaair.com/content/route-map.aspx> and <https://www.alaskaair.com/content/about-us/history/history-by-decade.aspx>, respectively.

⁹ Southwest annual report summaries are located at <http://www.swamedia.com/channels/By-Category/pages/yearend-summary>

costs to begin flying a new route are minimal. When that airline on both ends is a LCC such as Southwest, incumbent carriers will often lower their fares in an effort to deter the LCC from realizing the potential profit to be had from entering and further undercutting the incumbents (Brueckner et al., 2012). Morrison's study found that in the aggregate, airline passengers were saving \$12.9 billion total from the reduced fares of Southwest combined with the reduced fares of the legacy airlines competing with Southwest.

Eleven years after Morrison's work, John Kwoka et al., traced the initial presence of Southwest to a drop in legacy airlines' fares by 24.0% over 2009-2010 (Kwoka et al., 2013). This compares to a 3.4% drop in fares when a second legacy carrier enters to compete against another dominant legacy carrier. They also found that the effect on fares as a result of Southwest's presence varied with Southwest's own market share, as average fares decline by 9.7-10% when Southwest moved from a 25% to a 50% market share, but by only 5.4% when the airline moved to a 75% market share. In the broader LCC framework, they ultimately determined that the fare reduction marginally diminishes as LCCs gain market share, bottoming out when the LCC possess 65.7% of the market. Similarly it has also been substantiated that the presence of an LCC in a market increases the incumbent carriers' probability of codesharing with each other on the same aircraft, a form of revenue cooperation (Goetz and Shapiro, 2012; Brueckner et al., 2012).

While the "Southwest effect" has been shown to be true and effective at reducing fares over the majority of the airline's history, more recently it has been observed that the effect is potentially weakening. Wittman and Swelbar reported as much in a study in their 2012 paper, "Evolving Trends of U.S. Domestic Airfares: The Impacts of Competition, Consolidation, and Low-Cost Carriers." Their study used a linear regression to explore the relationship between average ticket prices for flights originating at a given airport, the average distance covered by such flights, and the presence of Southwest and other LCCs at the airport. They ran regressions separately for the years 2007 and 2012 and compared the results. As one could expect, a significant positive effect was found for average itinerary distance. The longer the flight, the higher the price. Also, they found a significantly negative effect of the presence of Southwest at an airport on mean fares, thus substantiating the existence of the "Southwest effect." The presence of Southwest at airport reduced the airport's mean fares by \$36 in 2007 and \$17 in 2012.

Though not conclusive, this change in the Southwest-specific drop in average fares from 2007 to 2012 indicates that the impact of Southwest to lower market prices is diminishing. To test whether that trend has continued, or whether the Southwest effect is still holding, this study includes a model of the year 2015 that can be compared to results for earlier years.

Southwest in 2015 carried 144 million passengers, roughly four times the 35 million carried by JetBlue, its closest LCC competitor, and ahead of United's total of 140 million.¹⁰

¹⁰ Passenger numbers for these three airlines can be observed at <http://www.swamedia.com/releases/southwest-airlines-reports-record-fourth-quarter-and-annual-profit-43rd-consecutive-year-of-profitability?l=en-US>, http://otp.investis.com/clients/us/jetblue_airways/usn/usnews-story.aspx?cid=981&newsid=32404, and <http://www.wsj.com/articles/delta-surpasses-united-for-no-2-airline-spot-by-traffic-1452626778>, respectively.

With such a significant presence that it is operating at about the same size as the consolidated legacy carriers it once sought to disrupt, any decrease in the ability of its presence to lower fares in a given market lends credence to the idea that Southwest may be drifting away from its low-cost core. According to the *Los Angeles Times* and *USA Today*, over the course of January and February 2016, Southwest raised its general fares by a cumulative \$22. These incremental price hikes were immediately matched by American, Delta, and United. However, while Southwest's fare hikes were matched by competitors, when Delta and JetBlue independently raised their fares by small amounts earlier in February, they were forced to reverse course after the rest of the industry did not go along. Likewise in September 2012, a general \$10 hike by Southwest was also immediately matched, indicating that the airline's blossoming dominance over pricing strategy could be solidifying (as in 2012 Southwest also carried four times the number of passengers of its nearest non-legacy competitor JetBlue).¹¹ Thus it will be informative to investigate whether or not the Southwest effect has weakened in recent years.

III. Empirical Analysis

In this section a multivariate econometric model to test the effect of LCCs on domestic airfares is developed and results are presented.

DATA SOURCE

Wittman and Swelbar used data based on government datasets, adjusted by data from a private aviation data company to account for the itinerary characteristics of each ticket. In context of this update for 2015, however, the ability to access a private consultancy to clean and fully standardize the data is limited. Thus the data used are straight from the databases of the Department of Transportation's Bureau of Transportation Research (BTS). The main data on airport mean airfare were acquired through BTS's DB1B market dataset that records a 10% sample of the itineraries purchased on domestically-operating airlines. The data therein include the operating carrier, origin, destination, distance, and total fare paid and allow for the examination of the average itinerary price originating from the top 418 domestic airports. The most recent time period for which this full set of cross-sectional data is available is for Q3 2015 and these data are used for this econometric model.¹² The econometric model was run using the software STATA.

¹¹ Newspaper articles that highlight all of these fare shifts are located at <http://www.latimes.com/business/la-fi-airlines-match-third-airfare-hike-of-2016-20160223-story.html>, <http://www.usatoday.com/story/money/2016/02/22/airlines-attempt-raise-fares-fifth-time-year/80759150>, <http://www.latimes.com/business/la-fi-airlines-match-third-airfare-hike-of-2016-20160223-story.html>, and <http://articles.latimes.com/2012/sep/14/business/la-fi-southwest-fare-hike-20120914>.

¹² Data is publically available at http://www.transtats.bts.gov/databases.asp?Mode_ID=1&Mode_Desc=Aviation&Subject_ID2=0

MODEL

The model has the following structure:

$$\text{Average fare} = f[\text{average itinerary distance; LCC dummies; sum of major carriers at the airport; dummy for vacation destination of the airport}]$$

Average Fare: The average fare represents the dollar amount that is paid for the itinerary. It includes fuel surcharges and taxes, but importantly it does not include ancillary fees such as baggage fees. Q3 data are used in all cases to control for seasonality.

In this study, average fare is analyzed over three levels of analysis and across three time periods: 2007, 2012, and 2015. The first analysis of average fare is at the airport level, the second at the airport-pair level, and the third at the city-pair level. The airport-pair and city-pair analyses attempt to uncover the effect of LCCs' presence in particular passenger markets, and differ from the airport-only analysis by going beyond simply evaluating the effect of a LCC's presence at any airport. Fare competition most likely will be heightened between competitors on identical routes between airports and between cities. The study will examine these three levels of competition in turn for each year.

Average price is used in the present model; however, it would be additionally informative to analyze the impact of LCC prices on the specific ticket prices of legacy airlines rather than using the average over all airlines. Doing so would allow the identification of fare decreases from incumbent carriers that are a specific result of LCC competitive presence. This is beyond the scope of the present study which follows the methodology of recent studies that have used average fare.

Average itinerary distance: The average itinerary distance, measured in flight miles from point of departure to point of arrival for flights originating at each airport. An itinerary refers to one full trip to a destination, allowing intermediate connections. As mentioned, a longer route traditionally implies a higher price. Expectations for the 2015 model are that average itinerary distance will display a significantly positive impact on airfares, as it did for Wittman and Swelbar.

LCC-specific dummy variables: Dummy Variables for LCCs are set with 1 equal to presence and 0 equal to no presence at the airport. The carriers are Southwest, JetBlue, Spirit, Frontier, Allegiant, Virgin America, and Alaska. For the carrier-specific dummy variables, it is expected that their effect will each be negative; the operating presence of any of six LCCs at an airport will generally lower the average price a passenger will pay to fly from that airport. Although the presence of Southwest is expected to affect average price negatively, if the Southwest effect is decreasing, the effect of the presence of Southwest at an airport should be relatively small. It is appropriate to expect that this decreasing trend probably has continued to 2015 from the 2007 to 2012 period, given the continued growth

of the airline, the further consolidation in the domestic LCC and legacy market, and Southwest's general price increases between 2012 and 2015. The additional market share taken by Southwest via its 2011-2014 merger with AirTran will also give it competitive power to reduce fares less. The presence of Alaska and Virgin America are included in later route and city-based regressions, with identical regression equations to Wittman and Swelbar's used to start. It is necessary to remember that the amount paid per passenger in ancillary fees is proportionately much higher on ULCC airlines. Big fare reductions are not all consumer savings.

Sum of major competitors serving an airport: The number of major airlines serving a given airport was used in certain regressions. It was not included in Wittman and Swelbar's model. Sum is predicted to vary inversely with price but not to be extreme, considering that as the number of airlines at a given airport initially increases, consumers should face more options for flights, but from then on competition's marginal increase should theoretically see the size of the reduction in average airfare diminish. It must also be noted that the total number of passengers using a given airport could be a function of the average airfare, and thus it is not included over endogeneity concerns.

Vacation destination of the airport: This variable is included to capture the lower average prices associated with flights that include Florida and Nevada. The decision to use these states as indicators of vacation flights follows Wittman and Swelbar's model.

RESULTS: FARE IMPACT BY AIRPORT

The data on ticket fares can be reported in several ways. This section discusses results when the dependent variable, average fare, refers to the fares for all flights originating at a given airport. Table 1 shows results of regressions for the years 2007, 2012, and 2015. The first two columns, for 2007 and 2012, show results similar to Wittman and Swelbar.

In 2007 both Southwest and JetBlue are associated with strong negative impacts on average prices, with Southwest's \$63 indicated reduction demonstrating that the carrier was the most influential in lowering average route fares. 2007 was just before the start of an industry-wide wave of consolidation. At the time the legacy carriers were weak, trying to cut costs that had allowed LCC's to extract substantial pricing advantages over the course of the early 2000s. Minor disparities between the results and those of Wittman and Swelbar remain present, but in the 2007 data the order of the carriers and their proportional impacts relative to each other are analogous. Southwest's impact is largest.

TABLE 1

Regression Results, Impacts by Airport¹³

Variable	Coefficient		
	2007 Q3	2012 Q3	2015 Q3
Constant	74.99** (11.894)	134.64** (9.920)	112.08** (20.970)
Air Itinerary	0.16** (0.007)	0.13** (0.006)	0.17** (0.013)
Vacation	-7.29 (20.557)	-35.85** (17.415)	-30.98 (32.627)
Southwest	-63.32** (16.097)	-43.30** (11.727)	-31.36 (22.199)
JetBlue	-45.55** (18.266)	-37.76** (13.972)	-39.00 (25.430)
Spirit	-30.27 (24.761)	-30.83* (16.489)	-47.77 (32.560)
Frontier	20.15 (13.786)	16.43* (9.735)	17.34 (27.595)
Allegiant	-33.84** (13.879)	-50.50** (9.702)	-56.74** (17.146)
AirTran	-20.83 (16.824)	-9.48 (14.132)	n/a
N	465	446	418
Adjusted R ²	0.51	0.57	0.32

*statistically significant at 10%, **statistically significant at 5%

In 2012, all of the LCCs had a statistically significant impact on airport average prices, except for AirTran. In addition, a vacation location had a significant negative impact on fares. Spirit's low fares, modeled on a zero-frills model like those of Allegiant, exerted a modest price reduction of \$31 on the fares of a given airport in 2012. At the time, Bill Franke, the chairman of Spirit and former CEO of America West, had not yet decided to sell his stake in Spirit and purchase Frontier to also develop it into an ULCC (Nicas, 2013). This would occur a year later, and is the most plausible reason that the presence of Frontier in a given airport was associated with a fare increase in 2012. Both Southwest and JetBlue had significant, downward impact on airport ticket prices, with Southwest's \$43 associated reduction surpassing JetBlue's \$38. The reduction of Southwest's indicated by this regres-

¹³ The means for the regression variables for 2007 are: average fare (269.66), average itinerary distance (1327.25), vacation (0.05), Southwest (0.14), Jet Blue (0.10), Spirit (0.04), Frontier (0.17), Allegiant (0.12), AirTran (0.12). Means for the other regressions are comparable.

sion was significantly larger than the \$17 reduction that Wittman and Swelbar recorded. Allegiant again demonstrated the greatest negative impact on fares, with a \$51 dollar reduction significantly higher than the \$29 reduction recorded by Wittman and Swelbar in the same year. The differences may be partly due to differences in the data used, as the firm that standardized Wittman and Swelbar's data most likely controlled for additional factors given the peculiarity of the airports Allegiant serves.

Results of the regression for 2015 are shown in the third column of Table 1. While the resulting coefficients displayed the expected negative values and the relationship between average itinerary distance and mean fare was highly significant and positive, the only other variable significant in determining the mean fare out of a given airport was the presence of Allegiant Air. The adjusted R^2 term shows that 32% of the variations in average itinerary fare from an airport can be explained by these particular variables.

While the 2015 results are not compelling with their modest significance levels, they do pass the F-distribution test of joint significance (seen by the "Prob > F = 0.000"), indicating that they are collectively significant in determining the mean airfare when flying out of a given airport. Allegiant's presence was associated with a reduction in mean airfares by \$57. For Allegiant, such a reduction is expected considering that its no-frills, non-connecting vacation model that focuses on keeping its cost base low allows for lower fares that raise price competition at the airports where it serves (which are mostly secondary airports). The remaining carriers, though not significant, display negative price influences associated with their presence at an airport, with Southwest indicating a fare reduction of \$31 and JetBlue a \$39 decrease. Southwest's presence from this piece of data proposes that the airline was generally acting to lower airport prices slightly less than the other LCCs in 2015, but a better conclusion requires upcoming analyses.

A very interesting question is why the statistical significance of the results for 2007 and 2012 are strong but appears weaker in 2015. The most plausible explanation for such behavior is based on the two-thirds reduction in the price of jet fuel since the summer of 2014. With lower input prices, all carriers have seen record profitability such that the presence of any carrier at an airport in 2015 may not necessarily have been associated with as much of a unique change in fares than in previous periods of tighter margins.¹⁴ Because fares have not seen declines proportional to the recent decrease in fuel costs, the presence of particular carriers at an airport may possess less of an effect at an airport because the carriers are not as susceptible to losing money even when LCC competition is present. Essentially, the larger profit margins give each carrier a buffer from setting fares that finely track their cost bases. There could also simply be more airport fare fluctuations that are not linked to a specific carrier and a result of greater disparities between the regions the airports serve, such as varied airport taxes. This would imply that airlines possessed less of a unique pressure on airfares at a given airport in 2015.

¹⁴ 2015 margins were up 6.8% on average for legacy carries and 6.2% for LCCs since 2014. See Hazel et al., 2016.

RESULTS: FARE IMPACT BY ROUTE

The analysis above is based on average fares over all flights originating at given airports. But head-to-head competition normally occurs on certain flights between common origins and endpoints. A more detailed analysis comes by examining mean fares between unique endpoints (Refer to Table 2).

The dependent variable in this case is the mean fare on an itinerary over two unique endpoint airports. In the Q3 2015 data, there were 68,213 unique domestic routes contained in the DB1B database. All average fares in the original data that displayed prices of less than \$20 were removed in order to eliminate flash deals and other super-low-priced marketing deals. This regression was built with the same DB1B database and data, in the statistical programming language R.

In this regression, all the variables remained the same, except Vacation was discarded. Within one route, whether or not that route is to or from a vacation destination will not have a material impact on the carriers' differences in fares. All carriers' fares will reflect a price shift if the route is a vacation route. The average itinerary distance is the average distance flown between two airports, which depends on connections.

Results from Q3 2007 indicate that the LCC effect of reducing fares on airport-pair routes was healthy and present and included a relative strong impact from Southwest. Southwest's presence was associated with a \$34 reduction in mean route fares, whereas JetBlue was associated with only a \$24 reduction. Frontier, at the time a smaller regional legacy carrier, was expectedly associated with an increase in fares (\$29), while Allegiant, again the strongest force at pulling down fares, was associated with a \$99 reduction in fares on the routes that it operated. Spirit was interestingly associated with more modest fare declines of \$17, and Alaska, which in 2007 was a large regional carrier similar to Frontier, was expectedly associated with a fare increase of \$36 over the routes that it operated. Virgin America was still being founded. The adjusted R^2 of the model in 2007 demonstrated that it would explain 33% of the variation in average route fare with these variables, in the same range as that of 2012.

TABLE 2

Regression Results, Unique Domestic Airport-Pair Routes

Variable	Coefficient		
	2007 Q3	2012 Q3	2015 Q3
Constant	203.65** (0.910)	219.7** (1.020)	222.6** (2.300)
Air Itinerary	0.084** (0.0005)	0.096** (0.0005)	0.099** (0.001)
Sum of major carriers	-27.44** (0.841)	-28.34** (0.725)	-18.82** (1.711)
Alaska	35.89** (3.797)	44.25** (3.536)	-5.19 (7.157)
Southwest	-33.82** (2.884)	-7.62** (2.701)	-24.00** (6.006)
JetBlue	-24.25** (4.262)	-30.55** (4.229)	-33.62** (8.692)
Spirit	-16.94* (9.542)	-34.09** (7.435)	-45.55** (13.260)
Frontier	28.98** (4.193)	3.08 (3.410)	21.43* (12.910)
Allegiant	-99.05** (10.429)	-135.1** (7.545)	-157.9** (14.050)
Virgin America	n/a	79.67** (13.850)	41.17* (25.650)
n	70,163	70,784	68,213
Adjusted R ²	0.33	0.34	0.10

*statistically significant at 10%, **statistically significant at 5%

With Q3 2012 data, the results demonstrate that in that year there was still a solid degree of fare reduction when LCCs were present on an airport-pair route. Southwest was associated with an \$8 fare reduction, and JetBlue, Spirit, and Allegiant were associated with reductions of \$31, \$34, and \$135 respectively. Frontier was associated with a \$3 increase in fares as this was prior to its rebranding as a ULCC. Alaska (now significant) was associated with a \$44 fare increase. The addition of an extra major carrier to a route pair was associated with a \$28 fare decrease. All of the variables were significant except for the impact of Frontier.

The reduction in the Southwest coefficient from 2007 to 2012 is the trend that led Witman and Swelbar to conclude that the Southwest effect might be diminishing.

Table 2 shows that, in 2015, the Southwest effect has recovered to its 2007 levels. In that year the presence of Southwest on an airport-pair route is associated with a reduction

in fares by \$24, which is relatively less than the JetBlue's, Spirit's, and Allegiant's associated reductions of \$34, \$46, and \$158 respectively. Frontier, two years after its restructuring as a ULCC, strangely shows that its presence increases mean route fare by \$21, and Virgin America has a similar effect but with a magnitude of \$41. The addition of an extra major air carrier to a route is associated with a \$19 average fare reduction. These airport-pair specific results indicate that Southwest's presence in a market is not associated with as large a reduction as JetBlue or any of the other three ULCC's (Spirit, Frontier, and Allegiant). Allegiant's very large value is most likely related to the fact that it directly competes with few other carriers on its routes, being almost exclusively oriented to a leisure market between rural areas and Florida and Las Vegas, and it sees significant income for its size from ancillary fees (baggage fees, seat selection, beverages, entertainment fees).^{15,16} The adjusted R² demonstrates that the particular set of variables accounts for 10% of the variance in route fares.

These route data indicate that since 2012, the effect of the presence of Southwest on a route has actually increased; giving credence to the notion that price competition from the airline might be on resurgence today. With many more highly significant coefficients, these results are more credible than those from the airport-level analysis.

While through the era of large industry mergers that lasted until 2013, the route fare data from 2012 demonstrate that the price effect of Southwest and other LCCs was decreasing, this has not appeared to continue through 2015, when the relative price reductions increased. A significant external factor over the period since 2012 has been the 2014-2016 fall in oil prices, which has saved the airline industry huge levels of input costs, resulted in exceptionally high profits, and perhaps has reinvigorated pricing competition in pursuit of market share. The airport-specific route data seem to indicate as much could be the case. The LCCs can continue to ensure their prices are below market averages. By measuring so many dummy variables and only two truly floating variables (the average itinerary distance and the number of total carriers operating on a specific route) it is necessarily implied that the explanatory power of the model (the adjusted R² value) will not be as high. The point of this analysis is to observe the direction and magnitudes of the dummy variable coefficients, rather than predicting fares.

RESULTS: LEGACY CARRIERS

To place these regressions in context, the Q3 2015 analysis on the legacy airlines was undertaken by replacing the LCC dummies with dummies for the legacy carriers. The results are shown in Table 3. The presence of legacy carriers (American, Delta, United, and Alaska) has significant positive impacts on the mean route fare. This is expected, because legacy carriers are generally associated with higher fares. AA, DL, and UA are associated with increases of average route fare by \$28, \$25, and \$48 respectively. This is compared to the other carriers (essentially all the historical LCC's) that are not included in this model

¹⁵ Allegiant's route map can be viewed at <https://www.allegiantair.com/interactive-routemap>

¹⁶ Baggage and other reservations fees are recorded by carrier at https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/subject_areas/airline_information/baggage_fees/index.

and thus make up the reference group. While Alaska's presence results average fare increases of \$26 compared to the reference group, this can be combined with the findings of the previous regression to conclude that Alaska does not possess the same ability to lower fares over an airport-pair as the pure LCCs do. It is on par with Delta in this linear model. The \$48 reduction associated with an extra major carrier on a route is a result of the positive bias from including high-fare legacy carriers. One more carrier in a legacy carrier framework will result in a larger decrease in fares than a many-LCC market.

RESULTS: FARE IMPACT BY CITY-PAIR ROUTE

By analyzing each airport itinerary routing's average price, price competition provided by flying to and from alternate airports is not included. A market for air travel is ultimately created by participants that want to fly from one city to another city. When a city possesses competing airport alternatives, a rational decision-maker will evaluate the price of airfares at the array of origin and destination airports. In the case of the LCCs, where Southwest pioneered the use of secondary airports to serve primary locations, comparing routes on a strictly airport-for-airport basis fails to capture the entire market. For example, a trip from New York to Dallas can be accomplished by departing Kennedy, LaGuardia, or Newark, and arriving at DFW or Love Field. Were Southwest to offer lower fares from Newark to Love, it would effectively require that American reduce its prices from LaGuardia to DFW, or lose customers to Southwest's flight. Kwoka et al. describe this appropriately as "adjacent airport competition."

To continue analyzing the current state of the "Southwest effect" and the effects of other US LCC's on airfares in 2015, this analysis runs the same regression using city pairs instead of airport pairs to capture the full extent of multi-airport markets. Cities are defined by the BTS and include multiple-airport districts in Atlanta, Boston, Cleveland, Washington D.C., New York City, San Francisco, Los Angeles, Miami and Tampa. The same regression equation is used.

TABLE 3

**Regression Results, Unique Domestic Airport-Pair
Routes for Legacy Carriers, 2015 Q3**

Variable	Coefficient
Constant	221.83** (2.323)
Air Itinerary Distance	.099** (0.001)
Sum of major carriers	-48.30** (3.317)
American	27.82** (6.255)
Delta	25.14** (5.009)
United	47.80** (6.663)
Alaska	25.52** (7.684)
n	68,213
Adjusted R ²	0.10

*statistically significant at 10% **statistically significant at 5%

Results for the three time periods are shown in Table 4. In 2007, city-pair results again followed the trend of being associated with larger absolute price changes due to the operating LCCs. A major difference, however, was the weak impact of Southwest's presence in city-pair markets in 2007. The \$9 reduction associated with Southwest's presence over a city-pair is significantly smaller than the \$34 reduction it was associated with over airport-pair markets in the same time period. It is very possible that the shift from airport-pair to city-pair markets results in Southwest incurring less of a pricing impact because the probability of the existence other LCCs in the larger city-pair markets is greater. 2007 was before Southwest purchased its next-largest LCC competitor, AirTran, whose presence at the time most likely served to keep city fares low because it operated into most major cities. By adding Southwest to a city, this most likely meant that city-pair fares would not move exceptionally lower. Alaska still followed legacy behavior in 2007 in city-pair markets.

In 2012, the price-decrease effect of most LCCs in city-route markets is smaller than the comparable results for airport routes. The major exception to this trend is Southwest, which displayed an \$8 reduction when it was present on airport-routes in 2012 but a \$22 increase in average fares on city-routes. Such a stark contrast is strange, but it emphasizes the fact that the "Southwest effect" had become significantly smaller in 2012. The presence of JetBlue, Spirit, and Allegiant all possess smaller impacts on average fare in city-pair markets than in airport pair markets.

TABLE 4

Regression Results, Unique Domestic City-Pair Routes			
Variables	Coefficients		
	2007 Q3	2012 Q3	2015 Q3
Constant	206.6** (0.981)	222.5** (1.109)	224.2** (2.639)
Air Itinerary	0.09** (0.0005)	0.10** (0.0006)	0.10** (0.001)
Sum of major carriers	-31.33** (1.003)	-32.39** (0.835)	-21.57** (2.096)
Alaska	40.0** (4.216)	47.60** (3.933)	-5.03 (8.439)
Southwest	-9.38** (3.571)	21.51** (3.384)	-0.88 (7.941)
JetBlue	-2.96 (4.925)	-17.28** (5.070)	-24.72** (10.840)
Spirit	-7.51 (9.771)	-12.14 (7.910)	-35.48** (15.070)
Frontier	37.36** (4.677)	10.43** (3.655)	22.63 (14.810)
Allegiant	-85.58** (10.540)	-95.69** (7.711)	-113.6** (15.230)
Virgin America	n/a	64.53** (14.840)	62.32* (31.910)
n	59,710	60,413	58,179
Adjusted R ²	0.33	0.35	0.09

*statistically significant at 10%, **statistically significant at 5%

The 2015 analysis includes 58,179 unique city-pair routes. The effects of the presence of Southwest, Frontier, and Alaska are not significant, but the directions and most of the magnitudes are on par with what was witnessed in the airport-specific route regression. The coefficient of Southwest is significantly smaller, indicating only a \$0.88 decrease over city-pairs where the carrier operates, but without a high significance the conclusions that can be drawn are limited. JetBlue, Spirit, and Allegiant all indicate slightly smaller impacts than before, which can be interpreted to mean that the associated fare decrease with the presence of these LCCs is higher if the airline is presently operating between two specific airports than just between two specific cities. This is intuitive. Alaska and the number of major carriers exhibited similar effects to the airport-pair market data.

Alaska's presence in both airport and city-pair markets indicates that the impact of its presence on fares has moved from an associated increase in 2012 of around \$40-50, to a

much more negative effect on the order of a \$5 reduction. Such a trend was also visible in the airport-pair route data. Such a decline signifies the increasing presence of Alaska as a unique type of legacy whose fares are looking more and more like those of a LCC acting to capture share.

IV. Conclusions

The most compelling 2015 results in this section are from the airport-pair data, because with these data the majority of domestic LCCs have an effect on price that is statistically significant. Southwest is undeniably the most important carrier to examine by virtue of its dominant market position and its history of shaping lower domestic fares via the “Southwest effect.” Accordingly, Southwest’s association with a \$24 decrease in average fares between two airports over the course of last summer and early fall can be compared to associated decreases of \$34 and \$8 in the same time periods of 2007 and 2012, respectively. This indicates that the Southwest effect decreased to 2012 but has since reappeared. While the domestic airline industry changed in 2013, when American and US Airways began the final legacy merger to form the world’s largest airline, this was not a shift that, on its own, would have allowed Southwest to charge relatively lower fares. On the contrary, the re-appearance of the Southwest effect is most likely tied to the major recovery of the industry’s profitability as a result of vastly lower jet fuel costs. With lower fuel prices, Southwest saw an increase in net income in 2015 of 92% in its annual report. The other major carriers also experienced similar windfalls, and this type of exogenous impact on financial sustainability has allowed the carriers more leeway in maintaining fares and market share.¹⁷ Essentially, despite the consolidation in the industry, the fact that all carriers are suddenly flush with cash has seen them become more flexible in competing with one another on fares. The price reductions associated with Allegiant, Spirit, and JetBlue’s presence in markets increased from 2012 to 2015 on all of the regressions run, indicating that the larger “LCC effect” was alive in 2015. However, these carriers still are very small compared to Southwest. Additionally, the decrease in ticket price associated with one extra major carrier decreased in both route regressions from 2012 to 2015. This result can be interpreted to mean that price competition has increased, as a smaller impact by an additional carrier implies that the incumbent carriers are competing more heavily on price than in 2012. Another participant means less of a price effect.

This study has shown that the Southwest effect is again quite strong in 2015. The industry could be observed after major mergers had been achieved and compared to both 2007 pre-merger and 2012 mid-merger data. Also, the comparison of data from both airport-pair and city-pair itineraries took the study a step beyond the airport-specific data that was described in Wittman and Swelbar’s similar analysis.

These conclusions are contrary to what was expected as a result of industry consolidation. Lower fuel prices have led to much higher profits by the legacy carriers, who have

¹⁷ Among other carriers, American, Delta, United, JetBlue, Alaska, and Spirit saw net income increase by 164%, 587%, 548%, 69%, 40%, and 41% in 2015, respectively. See financials at quotes.wsj.com

tried to maintain “capacity discipline.” But these low fuel prices have also given the LCCs the opportunity to compete on price, which is why there has been a resurgence of the Southwest effect. Many more consumers can afford to fly today thanks to the business model of Southwest and smaller LCCs. While recently the industry has heavily consolidated, the impact on its pricing behavior has been affected simultaneously by an enormous drop in input prices that has resulted in substantial improvements in financial performance, and most likely the renewed competitive pressure in the industry. To maintain price competition alongside the greater market shares of the four dominant US airlines, it is essential that the low-cost model remains viable. A broader takeaway from this study is the importance of having a competitive fringe in industries experiencing consolidation. Vigorous competition can be provided by relatively small companies with strong innovation efforts, such as the LCC sector in the airline industry.

The analysis did have certain scope limitations in terms of the extent to which additional variables might have improved the results. Adjusting for market share, airport concentration, and the marketing carrier in the case of codeshared flights, would have provided a more representative analysis of the true effects of each airline. A wider panel of data than just the third quarter of three selected years would furthermore have aided the empirical analysis in clarifying the competitive evolution of each carrier. In addition, analysis of LCC effects on the actual prices of tickets on legacy airlines, instead of their impact on average ticket prices, would be an important subject for future research.

References

- Borenstein, S., and Rose, N. L.** (2007). “How airline markets work... or do they? Regulatory Reform in the Airline Industry.” National Bureau of Economic Research Working Paper No. w13452.
- Goetz, C., and Shapiro, A. H.** (2012). “Strategic Alliance as a Response to the Threat of Entry: Evidence from Airline Codesharing.” *International Journal of Industrial Organization*, 30(6), 735-747.
- Brueckner, J. K., Lee, D., and Singer, E. S.** (2012). “Airline Competition and Domestic US Airfares: A Comprehensive Reappraisal.” *Economics of Transportation*, 2(1), 1-17.
- Davies, R. E. G.** (2011). *Airlines of the Jet Age: A History*. Lanham, MD: Smithsonian Institution Scholarly Press in cooperation with Rowman and Littlefield Publishers.
- Hazel, B., Stalnaker, T., Taylor, A., and Usman, K.** (2016). *Airline Economic Analysis 2015-2016*. New York, NY: Oliver Wyman. <http://www.oliverwyman.com/insights/publications/2016/jan/airline-economic-analysis-2016.html#VxBPcvkrKM8>

- Kwoka, J. E. Jr., Alepin, P., and Hearle, K.** (2013). "Segmented Competition in Airlines: The Changing Roles of Low-Cost and Legacy Carriers in Fare Determination." Social Science Research Network, SSRN 2212860.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2212860
- Lordan, O.** (2014). "Study of the Full-Service and Low-Cost Carriers Network Configuration." *Journal of Industrial Engineering and Management*, 7(5), 1112.
- Morrison, S. A.** (2001). "Actual, Adjacent, and Potential Competition: Estimating the Full Effect of Southwest Airlines." *Journal of Transport Economics and Policy*, 31(2), 239-256.
- Mouawad, J.** (2016). "Airlines Reap Record Profits, and Passengers Get Peanuts." *The New York Times*, February 6, 2016, p. A1.
- Nicas, J.** (2013). "Bill Franke's Indigo to Acquire Frontier Airlines." *The Wall Street Journal*, October 1, 2013.
- Rosenstein, D. E.** (2013). "The Changing Low-Cost Airline Model: An Analysis of Spirit Airlines." *Aviation Technology Graduate Student Publications*. Paper 19.
- Speaker, A.** (2016). "Dallas Love Field." *Airways, A Global Review of Commercial Flight*, edition March 2016.
- Unnikrishnan, M.** (2016). "Parker: U.S. Industry 'unlikely' to Post Losses Again." *Aviation Week and Space Technology*, 178(7).
- Vasigh, B., Fleming, K., and Tacker, T.** (2013). "Low-cost and Start-up Airlines: A New Paradigm." From *Introduction to Air Transport Economics: from Theory to Applications*, Burlington, VT: Ashgate Publishing, 373-399.
- Wittman, M. D., and Swelbar, W. S.** (2013). "Evolving Trends of U.S. Domestic Airfares: The Impacts of Competition, Consolidation, and Low-Cost Carriers." Massachusetts Institute of Technology Small Community Air Service White Paper No. 3 Report No. ICAT-2013-07.

THE IMPACT OF APPEARANCE IN EVALUATING POLITICAL CANDIDATES

By Julia T. Paradis*

In addition to relying on partisan attachments to evaluate political candidates, the proliferation of media outlets and accessibility of social media may encourage individuals to form judgments about candidates based on their physical appearance. Research in psychology and political science have found that attractiveness acts as a heuristic to judge a candidate's character and that candidates that are perceived to be more attractive receive benefit electorally. This present study tests the hypothesis that attractiveness impacts election results and analyzes variation by race and gender. Data were obtained by asking respondents to rate 40 candidates on a 1-to-10 physical attractiveness scale. Some respondents were shown candidates on a computer screen, while others were shown the same candidates on printed paper to test a potential impact of different survey methodologies. Contrary to past studies, the relationship between candidate attractiveness and vote share was found to be statistically insignificant. There were no differences in the mean ratings for men and women and minorities and non-minorities. There also were no differences in ratings based on how the candidates were presented to respondents. These insignificant findings are encouraging by suggesting less of a bias against minorities and women and that vote share may be determined by factors or heuristics other than physical appearance.

Keywords: Elections; Politics; Psychology; Political Science; Attractiveness; Race; Gender; Physical Appearance; Voting; Social Media; Vote Share

I. Introduction

The United States population is inundated with elections. Between local, regional and national contests, it seems that Americans are constantly voting for politicians to represent our interests and goals for the future. In the months leading up to big election days, individuals are exposed to the candidates for whom they will be voting in a variety of ways. While a candidate may come to a voter's home area for a rally or a meeting that the voter could attend, voters are more likely to interact with candidates and the political world through various media outlets and social media (Trent, 1993). With the growth of television and the explosion of social platforms, candidates dominate these channels in the months, days and weeks leading up to an election (Hayes, 2014). Televised interviews, debates streamed online, tweets, Facebook pages, Instagram posts, news coverage, BuzzFeed articles and a multitude of other mediums have become vehicles for candidates to communicate with

*This research would not have been possible without the tireless support of Professor Jeff Gulati. Endless thanks for heightening my interest in all things politics and believing in me when I certainly didn't! Email: julia.paradis@yahoo.com

the public and for the public to talk about candidates (De Landsheer, 2008). In the years, months, weeks and days leading up to a vote, candidates' faces are constantly plastered across all media outlets and social media. It seems impossible for voters to escape constant visual political imaging during voting season.

Public relations and media companies do not just place images of candidates in these many outlets haphazardly. Research in psychology found that when humans see images of individuals, visceral reactions based on perceptions and stereotypes drive subsequent opinions and judgments about that individual (Lippmann, 1922). In politics, due to constant mass media exposure and reduced partisan focus, researchers believed that candidate appearance might have a similar impact on whom voters select. As a result of research and study over the last 30 years, political scientists largely agree that automatic candidate appearance evaluations are often based on how attractive the candidate is perceived to be (Langlois, 2000). Attractiveness acts as a heuristic to that individual's character (Iredale, 2010). Further research has shown that higher attractiveness ratings lead to increased vote share for that individual. Therefore, studies show that better-looking candidates have an electoral advantage over candidates who are less visually appealing (Kelley, 1974).

Despite these continual findings about a verifiable relationship between candidate attractiveness and vote share, the structure of the studies in the literature vary extensively and are inconsistent in questions asked to respondents, the amount of information given about a candidate and the inclusion of key stereotypical identifiers such as race or gender (Banducci, 2008). Many studies also evaluate only low-information elections and assume that respondents will not recognize candidates without actively controlling for this in the data analysis (Praino, 2014). Most importantly, several of these widely regarded studies do not use real politicians and simply have respondents rate images of any individual and claim that they are a politician (Rosenberg, 1986). These inconsistencies challenge the former literature and beg researchers to continue to study the relationship between attractiveness and electoral share in a more comprehensive and standardized manner.

To test the current hypothesis of the relationship between candidate attractiveness and vote share and to control for various structural errors in previous studies, the present experiment examines how college-aged respondents rate Congress members' attractiveness on a scale of 1 to 10. In addition to a test of the wider hypothesis of the relationship between candidate attractiveness and vote share using correlational analysis, the variables of race, gender, candidate recognition and survey methodology will also be analyzed to learn more about how this phenomenon could be better studied and analyzed.

II. Literature Review

PSYCHOLOGICAL RESEARCH ON THE IMPACT OF APPEARANCES IN SOCIETY

Research in psychology dating as early as 1920 found that individuals innately rely on perception-based knowledge and mental images to evaluate others. This observation is due to conditions in the United States in the 1900s where Americans lived on sparsely settled

farms miles away from their neighbors and learned of news through scattered telegraph and newspaper communication. This situation gave ordinary citizens few opportunities in their lifetimes to see famous leaders. Limited access to contact with figureheads forced individuals to compensate for the inability to obtain concrete first-hand knowledge on which they could base their decisions. Instead, individuals used the few visuals of these people that they had and perceptions from them to build a self-created image of that person and who they were. These mental pictures are called a pseudo-environment, or an individual's interior representation of the world inside their head. The pseudo-environment is a key factor to determine an individual's thoughts and actions related to that person (Lippmann, 1922). Over time, continued use of perception-based evaluations to analyze the vast world and form opinions from a remote location on a farm or small town grew into an automatic judgment of appearances that these non-visual societies relied on for decision-making.

Yet individuals no longer have limited access to leaders and information about them; rather, today's society is saturated with an excess of both images and information through innumerable forms of media that is increasing daily. The days of waiting for a book, newspaper, magazine or sporadic television event to stay attuned to the on goings in the world are long over. Since the turn of the 20th century, the amount and frequency of media have grown immensely. The growth of television, the development of visual campaigns through advertising and the explosion of the smart phone and social media produced a day-to-day life saturated with images. With constant and easy access to information available on any subject at all hours of the day, one might logically assume that individuals no longer require their evolutionary skill of using perceptions to create a mental image and opinion of a person based on their appearance.

Despite society's cultural growth, immense amounts of easily available information do not drive individuals to take time to search and verify knowledge to guide their decisions. Though this is a commodity that their ancestors dreamed of, individuals still automatically and constantly navigate the modern world by relying on their pseudo-environment to notice one aspect of a person, place or thing and subsequently create mental pictures using little perceptible information and mental stereotypes to interpret the world (Lippmann, 1922). Recent psychological research shows that the increased amount of visuals in today's society has even amplified this conditioned skill of relying on limited information gained through appearance perceptions, initial impressions, stereotypes and personal preferences to create a mental picture that can be used to interpret the world (Little, 2007).

The tendency to rely on appearance and perceptions to quickly evaluate subjects has been further examined by psychologists in recent years to understand what factors of appearance trigger the formation of a mental picture and subsequent appearance-based judgments. Many found that how a person is perceived is largely a result of his or her attractiveness and that a person's attractiveness is used as an indicator of character (Langlois, 2000). This hypothesis derives from early psychological research that found that what is beautiful is considered to be good. Beauty and more attractive faces are more likely to be associated with more positive qualities and characteristics (Dion, 1972). Beauty appears to be a universal standard; a number of studies found that both attractive children and adults were

evaluated significantly more favorably than unattractive children and adults (Langlois, 1991). Additionally, raters agree who is and who is not attractive both within and across cultures (Winston, 2007). The psychological research provides convincing evidence that humans have a tendency to make rapid evaluations of character based on the attractiveness of individuals they encounter in their everyday lives (Rule, 2010; Olivola, 2010).

APPLICATIONS OF PSYCHOLOGICAL RESEARCH TO POLITICS

To extend this research, political scientists hypothesized that appearance evaluations and judgments of character based on attractiveness may be a relevant phenomenon in the political realm and may logically apply to voting decisions as well. For years, political science research had instead shown that voters quickly evaluate candidates and simplify the voting process using heuristics or shortcuts (Campbell, 1960). Studies found heuristics to be especially prevalent in the United States where constant elections increase “voter fatigue” and decrease the likelihood of information gathering on candidates (Praino, 2014). Though large amounts of information are available, most voters have gaps in their political knowledge that they rely on these simple shortcuts to overcome (Popkin, 1991).

At the turn of the 1960s, research found partisan attachment to be the most common heuristic. Studies of several electoral campaigns show that voting is not seen as a decision that must be made for each and every election. Party loyalties are not easily changed and provide voters with a group that shares certain ideological values that they can align with as a voting shortcut. The partisan heuristic acts as a long-term party allegiance and often frames attitudes towards candidates and psychological attachments to that political party (Popkin, 1991). Research also shows that voters will use incumbency as another quick heuristic to judge if a candidate would be a competent leader without taking time to gain updated information on them several years later (Hayes, 2010). These heuristics simplified voting and allowed individuals to easily understand which candidates they preferred.

But, as noted in the psychological research section, society changed immensely over the last 50 years and the political world changed as well. The most critical changes are the introduction of the American Independent Party and the introduction and massive growth of political media outlets and social media. In 1967, the American Independent Party was founded (Schaffner, 2002). This gave voters a third option and allowed more partisan mobility and freedom than was previously allowed by the dominance of a polarized two-party system. At the same time, throughout the 1960s, the television was growing in popularity and could be found in every American home (Johns, 2011). Political debates and major events could now be broadcast on television to the American people as they sat in the comfort of their living room. With the television as a vehicle to easily and intimately connect voters with visual images of candidates, it became easier to consider additional information beyond that candidate’s political party when voting (Hayes, 2014). The Independent Party, the decrease in partisan affiliation, the growth of media outlets and social media and the subsequent increased candidate exposure in this era drove political scientists to hypothesize that appearance could be a more relevant heuristic to quickly evaluate political candidates.

The political landscape continued to change at the turn of the century with the development of the Internet, social media and other communications channels. These technologies continued to amplify the constant exchange of images of political candidates to voters through various outlets. Further, these online capabilities enabled more sophisticated editing capabilities that could portray candidates in both flattering and unflattering ways. This drove campaign managers to redesign their publicity efforts to integrate and maximize positive opportunities and exposure in these media forms. Political advertisements on television and online became a way for candidates to reach voters and maintain consistent exposure. Candidates could now remain constantly present during an election season by simply spending money to expose their face to the public through advertising, commercials, participating in debates broadcast nationwide, speaking on talk shows, news channels interviews and engaging with voters on all forms of social media from Twitter to Facebook and more (Martin 1978). In a world dominated by visuals, the massive growth and constant proliferation of political imaging across these many media channels drove political scientists to assess the viability of the psychological research on the impact of appearance in judging individuals in a political context. Researchers dedicated a serious body of political research and energy to the idea that appearance-based judgments of political candidates may be a key heuristic that impacts candidate preference and subsequent voting behavior.

POLITICAL RESEARCH ON IMPACT OF APPEARANCE ON VOTE CHOICE

To test the impact of appearance on evaluations of and preference for political candidates, political scientists conducted studies that asked participants to rate images of political individuals based on appearance. As a whole, results found that appearance impacted which candidates voters preferred and that higher appearance ratings drove electoral preference and increased vote share (Lau, 2001). Just as psychologists noted in society at large, when making decisions in the political sphere, research found that voters create mental pictures of and opinions about candidate based on certain aspects of their physical appearance obtained from an image of the candidate (Rosenberg, 1986). These physical images of candidates, whether on television, in the newspaper, on a ballot or in some other media outlet are easy for individuals to obtain in an age where imaging, social media and other media channels are providing ever-present and non-stop content. An image of a candidate is what first attracts voter attention and garners a first impression that influences all further judgments made about a candidate (Olivola, 2010). This relationship between increased appearance ratings and vote share holds whether responses are recorded rapidly or over longer periods of times. The correlation also remains regardless of candidate age (Buckley, 2007). These findings echo psychological research that states that subconsciously rating faces and creating mental pictures is an evolutionary perceptual process and a skill that humans innately possess from a young age (Rosenberg, 1991).

Political scientists then found that attractiveness is the primary feature of appearance that triggers positive candidate reception by voters and drives which candidates voters prefer and thus select (Berggren, 2009). Similar to psychological research that found that

what is beautiful is mentally connected to what is good, political research found that attractive candidates were associated with having a more positive character (Rosenberg, 1991). Studies found that voters responded more favorably to candidates that are more nonverbally desirable. Further, individuals that were given a higher attractiveness score received a larger portion of the vote than those with a lower attractiveness score (Lawson, 2010; Lenz, 2011).

Studies also define what attractiveness looks like and what aspects of appearance are more visually desired. Some of the basic characteristics of appearance that are attractive are a high forehead and dark or grey hair for males. For females, dark or grey hair in a classic fashion is most visually appealing (De Landtsheer, 2008). In addition, research shows that voters are naturally drawn to individuals with a healthy appearance, symmetrical features and more of a baby face (Rhodes, 2002; Little, 2005; Zebrowitz, 2005). To further illustrate the features that define attractiveness, one study created automated face recognition technology to develop a smile index of photographs of candidates. This tool objectively measured facial traits and found a smile to be a positive and significant cause of increased vote share in Japanese and Australian communities. The significance of a smile in attracting one individual to another was noted through increased activity and activation in the medial orbitofrontal cortex of the brain when voters were presented with a smiling face (Horiuchi, 2011). Testing therefore shows that a smile is one of the most important features and that this feature automatically triggers a favorable response for respondents evaluating candidate appearance. In summary, research verifies that attractiveness is a determinant of character and a heuristic that voters use when selecting which candidates they prefer and for whom they will vote. Attractiveness, therefore, is found to lead to political success.

III. Shortcomings in Present Studies

METHODOLOGICAL SHORTCOMINGS

Though present research has concluded that candidate appearance and specifically attractiveness impacts how voters evaluate candidates and leads to increased vote share, there are several inconsistencies among these studies that are essential to note and understand. An evident shortcoming of the current literature is that present studies display an inconsistency in both structure and methodologies yet make subsequent widespread claims made despite these differences. First, the studies in the literature question survey respondents about attractiveness in different ways. Some studies ask “Is this the kind of person that you want to represent you in the US Congress?” or pose similar questions about a candidate’s suitability for office (Rosenberg, 1986; Lawson, 2010). With these more general questions, it is unknown what aspects of appearance respondents are judging. Other studies ask respondents to vote in a mock election, to pick which candidate is most attractive, to rate candidates on certain characteristics or to guess the outcome of a political election (Praino, 2014). No consensus can be made regarding what aspect of a candidate respondents are analyzing due to the lack of standardization in the overall research.

Second, all studies do not use real politicians and instead ask respondents to rate images of random or made-up individuals (Rosenberg, 1986). Other research instead utilizes pictures of candidates from actual elections such as the U.S. House of Representatives or smaller local elections (Praino, 2014). In addition, some studies deviate further from reality and manipulate candidate photographs (Rosenberg, 1991; Terkildsen, 1993; Weaver, 2009). Despite these varying methods, all results are lumped into generalized conclusions about the relationship between attractiveness and vote share.

Third, there are also only a few studies that use high-information elections and consider ratings of candidates that are more likely to have wider reception by various media outlets. Most studies instead choose to analyze obscure election contests that are not widely publicized (Banducci, 2008). Conclusions from studies that analyze appearance in low-information elections only account for candidates that are less popular and provide different results than an analysis of high-information elections where candidates have massive coverage in several media channels. Again, both high-information and low-information contests are generalized into over-arching results (Buckley, 2007).

Fourth, most studies assume that candidates are not recognized by respondents and do not institute a control to test the impact of candidate recognition. To prevent recognition of politicians and to ensure judgments based solely on appearance, some studies travel abroad to attempt to control biases (Antonakis, 2009; Lawson, 2010). Others ignore this variable entirely and simply hope that candidates will not be recognized (Praino, 2014). There is currently no research that controls for the impact of candidate recognition in the relationship between candidate attractiveness and vote share.

Fifth, the presentation of candidates to survey respondents is not standardized across the literature. Some studies provide only a photo of a candidate (Weaver, 2009). Others instead provide names, party information or a brief description about the candidate and their stance on several issues (Lawson, 2010). Studies also differed in the choice to present only one candidate on a page, multiple candidates on a page or competing candidates on a page (Lau, 2001). It cannot be assumed that the respondent is rating appearance when other variables are presented to respondents that could potentially impact candidate evaluations.

Sixth, presentation style varies in the literature. Some studies were conducted in-person while others were delivered to respondents online (Terkildsen, 1993). Further, manual or in-person studies are divided into a self-run study or a study facilitated by another individual (Praino, 2014). The impact of these different vehicles is currently unknown and there are no best practices on how this data should be collected to ensure quality and consistency among research.

With this significant variance in methodology across the literature, the many inconsistencies make it difficult to create generalized and widespread conclusions that aggregate these vastly different studies to answer the same overall question about the relationship between candidate attractiveness and appearance. A study in a low-information election that asks a group of college students to predict the winning candidate and includes information on a candidate's issues is vastly different from a study in a presidential election that asks for an attractiveness rating by an age-variant sample and only includes a picture of a can-

didate. These varying methodologies analyze different aspects of the appearance in politics are an evident shortcoming of current research that invites further study to standardize and replicate some of these structures to create conclusive results.

GENDER AND RACE VARIABLE SHORTCOMINGS

Methodology also includes the variables examined in a particular study. In present studies concerning the relationship between candidate attractiveness and vote share, an additional methodological shortcoming is the lack of inclusion of the race and gender variables. Race and gender are two demographic variables that are rapidly gaining both political vote share and mayoral, gubernatorial, House and Senate representation (Banducci, 2008). In terms of race, white Americans have presently expressed increased willingness to vote for black candidates as evidenced by Barack Obama's two terms as the first black President of the United States (Philpot, 2009). The same increased preference has been noted for female candidates (Berggren, 2009).

Despite achievements of individuals like Barack Obama and other female politicians like Hillary Clinton and Condoleezza Rice, a limited number of females and even fewer blacks and minority candidates hold high political offices in the United States. Research supports this lack of representation and shows that the election of black candidates is directly correlated with the proportion of blacks in that electoral district's population (Hahn, 1976). Furthermore, the differences between light-skinned blacks and darker-skinned blacks and the extent to which a person appears stereotypically black were also found to be statistically significant in cognitive perception and unconscious judgments of character and criminality (Terkildsen, 1993; Weaver, 2009). Psychologists and political scientists hypothesize that the stereotypical responses heuristic is triggered when an individual sees these demographic cues in an image of a candidate and that this heuristic impacts candidate evaluations and vote choice (Campell, 2014). With these current conditions, research to understand how exactly race and gender impact voters' evaluation of candidate appearance is essential. Stereotypes can clearly bias voters' evaluations of candidate appearance and therefore must be controlled for to produce non-biased results.

In spite of these theorized impacts, few studies strongly considered these variables in their quest to understand how appearance impacts candidate appearance evaluation and vote share. Demographic variables are continually overlooked despite clear visual evidence that different races and genders present noticeable physical differences. Most studies mention the need to examine these demographic variables due to potential prejudices that can drive a different evaluation of a minority candidate (Sparks, 2009). Fewer put this idea into practice and actually added these variables to understand if they are another heuristic that individuals rely on through automatic stereotypes upon seeing an image; most hold that these stereotypical reactions are not significant or can be controlled for (Trent, 2010). In addition to these inconsistent results, studies that include the race variable are often incomplete or falsely conclusive (Sigelman, 1995). For example, a study may evaluate an election in California, one specific geographic market, and subsequently make a wide-

spread claim about the behaviors of all populations. On the whole, though, results about the impact of the demographic variables of race and gender are ambiguous and need to be confirmed by further comparative studies before a clear and comprehensive assessment of the impact of these variables is made.

To start, hypotheses about the impact of the race variable are widely researched in psychology. Research shows that racial stereotypes are triggered automatically when people are asked to evaluate someone who is black even if they claim that they are not racist (Gaertner and McLaughlin, 1983). Studies show that individuals subsequently let their automatic processing dominate and either consciously or subconsciously rely on the activated racial judgments when evaluating minorities instead of utilizing cognitive effort to suppress the activated group in a controlled processing stage (Terkildsen, 1993).

Though research is limited, studies in the political realm that have tested if the race variable impacts the relationship between appearance and attractiveness and vote share show that race could potentially be a negative bias for minority candidates. Some researchers studied existing cases of minority politicians such as Jesse Jackson, Tom Bradley, Wilson Goode and Harold Washington and their pitch for national, state and local positions (Citrin, 1990). Studies show that white citizens' evaluations of a minority candidate's appearance may be partly determined by their racial attitudes and stereotypes; researchers found that white voters preferred the white politician to an identical black candidate (Devine, 1989) and that whites' attitudes of black candidates are often influenced by negative racial considerations (Terkildsen, 1993). These studies show that race may influence assessments made about the black candidate's personality and that white voters may evaluate black candidates more harshly (Sigelman, 1984; Moskowitz and Stroh, 1991; Sigelman, 1995; Banducci, 2008). Despite these confident conclusions, conflicting results arise from additional research that found that race does not impact the attractiveness ratings and candidate vote share. Other studies found that white voters are willing to vote for a hypothetical "qualified" black candidate for president (Colleau, 1990). The current overall weak analysis of the race variable demands additional attention as a possible factor in the attractiveness and candidate vote share hypothesis.

While studies of the impact of gender, the second key demographic variable, on politics are far more widespread than studies of the impact of race on politics, agreement on the impact of gender on attractiveness ratings and candidate vote share is also widely disputed. Some experiments that include candidate gender generally find a strong positive correlation between physical attractiveness and electability for male candidates and a strong negative correlation for female candidates. Most hold that attractiveness is measured differently for men and women (Johns, 2007). Studies suggest that more attractive women tend to be penalized at the ballot box while attractive men tend to be rewarded (Rosenberg, 1991). Yet other political scientists disagree and find that attractiveness impacts females' perceived femininity and thus enhances their perception as pleasant people. As a result of this evaluation, females may be rewarded for their beauty and may fare better by posting higher candidate ratings at the polls (Berggren, 2009). Still more studies find that gender can be controlled for and that other aspects of appearance and attractiveness beyond candidate

gender determine voters' candidate preference (Johns, 2011). Some studies ignore gender all together and survey only white males.

The gender variable and its relationship to attractiveness and vote share complicate further when voters are asked to evaluate intra-gender political contests. Studies found that physical attractiveness plays a large role when both candidates running for office are of the same gender. Voters seem to be easily influenced by good-looking candidates when it is easy for them to choose which man or which woman looks best in a set of two candidates of the same gender. Yet, like the race variable, the gender variable is not a staple in every study and the true impact of gender on this question is far from resolved beyond these mixed findings. Just as with methodological choices, these shortcomings related to the stereotypical identifiers of race and gender point out inconsistencies across studies that call for further research and analysis.

SHORTCOMINGS SUMMARY

After over a hundred years of psychology research on the impact of beauty in perception judgments and nearly thirty years of political research about the relationship between candidate attractiveness and vote share, the literature as a whole states that appearance impacts how voters evaluate candidates and which candidates they select. Further, research shows that voters use attractiveness as a determinant of character and a way to measure appearance. Though there is some disagreement from recent research that has found competence judgments to be more dominant than attractiveness judgments in evaluating appearance and determining vote share (Todorov, 2005; Johns, 2007; Ballew, 2007; Atkinson; 2009; Antonakis, 2009; Olivola, 2010; Riggio, 2010; Armstrong, 2010; Verhulst, 2010; Lautsen, 2013; Goncalves, 2014; Praino, 2014; Na, 2015), the relationship between candidate attractiveness and vote share is largely accepted by the political research community.

Yet, despite these conclusions, inconsistent methodological differences among these past studies motivates this new research and establishes the importance and relevance of the current project. Consistency must be established in these studies moving forward relating to methodological structure of studies and the inclusion of stereotypical identifiers such as race and gender for widespread claims to be made.

IV. Experiment Methodology

As the literature shows, many studies have contributed to research on the relationship between candidate attractiveness and vote share. Thus, in order to verify the positive correlation between attractiveness and vote choice, this research will test this widely accepted hypothesis to reaffirm that increased attractiveness ratings do in fact produce increased vote share. This experiment will utilize practices from a comprehensive 2014 study by Rodrigo Praino, Daniel Stockemer and James Ratis that asked college-aged students to rate images of candidates in the 2008 U.S. House of Representatives election on a scale of 1 to 10, a model selected for its comparability to my current situation as an undergraduate

student at Bentley University. Modeling this comparable case, the present research will use images of 2014 U.S. House of Representatives candidates test the hypothesis on college-aged students and create a substantive analysis to study the relationship between candidate attractiveness and vote share.

In addition, the present study will control for several of the methodological inconsistencies that characterize the previous literature to determine best practices in terms of experimental structure. As Praino did in his 2014 study, the present research will only provide respondents with candidate photos to ensure that the visual image is the only determinant of the respondent's given attractiveness rating. Respondents will also be able to indicate if they recognize a candidate to enable a control for this critical yet widely ignored factor that is not present in any literature. Real politicians will be used in the present study to reassert the need to simulate evaluations of actual political candidates to follow Praino's best practices. Finally, to again echo Praino's dual methodology research structure, this study will use two different methodologies, an online and a manual component, to test if there is a discrepancy between methods or if one vehicle is more effective.

Together, these elements will create a methodological analysis of race, gender, candidate recognition and survey methodology in a holistic manner similar to a 2008 study by Susan A. Banducci, Jeffrey A. Karp, Michael Thrasher and Colin Rallings that analyzed a comprehensive list of potential factors that could impact appearance judgments in low information elections. The comprehensive results will be used as call to action to stress the need to streamline all future studies to produce consistent and reliable overall conclusions.

EXPERIMENT DESIGN

To assess the impact of candidate attractiveness on vote share, candidates from the 2014 U.S. House of Representatives, a body diverse in both gender and race, were used as subjects for the present study. The 875 candidates and challengers were reduced to include only incumbents running for reelection. Candidates were randomly selected using Excel's random number generator and that number was copied into a new column to prevent refreshing upon worksheet changes and then assembled in ascending order. From this group, 200 candidates were selected to create a sample large enough to both measure respondents' opinions about candidate attractiveness and feasible enough for a semester long study.

Next, the sample was checked for racial and gender diversity. Though respondents should constantly see white males during the survey to simulate actual societal conditions of a white male dominated Congress, minorities and females were required to assess candidate evaluations of minority individuals. The sample was slightly amended as Andy Baker, Chappell Lawson, Gabriel S. Lenz and Michael Meyers did to ensure minorities in their 2010 study of candidate appearance and electoral success. The final group included 108 males (54%), 51 minorities (25.5%) and 61 females (30.5%).

Images of the chosen candidates were presented as a PowerPoint presentation. Official pictures of candidates were taken directly from Congress.gov. If a candidate did not have this image available, then a picture with as neutral of a background as possible was used.

To again follow Lawson, Banducci and Praino, it was essential to provide a standardized set of pictures across all candidates to minimize any biases for different photo quality or backgrounds. All candidates had a color photograph available for use to ensure standardization across candidates. Each candidate was given a number that was placed next to his or her picture. Four pictures were randomly placed on each slide as Lawson's study previously did and produced a total of fifty pages of candidates (see Appendix A for an example). To eliminate the possibility of comparison and the threat that the first candidate shown will serve as a point of reference for subsequent attractiveness evaluations, this study follows Praino's best practice and provides each respondent with a different first page of their survey.

An instruction packet was created to supplement the survey and did not mention anything about politics (see Appendix B). This packet was presented to each participant at the inception of the survey without verbal instructions and asked each candidate to rate 40 individuals, one on each line of the respondent sheet (similar to the example response table in the instruction sheet). The instruction sheet also provided a column for respondents to indicate if they recognized a particular candidate. This variable was included to understand if recognizing a candidate (knowing their name, knowing their face or knowing more than this about them) impacted the given attractiveness rating and to eliminate the threat of biased candidate evaluations for respondents who may know additional information about a certain candidate beyond the picture presented.

In addition to the instruction sheet, a demographic sheet was also designed to gain additional information about each respondent (see Appendix C). All respondents were required to turn in this survey with their candidate attractiveness scores. Demographic respondent questionnaires are commonplace throughout the literature and are essential to note information like the age, gender, race, partisan affiliation and if the respondent is following the current presidential elections or not. The current responder sheet mimics that of Banducci's 2008 study and provides respondent information that will be added to the final aggregate data as an additional control. Survey respondents were students at Bentley University, a four-year higher education institution located in Waltham, Massachusetts. Bentley's diversity is similar to the U.S. House of Representatives. While the House is approximately 70% white, 30% female and 30% minority, Bentley University is 40% female, 60% white and 40% minority.¹ The diversity on campus provided adequate respondent diversity across both race and gender. At the conclusion of data collection, this data was compared to participant data to show a similar representative sample of respondents for the present study.

TESTING

For the testing stage of this study, testing was completed in three stages: a pre-test, computer testing and manual testing. The aggregate computer and manual testing produced 161 reliable and randomly selected respondents who reacted viscerally to images of candidates presented in both an online format (47 responses) and a printed format (114

¹ Forbes. America's Best Value Colleges. Retrieved from <http://www.forbes.com/colleges/bentley-university/>

responses). As an incentive to participate, each student who took the survey provided their email and was entered to win one of four \$25 Amazon gift cards. The sections below outline the testing process and explain how the 161 respondents were targeted and surveyed and how this data was codified for analysis.

PRE-TEST

A Pre-Test was conducted to ensure that all documents were effective and that the experiment ran smoothly. The Pre-Test was given to a group of Sophomore Honors Students at Bentley University. The group completed the Pre-Test using the manual test method that provided a printout of candidate images randomly assembled in groups of 40 candidates. During the Pre-Test, several design choices were made that would remain standardized throughout the distribution of the survey. First, the response sheet would be paired with the PowerPoint at the start of the session and would be passed out to respondents as such. Second, respondents would not be told that the images are of political candidates. Third, no verbal instructions would be given to respondents and questions would only be answered if they arose. Finally, it became necessary to check and remind respondents to fill out the demographic sheet at the conclusion of the survey. The average completion time of the entire survey (reading instructions, rating 40 candidates, filling out the demographic sheet) was about five minutes. The maximum time was 10 minutes. This smooth Pre-Test ensured that the experiment was ready to be tested with a larger sample.

COMPUTER TESTING

To attract respondents, a lab session was advertised across Bentley's campus to attract respondents to take the survey in The Center for Marketing and Technology lab on Wednesday, March 30th. Pizza was provided at the generous donation of the Bentley Honors Program to feed and reward participants. Upon entering the lab, each individual checked in at the front and was given the instructions sheet and a computer to sit at. Respondents were then asked to scroll to a randomly assigned slide of the 200 slide PowerPoint of candidate images and to press play. This ensured that each respondent was not rating the same group of individuals and that all 200 candidates received an adequate and statistically relevant number of responses. Individuals were instructed to work quickly and to note the candidate's number, how attractive they found them to be on a scale of 1 to 10 and then to circle YES if they recognized the candidate.

MANUAL TESTING

To collect data beyond the lab session, the printed PowerPoint packets of randomly assembled groups of candidate images mentioned in the Pre-Test section were distributed to additional subjects. This group utilized the same response sheet as computer respondents.

RESPONSE CODIFICATION

Respondent ratings of each candidate and a notation of whether they were recognized or not were inputted into Excel with a numeric system to allow for further analysis. Each candidate was given a number (C1-C200) and each respondent was given a number that was noted at the top of his or her response sheet when they completed the survey. Age was recorded as the number of the respondent's age. For gender, male was indicated with a 1 and female with a 2. For race, Caucasian or white was indicated with a 1, Hispanic with a 2, black with a 3, native American with a 4, Asian/Pacific Islander with a 5 and Other with a 6. For party, Democratic was indicated with a 1, Republican with a 2, Independent with a 3, Libertarian with a 4 and Neither with 5. For how closely the individual is following the 2016 election, Very Closely was indicated with a 1, Somewhat Closely with a 2, A Little with a 3 and Not At All with a 4. For method of testing, Computer was indicated with a 1 and Manual was shown with a 2.

To code whether a respondent used the computer or the packet survey model, each respondent sheet was marked to indicate which methodology that particular respondent used. These data were subsequently inputted into Excel for each respondent. 'IF functions' were used to compute mean candidate attractiveness for aggregate packet and computer responses.

Once all respondent results were coded, an average numerical attractiveness score across respondents for each candidate was generated. This data was then paired with information about the 2014 U.S. House of Representatives candidates that included their name, party, birth year, vote share, state and other relevant data points. Candidates were sorted in ascending order by mean attractiveness and an additional column was added for the candidates' attractiveness rating (1 to 200). The codification stage prepared the survey results for analysis.

V. Results

The explanations of this study's results are divided into two sections. Each analysis aims to understand particular key factors that may or may not determine the relationship between candidate attractiveness and vote share. First, the substantive analysis seeks to understand the impact and statistical significance of the relationship between attractiveness and vote share and the variables of race, gender and candidate recognition. To do this, most and least attractive candidates, the relationship between attractiveness and vote share, race and gender variables and candidate recognition will be analyzed in depth. Second, a methods analysis will analyze the difference between the computer method and the packet method that participants used to complete the survey to understand if this variable impacts candidate attractiveness ratings.

SUBSTANTIVE ANALYSIS

A. Most and Least Attractive Candidates

The list of the Top 10 Most Attractive Candidates in the sample, beginning with the most attractive candidate, is included in Appendix D. This list presents several insights about the characteristics of attractive candidates. Of the Top 10, 80% candidates were White, 10% were Black and 10% were Asian. 70% of candidates in the Top 10 were female. The mean birth year of the Top 10 was 1967, or 49 years old. 70% of candidates in the Top 10 were Republican and 10% were Democrat. These general insights will be further evaluated in several relevant upcoming discussions.

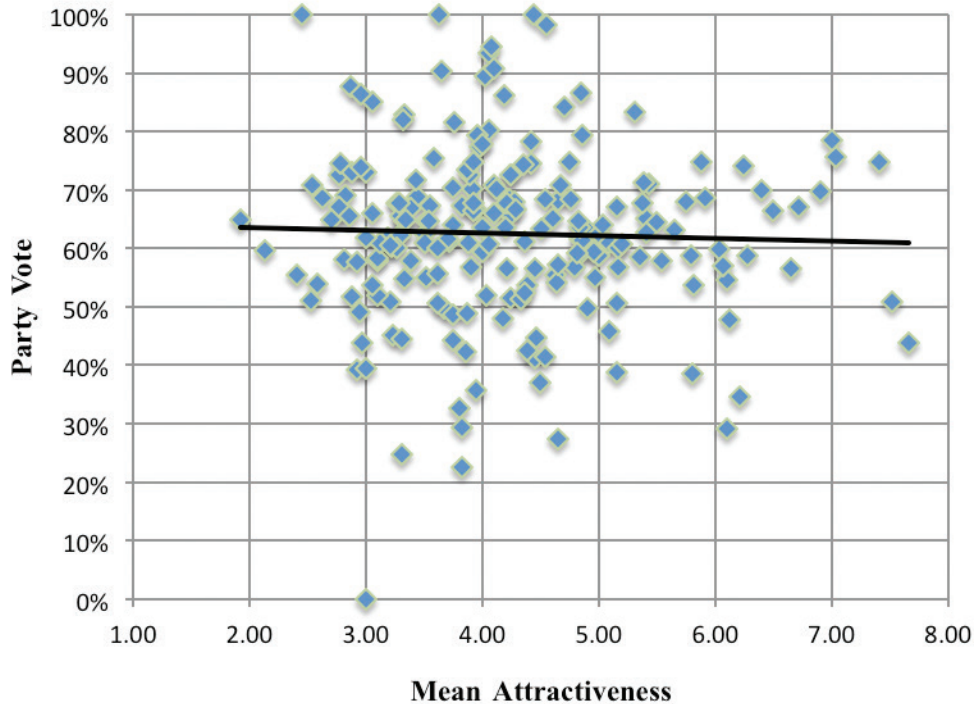
The Bottom 10 Least Attractive Candidates were also determined from the comprehensive data and are listed in Appendix E. The list begins with the least attractive candidate and again presents interesting insights. Of the Bottom 10, 70% of candidates were White, 10% were Black and 20% were Hispanic. 80% of candidates in the Bottom 10 were male. The mean birth year of the Bottom 10 was 1947, or 69 years old. 60% of candidates in the Bottom 10 were Republican and 40% were Democrat. As with the Top 10, these insights will be translated to several subsequent sections to understand the impact of these characteristics on candidate attractiveness and vote share.

B. Attractiveness and Vote Share

In addition to these basic insights about the most and least attractive candidates, all but one of the Top 10 received over 51% of the vote in their House contest. On the other hand, all of the candidates in the Bottom 10 achieved this benchmark. This questions if previous studies are correct in stating that higher attractiveness leads to increased vote share for political candidates.

To analyze this contradiction, mean candidate attractiveness and electoral share are shown in Figure 1. The figure shows each candidate's mean attractiveness on the horizontal axis and electoral share on the vertical axis. From this figure, a trend is unclear. It appears that most candidates receive about 60% of the vote regardless of attractiveness.

FIGURE 1
Party Vote vs. Mean Attractiveness



This inconclusive graph was supplemented with an SPSS analysis to test the statistical significance of the relationship between mean candidate attractiveness and electoral share using correlation. Candidates who received 100% vote share were removed from the test. This resulted in 168 candidates with a vote share that ranged from 35.71% to 98.37%. The t-statistic in the relationship between mean candidate attractiveness and electoral share is 0.614. This value falls within a normal curve and is not statistically significant. Therefore, contradictory to the wider literature, there is no relationship between mean attractiveness and vote share in this study.

C. Race/Gender Variable Analysis

In addition to the relationship between candidate attractiveness and vote share, it is essential to analyze if candidate attractiveness is correlated to other variables that may impact judgments of appearance of political candidates and subsequent vote share. First, race seems to have no effect on whether or not a person is considered attractive. Both the Top 10 and the Bottom 10 candidates were respectively only 30% and 20% minority representation. This is accurate based on the overall survey demographic that only included 25% minorities. Further, though the literature suggests that darker-skinned blacks are rated

more harshly, the 2nd most attractive candidate is black and darker-skinned. The only black individual in the bottom 10 was a light-skinned black, again contrary to the literature. Additionally, the mean attractiveness for whites, blacks and Hispanics, is 4.28, 4.15 and 3.87 respectively. For Hispanics, this number cannot be taken as a significant finding as there were only 7 Hispanic candidates, a sample too small to be reliable. Race was also found to be statistically insignificant at the 0.05 level in SPSS with a t-statistic of 0.512 for whites, 0.670 for blacks and 0.267 for Hispanics. Race is therefore not a factor in determining candidate attractiveness and vote share.

Gender also has no significant effect on candidate attractiveness. Though the Top 10 included 70% females while the Bottom 10 included 20% females, the mean attractiveness for males and females was 4.23 and 4.24 respectively. These values are nearly identical and show that both groups received similar attractiveness ratings, regardless of gender. Gender, mean attractiveness and vote share were also statistically insignificant at a 95% confidence interval at 0.566 for males and 0.093 for females as these numbers are not at the 0.05 significance level. While there is clearly no correlation between gender, mean attractiveness and electoral share for males, it appears that there is a 91% chance that gender is significant for females when determining candidate attractiveness and vote share and that attractiveness may be more important for females in gaining electoral share. This potential yet weaker correlation can be explained due to the smaller sample size (61 women of 200 total candidates). Despite a potential relationship, controlling for the candidate's partisan affiliation, challenger quality, partisan advantage and gender, the t-statistic is 0.605 and shows that gender for females is not significant. Additional data on females, mean attractiveness and electoral share is necessary to fully understand the impact of the gender variables for females in this study.

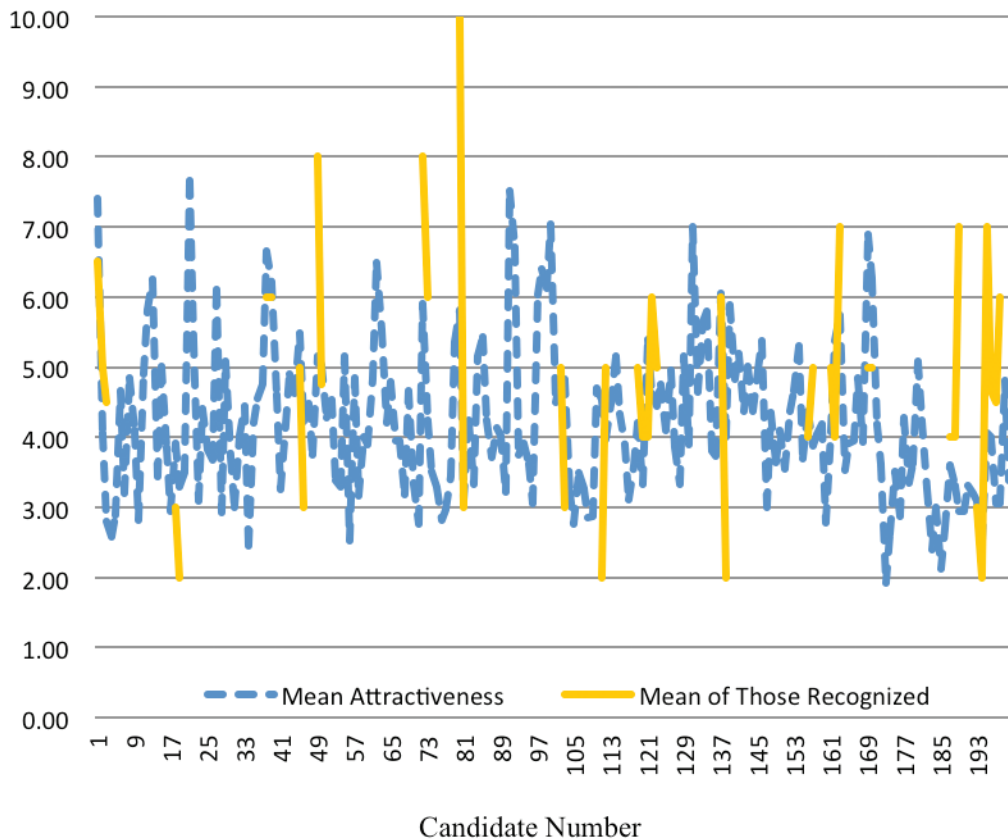
D. Candidate Recognition Variable

Though candidate attractiveness does not significantly correlate with vote share, race, or gender, this study also included a variable for candidate recognition. When rating candidate attractiveness, respondents could indicate if they recognized a particular candidate. 66 candidates were recognized at least once in the study. Some notable candidates who were recognized by respondents were: John Boehner, Speaker of the House, 2011 to 2015 (Republican), recognized 7 times or 4.3%; Nancy Pelosi, Speaker of the House, 2007 to 2011 (Democrat), recognized 6 times or 3.7%; Ann Wagner, St. Louis Congressman (Republican), recognized 4 times or 2.4%; and Charlie Rangel, New York Congressman (Democrat), recognized 3 times or 1.9%.

The mean attractiveness of candidates who were recognized by respondents was 4.57. The mean attractiveness of candidates who were not recognized by respondents was 4.34. A Paired Samples Test compared the mean of those recognized to the mean of those who are not recognized to give a t-statistic of 0.578. This is statistically insignificant and shows that recognition did not impact the mean attractiveness significantly between candidates recognized and those who are not. To graphically depict this, a line graph was constructed using Excel with the attractiveness rating for each candidate (dashed line) and the mean at-

tractiveness for each and any candidates who were recognized (solid line). Figure 2 echoes the findings that the difference in mean attractiveness between those recognized and those not is statistically insignificant. Though there is no significance between candidate recognition and candidate attractiveness ratings, when a candidate is recognized, the graph shows that individual respondent attractiveness ratings are polarized. When a respondent knew who a candidate was, they either rated that person significantly higher or lower than the average respondent who did not recognize the candidate. This polarization that could be due to partisan affiliation, geographic affiliation, or another variable is widely ignored in the literature but definitely a variable to consider when studying candidate attractiveness and vote share.

FIGURE 2
Mean Attractiveness vs. Those Who Are Recognized

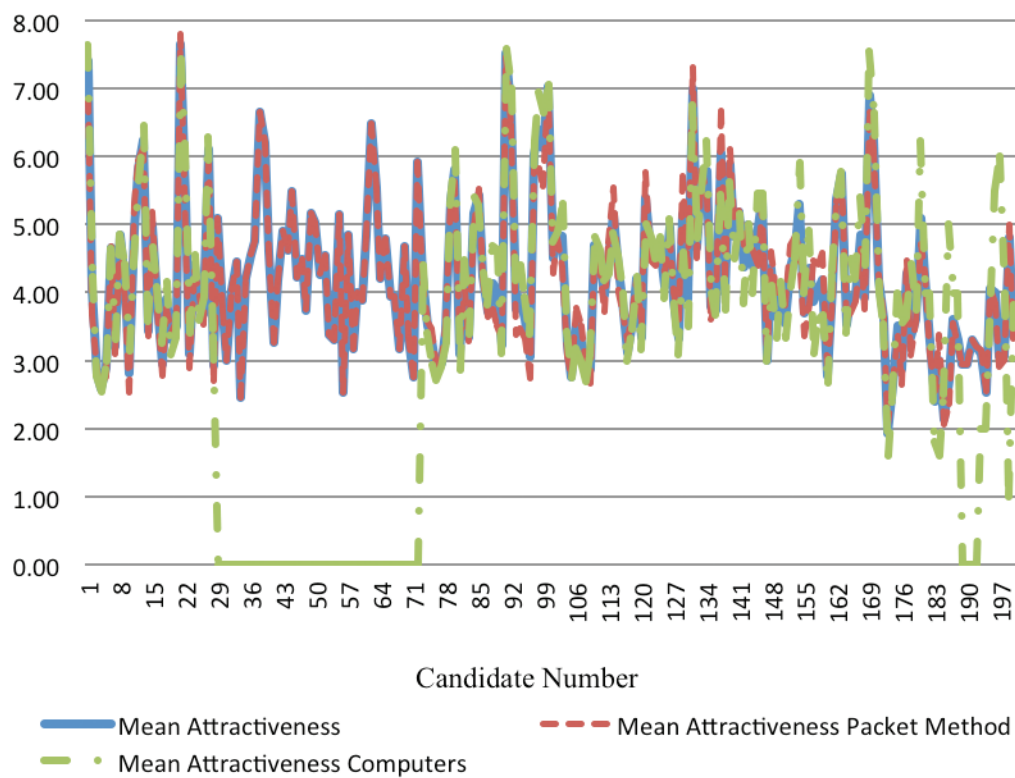


METHODS ANALYSIS

In addition to the substantive analysis, the methods analysis will show any potential difference in results between the two tools used to collect mean attractiveness ratings. Figure 3 shows a graph of the overall mean attractiveness (solid line), mean attractiveness for

packet responses (dashed line) and mean attractiveness for computer responses (dash-dot line) for each candidate. The lack of dash-dot ratings from about 29 to 72 indicates that these candidates' attractiveness scores were only captured with one method and that there was thus no data to compare. This graph indicates a difference between the two methods but is rather unclear about any sort of trend. While there appears to be slight variance, much of the mean attractiveness data appears to overlap across methods.

FIGURE 3
Mean Attractiveness – Different Methods



To supplement this inconclusive Excel graph, SPSS was used to further analyze and compare the mean candidate attractiveness for each method. The mean attractiveness of packet responses was 4.19 and the mean attractiveness of computer responses was 4.28. A Paired Samples Test was executed and produced a t -statistic of 0.092, which shows that using either method over the other is not significant at the 0.05 level. Therefore, the method used did not drastically alter candidate attractiveness ratings and the inconsistency across the literature between online and in-person methodologies clearly does not impact the results that are recorded.

VI. Conclusion

The present study partially modeled a 2014 study by Rodrigo Praino et al. to understand the potential relationship between candidate attractiveness and vote share for 2014 U.S. House of Representative candidates. 161 respondents were asked to rate these individuals based on how attractive they thought they were on a scale of 1 to 10. The study was constructed with two different survey methodologies to understand the potential impact of different survey tools while conducting a parallel analysis of the impact of candidate attractiveness on vote share. Further, the study deliberately included over 25% minority candidates, 30% female candidates and the ability to mark candidates that a respondent recognized to provide observations on these often ignored variables.

This study contributes to the wider literature by raising questions and urging additional research about a largely agreed-upon relationship in political science research. The results of this study disagreed with the current literature and found that the relationship between mean candidate attractiveness and vote share is not significant. The discrepancy between the conclusions in this study and the conclusions in the literature suggests that literature study respondents who claim to rate candidates based on attractiveness may instead be using other heuristics that have not been tested. One possibility is that those suggesting competence as a heuristic are in fact correct and that this is a dimension that respondents are reacting to when asked to rate images of political candidates. Future research that moves beyond the limits of this study's correlational analysis between only the attractiveness and vote share variables to instead apply a multivariate model will be more apt at considering a multitude of other potential factors. As a whole, though, this dissonant finding that contradicts the widely accepted relationship between candidate attractiveness and vote share suggests that further research is essential to continue to understand how attractiveness or other less studied heuristics impact which candidates respondents prefer and subsequently vote for.

In addition to highlighting the need to continue to analyze the relationship between candidate attractiveness and vote share, this study also improves to the wider literature by including several additional variables are historically ignored or undertreated. While the results of this study show insignificance for these race, gender, candidate recognition and survey methodology variables, this study simply scratches the surface in terms of comprehensive understanding of these key factors in appearance judgments. More research is clearly necessary, especially regarding both the impact of female attractiveness on vote share, a correlation that was significant at the 91% level and the difference between darker-skinned and lighter skinned-blacks. This study implores researchers who do not examine a wide variety of variables or claim insignificance without proof while studying this question to further investigate and manipulate potential determinants of attractiveness for a more holistic understanding.

Beyond the contributions and benefit of this study as a call to action for further research in this field and for the inclusion of several key variables that impact attractiveness ratings and electoral share, the findings of this study also provide interesting comments

about politics and elections. First, if attractiveness indeed does not matter in terms of vote share, this could drive a shift in our visually driven culture towards a focus on educating the public on political issues and nurturing this potential desire to vote based on information and knowledge instead of a heuristic like appearance or attractiveness. This shows that voter fatigue in the United States may not be as prominent as believed. As these results indicate, people may just want the information presented to them more directly and easier so that they can make informed decisions and ones that are not based on superficial considerations like attractiveness.

Second, findings that race does not matter in terms of attractiveness could show that progress towards eliminating racism and truly seeking the best candidate regardless of skin color is happening. This finding could predict for minority victories in non-minority dominated districts and indicate a focus on finding the best leader despite demographic differences.

Third, though recognition was not found to be statistically significant, the polarized results from the few respondents who recognized particular candidates show that recognition may be important in politics. While it may not matter how attractive a candidate is, this research shows that people clearly form strong opinions when they recognize someone. Politicians should therefore use any and all tools at their disposal to get themselves in front of potential voters who want that person-to-person connection from someone who may be representing them in the political sphere.

This study leaves significant opportunities for future studies to expand and develop the ideas presented here and in the literature. One area of study beyond the scope of the present study was the idea of similarity between the respondent and the candidate and the potential impact that this may have on vote choice. In the widespread literature on voting heuristics, this similarity principle is largely ignored. Several researchers have devoted entire studies to this idea and found that individuals rate those who are similar to them more favorably (Cutler, 2002; Bailenson, 2008; Zebrowitz, 2008). A future study should analyze this by collecting demographic information about respondents in a similar manner to the present study, comparing this information to the candidates and noting if this impacts attractiveness ratings and vote share. Additional information could be collected if sophisticated enough photography tools were available to compare facial images of respondents to facial images of candidates. Facial manipulation tools could also be used to warp images of candidates to look more like a respondent to measure if this change makes the respondent more likely to vote for them. Studies on the similarity principle could produce interesting results that may overshadow attractiveness ratings and instead present the idea that people are instead more apt to vote for those who are similar to them.

Another future study that could supplement this research would be an analysis on the impact of respondent age. In the present study, the average age of respondents was 20 years old and 58% were male. This may have caused the large number of younger attractive females in the Top 10 and the large number of older males in the Bottom 10. It would be interesting to replicate this study with a different age population to understand if this could change. If a population of older males and females were studied instead, with an average

age closer to the average age of the candidates, different results may occur. Previous studies like Olivola, Banducci, Praino and Lau show that college students can be an adequate representation of the population, but studies of various age demographics must confirm this finding to prove that college-aged students in 2016, with a good education and constant access to news, will not skew survey results. Future studies will allow the current research to influence political scientists and to produce a more complete understanding of the relationship between candidate attractiveness and vote share.

References

- Antonakis, J., & Dalgas, O.** (2009). "Predicting elections: Child's play!" *Science*, 323, 1183.
- Armstrong, J. S., Green, K. C., Jones Jr., R. J., Wright, M. J.** (2010). "Predicting Elections from Politicians' Faces." *International Journal of Public Opinion Research*, 22:4, 511-522.
- Banducci, S. A., Karp, J.A., Thrasher, M., & Rallings, C.** (2008). "Ballot Photographs as cues in low-information elections." *Political Psychology*, 29, 903-917.
- Ballem, C. C., & Todorov, A.** (2007). "Predicting political elections from rapid and unreflective face judgments." *Proceedings of the National Academy of Sciences of the USA*, 104, 1 7948-17953.
- Berggren, N., Jordahl, H., Poutvaara, P.** (2009). "The looks of a winner: Beauty and electoral success." *Journal of Public Economics*, 94, 8-15.
- Buckley, F., Collins, N., Reidy, T.** (2007). "Ballot Paper Photographs and Low-Information Elections in Ireland." *Politics*, 27:3, 174-181.
- Citrin, J., Green, D. P., Sears, D. O.** (1990). "White Reactions to Black Candidates: When Does Race Matter?" *The Public Opinion Quarterly*, 54:1, 74-96.
- Colleau, S. M., Glynn, K., Lybrand, S., Merelman, R. M., Mohan, P., Wall, J. E.** (1990). "Symbolic Racism in Candidate Evaluation: An Experiment." *Political Behavior*, 12:4, 385-402.
- De Landtsheer, C., De Vries, P., Vertessen, D.** (2008). "Political Impression Management: How Metaphors, Sound Bites, Appearance Effectiveness and Personality Traits Can Win Elections." *Journal of Political Marketing* 7:3-4, 21-238.
- Dion, K., Berscheid, E. Walster, E.** (1972). "What is Beautiful is Good." *Journal of Personality and Social Psychology*, 24:3, 285-290.
- Gaertner, S. L., McLaughlin, J. P.** (1983). "Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics." *Social Psychology Quarterly*, 46:1, 23-30.

- Horiuchi, Y., Komatsu, T., Nakaya, F.** (2011). "Should Candidates Smile to Win Elections? An Application of Automated Face Recognition Technology." *Australian National University Crawford School of Economics and Government*, 2.
- Johns, R., Shephard, M.** (2011). "Facing the Voters: The Potential Impact of Ballot Paper Photographs in British Elections." *Political Studies*, 59, 636-658.
- Johns, R., Shepard, M.** (2007). "Gender, Candidate Image and Electoral Preference." *Political Studies Association*, 9, 434-460.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., Smoot, M.** (2000). "Maxims or Myths of Beauty? A Meta-Analytic and Theoretical Review." *Psychological Bulletin*, 126:3, 390-423.
- Langlois, J. H., Ritter, J. M., Roggman, L. A., Vaughn, L. S.** (1991). "Facial Diversity and Infant Preferences for Attractive Faces." *Developmental Psychology*, 27:1, 79-84.
- Lau, R. R., Redlawsk, D. P.** (2001). "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making." *American Journal of Political Science*, 45:4, 951-971.
- Lautsen, L.** (2014). "Decomposing the Relationship Between Candidates' Facial Appearance and Electoral Success." *Journal of Political Behavior*, 236, 777-791.
- Lawson, C., Lenz, G. S., Baker, A., Myers, M.** (2010). "Looking Like a Winner: Candidate Appearance and Electoral Success in New Democracies." *World Politics*, 62, 561-593.
- Lenz, G. S., Lawson, C.** (2011). "Looking the Part: Television Leads Less Informed Citizens to Vote Based on Candidates' Appearance." *American Journal of Politics Research*, 55:3, 574-589.
- Little, A. C., R. P. Buriss, B. C. Jones and S. C. Roberts.** (2007). "Facial Appearance Affects Voting Decisions." *Evolution and Human Behavior* 28(1): 18-27.
- Moskowitz, D., Stroh, P.** (1994). "Psychological sources of electoral racism." *Political Psychology*, 15, 307-329.
- Na, J. Kim, S., Oh, H., Choi, I., O'Toole, A.** (2015). "Competence Judgments Based on Facial Appearance Are Better Predictors of American Elections Than of Korean Elections." *Journal of Psychological Science*, 26:7, 1107-1113
- Nixon, B., Pollom, S.** (2006). "Effects of Facial Maturity on Voting Preference." *Hanover College, PSY 401*.
- Olivola, C. Y., Todorov, A.** (2010). "Elected in 100 Milliseconds: Appearance-based Trait Inferences and Voting." *Journal of Nonverbal Behavior*, 34, 83-110.
- Popkin, S. L.** (1991). *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. University of Chicago Press.

- Praino, R., Stockemer, D., Ratis, J.** (2014). "Looking Good or Looking Competent? Physical Appearance and Electoral Success in the 2008 Congressional Elections." *American Politics Research*, 42:6, 1096-1117.
- Riggio, H. R., Riggio, R. E.** (2010). "Appearance-Based Trait Inferences and Voting: Evolutionary Roots and Implications for Leadership." *Journal of Nonverbal Behavior*, 34, 119-125.
- Rosenberg, S. W., Bohan, L., McCafferty, P., Harris, K.** (1986). "The Image and the Vote: The Effect of Candidate Presentation on Voter Preference." *American Journal of Political Science*, 30:1, 108-127.
- Rosenberg, S. W., & McCafferty, P.** (1987). "The image and the vote: Manipulating Voters' Preferences." *Public Opinion Quarterly*, 51, 1623-1626.
- Rosenberg, S. W., Kahn, S., & Tran, T.** (1991). "Creating a political image: Shaping appearance and manipulating the vote." *Political Behavior*, 13, 345-367.
- Rule, N. O., Ambady, N., Adams Jr., R. B.** (2010). "Polling the Face: Prediction and Consensus Across Cultures." *Journal of Personality and Social Psychology*, 98:1, 1-15.
- Schaffner B.F. and M.J. Streb** (2002). "The Partisan Heuristic in Low-Information Elections" *Public Opinion Quarterly* 66 (4), 559-581.
- Sigelman, C. K., Sigelman, L., Walkosz, B. J., Nitz, M.** (1995). "Black Candidates, White Voters: Understanding Racial Bias in Political Perceptions." *American Journal of Political Science*, 39:1, 243-265.
- Sparks, D. B., Watts, C. S.** (2011) "Race Still Matters: Racial Attitudes and Their Effects on Candidates Evaluations." *Duke University*.
- Spezio, M. L., Rangel, A., Alvarez, R. M., O'Doherty, J. P., Mattes, K., Todorov, A. et al.** (2008). "A neural basis for the effect of candidate appearance on election outcomes." *Social Cognitive and Affective Neuroscience*, 3, 344-352.
- Terkildsen, N.** (1993). "When White Voters Evaluate Black Candidates: The Processing Implications of Candidate Skin Color, Prejudice and Self-Monitoring." *American Journal of Political Science*, 37:4, 1032-1052.
- Todorov, A., Madisodza, A. N., Goren, A., & Hall, C.** (2005). "Inferences of competence from faces predict election outcomes." *Science*, 308, 1623-1626.
- Verhulst, B., Lodge, M., Lavine, H.** (2010). "The Attractiveness Halo: Why Some Candidates are Perceived More Favorably Than Others." *Journal of Nonverbal Behavior*, 34, 111-117.
- Weaver, V. M.** (2009). "Shades of Political Behavior: Examining Hidden Bias in Candidate Evaluation." *Woodrow Wilson Department of Politics, University of Virginia*.
- Zebrowitz, L. A., & Mmontepare, J. M.** (2005). "Appearance DOES Matter." *Science*, 308, 1565-1566.

Appendix A
Sample Experiment Images

#9



#10



#11



#12



Appendix B



#27

Instruction Sheet

You will rate a series of pages that display four individuals per page with accompanying case numbers.

For each individual:

1. Write down the accompanying number.
2. Rate their attractiveness on a scale of 1 to 10 (1= very unattractive, 10=very attractive)
3. In the third column, indicate whether you recognize the individual by circling Yes. Leave it blank if you do not recognize the individual.
4. Move on to the next photograph.

Try to work quickly and go with your gut instinct!

Example

After looking at the individual on your response sheet, write in the number, what you would rate their attractiveness and circle yes if you recognize them.

Candidate Number	Attractiveness Rating (scale of 1 to 10)	Do you recognize this candidate?
27	5	YES

Appendix C*Respondent Questionnaire*

Age: _____

Are you:

- a. Male
b. Female

Please check the box that best describes you:

- White/Caucasian
 Latino/Hispanic Origin
 Black/African American
 Native American/American Indian
 Asian/Pacific Islander
 Other _____

Do you consider yourself a Republican, Democrat, Independent, or something else?

If you answered Independent or something else, do you lean towards the Republican Party, Democratic Party, or neither?

How closely are you following this presidential election?

- Very Closely Somewhat Closely A Little Not At All

Appendix D

Most Attractive Candidates

Most Attractive Candidates

 1. Elise Stefanik	 2. Mia Love	 3. Aaron Schock	 4. Tulsi Gabbard
 5. Robin Kelly	 6. Eric Swalwell	 7. John Boehner	 8. Barbara Comstock
	 9. Kristi Noem	 10. Marsha Blackburn	

Appendix E

Least Attractive Candidates

Least Attractive Candidates

 1. Eliot Engel	 2. Grace Napolitano	 3. E. Scott Garrett	 4. John Lewis
 5. Rod Blum	 6. Tom Cole	 7. Ruben Hinojosa	 8. Candice Miller
	 9. Reid Ribble	 10. John Duncan Jr.	

THE SHORT-TERM EFFECT OF MONETARY POLICY ON FINANCIAL MARKETS

By Brian Levine*

This event-study analysis examines the short-term effect of changes in the target federal funds rate (FFR) on day-to-day equity returns, bond yields, and currency exchange rates. I utilize a methodology from Kuttner (2001) that distinguishes between anticipated and unanticipated changes in monetary policy using daily fed funds futures contract data. The analysis reveals that the unanticipated component of the FFR target changes was significant in almost all cases and the magnitude of the policy change's impact depends on the maturity of the Treasury security and level of corporate bond riskiness, but not on the type of currency exchange rate. The results from this study fuse the implications of central bank communication, monetary policy decision-making, and financial market performance, ultimately illustrating the importance of expectations on short-term volatility.

Keywords: Monetary policy; federal funds rate; asset prices; unexpected policy; central bank communication.

I. Introduction

The federal funds rate (FFR) is the interest rate at which U.S. banks make overnight loans to one another. Since the FFR represents the benchmark short-term U.S. interest rate, it has considerable influence over financial markets and other interest rates in the U.S. While the ultimate objectives of U.S. monetary policy are maximum sustainable employment and price stability, these goal variables are impacted with long and variable lags. The most direct and immediate effects of central bank actions are on asset prices and their returns, which are used to evaluate short-term policy effectiveness in the eyes of financial markets.

This event-study style of analysis examines the effect of changes in the target FFR on day-to-day equity returns, bond yields, and currency exchange rates. All policy actions, though, are not created equal. Estimating the market's overall response to monetary policy actions is complicated by the fact that the market is unlikely to respond significantly to policy that was already anticipated. Using a methodology from Kuttner (2001), I distinguish between anticipated and unanticipated changes in policy using daily fed funds futures contract data.

The linear multivariate regression output reveals that the unanticipated component of the FFR target changes on different financial securities is significant in almost all cases,

* I would like to graciously acknowledge and thank Aaron Jackson and David Gulley for their tremendous support and insight. They have served as outstanding mentors and given me the aptitude, passion, and confidence to further pursue monetary policy. Email: brian.levine.01@gmail.com

while the expected component is insignificant and closer to zero in almost all cases. Additionally, the magnitude of the policy change's impact depends on the maturity of the Treasury security and level of corporate bond riskiness, but does not depend on the type of currency exchange rate.

The results from this study illustrate the importance of central bank communication when conducting monetary policy, revealing that enhanced predictability of future policy actions can lead to less volatile short-term financial market outcomes. Comparing these results to those of Bernanke and Kuttner (2005) also provides insight into how Federal Reserve communication has generally improved over time. This is evident through a lower impact of unexpected policy changes on financial markets and a smaller standard deviation of this unexpected component over the more recent sample period. This research could be improved by incorporating the magnitude and direction of the monetary policy changes into the model, incorporating other variables that could affect asset returns and exchange rates, and accounting for risk premia that may bias the interpretation of using fed funds futures to proxy market expectations.

The rest of this paper is organized as follows: Section II provides a brief post-1994 history of FOMC communication and an overview of the fed funds futures market. Section III serves as a literature review on how changes in monetary policy impact both equity and bond markets. Section IV covers the methodology in greater detail, explaining the type of data used, how the unanticipated and anticipated variables were calculated, and the linear multivariate regression equation implemented to complete the analysis. Section V presents the findings of the regression analysis and is broken into subsections reflecting the particular dependent variable used. Section VI discusses the implications of this research for monetary policy via a closer examination of dates with the highest unexpected component values. Section VII summarizes the findings.

II. The Fed Funds Futures Market and FOMC Communication

Before discussing the relevant literature and research results, a short history on the how the Fed has communicated monetary policy and background on fed funds futures is helpful. Since the time frame of my research spans from February 1994 to December 2007, the inner workings of central bank policy at the zero lower bound will not be examined in detail.

The Fed operates in accordance to its dual mandate and works to fulfill these goals by setting a target for the FFR, which serves as a benchmark for other short-term interest rates and consequently influences financial market activity. The Federal Open Market Committee (FOMC) currently releases statements after each meeting explaining what policy action (or lack thereof) had been decided. This practice began in February 1994, instituted "to avoid any misunderstanding of the Committee's purposes," according to the statement. However, the statements were vague and did not incorporate specifics about the magnitude of the FFR change or its new level, forcing markets to infer what the change in the funds rate was only after it was actually implemented. In February 1995, the FOMC decided that

all changes in the stance of monetary policy would be announced after the meeting – as opposed to just the February 1994 one, which came during the first tightening of policy since early 1989 – but it was not until July of that year that mention was made of a specific FFR value.

In 1999, the FOMC initiated the practice of releasing a statement after every meeting regardless of whether the stance of policy had changed. Policy statements were released within a few minutes of 2:15pm the day that the FOMC meeting concludes, though today they are released within a few minutes of 2:00pm. Also starting in 1999, the Committee began to issue forward guidance in the form of perceived risks going forward, including phrases such as “biased toward a possible [future] firming of policy.” Forward guidance continued throughout the early 2000s, but this time in the form of the expected rate path. For example, by 2003, the FOMC noted that policy accommodation would likely be maintained “for a considerable period.” The language was again altered in 2004 with regards to the tightening cycle, stating that “policy accommodation can be removed at a pace that is likely to be measured.” Indeed, the FFR was raised by 25 basis points at its June meeting and at each of the subsequent 16 meetings.

The FOMC also releases minutes of their meetings, which help provide additional insight into what the Committee members are thinking, the state of the economy moving forward, and which factors weigh most on the Committee’s outlook and policy decision. Through 2004, the minutes were released two days after the subsequent meeting, but this release was pushed back to three weeks beginning the following year. The FOMC statements and minutes are important because they give the public context into what to expect moving forward and help financial market participants decipher the Fed’s reaction function.

Fed funds futures also give market participants context for future monetary policy actions. Fed funds futures contracts began trading on the Chicago Board of Trade in late 1988, initially designed as a hedging vehicle against future short-term interest rate movements. These contracts help determine the market probability of a policy maneuver by the Fed. Thus, by looking at the term structure of implied rates on fed funds futures, policymakers are able to gauge market participants’ expectations. This is useful when identifying whether or not a monetary policy decision was anticipated by markets and helps serve as one measure of central bank credibility.

The FFR implied by the futures contract is equal to 100 minus the contract’s settlement price. The fed funds futures quote can be thought of as the average price for fed funds in a particular contract month. For the current month, the contract price is equal to a weighted average of the actual Fed effective rates realized to date and the expected Fed effective rates for the remainder of the month. The contract pricing for future months is based only on expected rates.

As outlined by Keasler and Goff (2007), there are several approaches that can be utilized to determine expectations of Fed policy action. The first involves using fed funds futures contract prices to examine the market’s expectations relating to future interest rates, comparing the FFR implied by the contract to the actual FFR. This approach is similar to

the approach used later in this paper. Alternatively, fed funds futures prices can be used to calculate the probability of a Fed rate change. This method involves first determining the amount by which the implied FFR differs from the current FFR. This number is then divided by the amount of the anticipated policy change.¹ This calculation method may need to be adjusted based on when the FOMC is scheduled to meet relative to subsequent meetings.²

Despite some debate, the general consensus is that fed funds futures contracts are a natural, market-based proxy for expectations of Fed policy actions. Krueger and Kuttner (1996) found that FFR forecasts based on the futures price are “efficient” in that forecast errors are not significantly correlated with other variables known when the contract was priced. Additionally, Gürkaynak et al. (2002, 2005) showed that they are the best predictors of target funds rate changes one to five months ahead compared to other measures of financial market expectations such as term eurodollar rates or the eurodollar futures rate. Evans (1998) similarly documented how fed funds futures outperform forecasts based on alternative methods, such as sophisticated time series specifications or monetary policy rules.

There are a few minor complications, though, when working with fed funds futures contracts as a proxy for market-based expectations. As Robertson and Thornton (1997) point out, the futures rate is a forecast of the average FFR and not a forecast of the average FFR target. They also found the predictive accuracy of fed funds futures rates to be fairly weak, though the bulk of the data used was before the Fed began officially releasing its policy changes to the public in February 1994.

Additionally, using market-based measures of expectations may be biased by risk premia, which refer to the additional return investors require to hold a riskier security compared to that of a risk-free asset. Risk premia are embedded in fed funds futures prices and therefore affect the interpretation of the implied market expectations. Sack (2004) and Piazzesi and Swanson (2008) show that while risk premia primarily exhibit greater magnitude at longer maturities, they can still exert some influence over shorter horizons. Forward guidance theoretically should place downward pressure on risk premia via reduced economic and policy uncertainty. Ben Bernanke in a 2013 speech, however, noted that forward guidance primarily affects longer-term rates through influencing investors’ expectations of future short-term interest rates. Large scale asset purchases (LSAPs) most directly affect term and risk premia, but these occurred after December 2007 when the sample period for this paper ends. Piazzesi and Swanson (2008) explore a time-varying risk adjustment process that could improve the interpretation and accuracy of the results, though this methodology was not applied for this paper.

¹ This calculation is based on a formula developed by Geraty (2000).

² For more on this, see Keasler and Goff (2007).

III. Literature Review

The efficient market hypothesis (EMH) lies at the heart of examining monetary policy's impact on financial markets. The EMH encompasses the idea that when new information arises, this news is efficiently incorporated into the prices of securities without delay. As a result, prices should fully reflect all known information and, in theory, only new information (such as monetary policy actions) that is unanticipated will materially affect financial markets. In other words, if a policy change is well telegraphed in the weeks leading up to the FOMC meeting, market reaction to the policy change (or lack thereof) should not be as volatile because this information was already priced in and expected. This theory implies the unexpected component in my analysis – or the deviation of the fed funds future contract price before and after the meeting – should be close to zero.

There is much debate, however, as to the degree of which the EMH holds in practice. Fama (1965) shows that large daily price changes tend to be followed by additional large changes, but of unpredictable sign. This suggests that the initial first day's adjustment of prices to new information is unbiased. More recent research by Lo and MacKinlay (1999), however, find that there are short-run serial correlations in stock prices, implying prices may not adjust as "unbiasedly" as initially thought. Fama, Fisher, Jensen and Roll (1969) examine stock split announcements to prove the information is fully reflected in the price of a split share at the time of the split, which lends support to the semi-strong form of the EMH. Ball and Brown (1968) and Scholes (1969) come to similar conclusions with respect to the information contained in annual earnings announcements and new issues of common stock.

In relation to this, markets tend to exhibit a speculative behavior in that they drift in anticipation of upcoming news, with the S&P 500 index rising 49 basis points on average in the 24 hours before scheduled FOMC announcements.³ As documented by Lucca and Moench (2011), the realized volatility and trading volume of equities are lower in the hours before FOMC announcements compared to other days, but these indicators jump at the mid-day announcement as the new information becomes incorporated into the equity prices. While this could have at least some bearing on the day-to-day percent change of the equity returns I examine in my research, the average return on the S&P 500 index from right before the announcement until the market close is essentially zero and the authors found no such effects in U.S. Treasury securities. Balduzzi, Elton, and Green (2001) similarly found that bond markets adjust to public news releases relatively quickly, despite the fact that the news contributes to a substantial fraction of bond price volatility in the aftermath of these announcements.

The reaction of financial markets to monetary policy depends on whether this policy was anticipated or unanticipated leading up to the announcement. Bernanke and Kuttner (2005) measured and examined the stock market's response to monetary policy from June 1989 to December 2002. They utilized a methodology introduced in Kuttner (2001) to distinguish between policy changes that are anticipated versus unanticipated via fed funds

³ Lucca and Moench (2011) uses a sample period of 1980 to 2011.

futures contract prices.⁴ The authors found that a hypothetical unanticipated 25 basis point cut in the FFR target is associated with a 1% increase in broad stock indices. This research is beneficial in that it serves as a basis for comparison for more updated data used in my analysis.

Cook and Hahn (1989) also serves as a useful comparison tool, as the authors regressed the change in the Treasury bill, note, and bond rates on the change in the target FFR. Their results showed that the response to the target rate was positive and significant at all maturities, but smaller at the longer end of the yield curve. Kuttner (2001) conducted a similar analysis examining bond rates' response to changes in monetary policy, but this time accounted for whether the policy maneuver was expected. He found that bond rates' response to anticipated changes in the FFR is essentially zero, consistent with the efficient market hypothesis, while their response to unanticipated movements is large and highly significant. This confirmed the results of earlier work such as Roley and Sellon (1995), which revealed that bond prices set in forward-looking markets should respond only to the surprise element of monetary policy actions and not to anticipated FFR movements. Demiralp and Jordà (2004) used a similar methodology to draw conclusions about a security's maturity, finding that as the maturity of the Treasury security increases, the reaction to the surprise component of a target change diminishes. Additionally, target FFR changes executed at a regularly scheduled FOMC meeting and consistent with the general direction of policy has no statistically significant impact on Treasury rates except for the 3- and 6-month Treasury bills.

Central bank communication is an important feature of this research. The public's expectations about future monetary policy actions matter today because those actions have important effects on current financial conditions, which in turn affect output, employment, and inflation over time.⁵ While the existing literature on communication and monetary policy is varied, it supports the general notion that expectations matter. Woodford (2001, 2005) reveals that successful monetary policy is not so much a matter of effective control of overnight interest rates, but the evolution of market expectations. Enhanced transparency is valuable to the conduction of monetary policy – a view that became increasingly widespread amongst central bankers in the late 1990s and early 2000s – and heightened communication helps reduce uncertainty among economic actors and can improve the efficacy of policy.

Ferrero and Secchi (2007) reveal that both qualitative and quantitative communication of future policy intentions improved the ability of market actors to predict policy. One example of quantitative communication is through forecasts. Williams and Rudebusch (2006) find that publishing interest rate forecasts helps align the private sector's and central bank's expectations of future policy actions and reduces fluctuations in output and inflation. Interpretation of a central bank's reaction function is equally important. Bernanke (2004) asserted that when the public does not know, but instead must estimate, the central bank's reaction function, there is no guarantee that the economy will converge to the ratio-

⁴ This methodology will be explained in greater detail in Section IV.

⁵ See Ben Bernanke's November 2013 speech, "Communication and Monetary Policy."

nal expectations equilibrium because the public's learning process affects the economy's behavior.

It is important to point out, however, that what constitutes an "optimal" communication strategy is far from clear. Indeed, as Blinder et al. (2008) address, the key question is to look at to what extent communication contributes to the effectiveness of monetary policy by either moving short-term rates in a desired way or by lowering market uncertainty. The theoretical literature has not generated clear conclusions regarding the optimal level of transparency and acknowledges that there are limits to how much information can be digested by market participants and the public effectively.⁶ There are also instances where human error comes into play or the market misinterprets central bank intention. This occurred, for example, when bond yields skyrocketed in 2013 at the mere suggestion of an imminent reduction in Fed bond purchases, known colloquially as the "Taper Tantrum."

It is evident that communication and transparency matter to at least some degree for a central bank when aiming to conduct effective policy and minimize market instability. This paper uses updated data to examine the aforementioned literature and model how financial markets respond differently to anticipated and unanticipated target FFR changes.

IV. Methodology

The methodology for this paper is largely based off that of Kuttner (2001) as well as Bernanke and Kuttner (2005). Fed funds futures contract data is used to proxy market expectations of future monetary policy actions. Daily futures data spanning February 2, 1994 through December 31, 2007 was used for this research.

This time period was used for several reasons. February 2, 1994 was when the Fed first instituted its current policy of publically and formally announcing changes in the FFR target. The Fed's decision to announce policy changes was accompanied by a decline in futures market forecast errors and greater synchronicity between actual monetary policy and market expectations of monetary policy.⁷ Prior to 1994, investors had to indirectly infer policy actions through the size and type of open market operations in the days following each meeting. Beginning the sample at this date also mitigates uncertainty associated with the timing of policy actions and the ability of markets to interpret the policy change.⁸ Ending the sample at December 31, 2007 avoids the zero lower bound problem, which has been associated with endogenous financial volatility and increased uncertainty that could skew the results. This sample also allows for a diverse set of observations in the sense that it encompasses 31 target rate increases, 22 target rate decreases, and 112 total FOMC meetings.

The surprise element of any specific change in the FFR target can be measured by the change in the fed funds futures contract's price relative to the day prior to the policy action, which is scaled up by a factor related to the number of days in the month affected by the change:

⁶ See Crujeisen and Eijffinger (2007) and Kahneman (2003)

⁷ See Poole (1999).

⁸ Target rate changes occurred more frequently during intermeeting periods prior to 1994.

$$\Delta i^u = \frac{D}{D-d} (f_{m,d}^0 - f_{m,d-1}^0),$$

where Δi^u represents the unexpected component of the target rate change, f represents the current-month futures rate, d represents the day of the month on which the event takes place, and D represents the number of days in the month.⁹ The expected component can then be calculated by subtracting the unexpected component from the actual FFR change:

$$\Delta i^e = \Delta i - \Delta i^u .$$

These unexpected and expected measures of future policy changes serve as the independent variables in the linear regression analysis. While the linear multivariate regression is similar to what was used in Bernanke and Kuttner (2005), this research incorporates an interchangeable set of dependent variables to increase the scope of the analysis and uses a more updated data set. The observations concerned for this research include dates when the FOMC met and changed policy, dates when the FOMC met and did not change policy, and dates when policy was changed outside of an FOMC meeting (e.g. during a conference call), with dummy variables assigned to each observation date for a total of 117 observations.

It should be noted that five observations – 15 October 1998, 3 January 2001, 20 March 2001, 18 April 2001, and 17 September 2001 – are excluded from the sample because they exhibit particularly high influence statistics and are outliers in economic terms.¹⁰ In other words, I only want to examine reactions to monetary policy under “normal” circumstances. In particular, the outliers during the easing cycle of 2001 were characterized by unusually vehement market reactions, while the 1998 rate cut aligned with deteriorating situations in Asia and Russia and the September 17, 2001 observation followed the terrorist attacks which sparked unpredictable market volatility. An organized breakdown of the observations is shown in Table 1.

The linear regression is implemented using the form

where Y_t represents the day-to-day percent change in either equity returns, Treasury

$$Y_t = a + b^e \Delta i_t^e + b^u \Delta i_t^u + \varepsilon_t ,$$

yields of different maturities, corporate bond indices of different risk levels, or specific cur-

⁹ The unscaled change in the one-month futures rate is used to calculate the funds rate surprise when the change falls on one of the last three days of the month. This is to minimize the effect of any month-end noise in the effective funds rate.

¹⁰ Bernanke and Kuttner (2005) contains further detail on the influence statistic calculations. In summary, the authors observed the change in the estimated parameters of the regression that resulted from dropping a given observation, t . The distribution of these calculations revealed that statistics in excess of 0.3 exerted an unusually large influence on the estimates.

rency exchange rates. This general form allows for the comparison of how different assets are affected by an unexpected or expected policy action. An important feature of the regression model is that it imposes symmetry, meaning that a target rate increase or decrease has the same numerical impact according to the model. This is beneficial for simplicity purposes, but distinguishing between rate increases and decreases would likely alter the results and interpretation.

In particular, this research examines the effects of monetary policy on continuous S&P 500 returns, the percent change in the 2-, 5-, 10-, and 30-year Treasury yields,¹¹ investment grade and below investment grade yields of corporate bond indices,¹² as well as the U.S. dollar index, Yen/USD exchange rate, and USD/Pound exchange rate. Daily data is used for all data sets.

TABLE 1

Year	Maintained Target FFR	Adjusted Target FFR	
1994	3	6	
1995	5	3	
1996	7	1	
1997	7	1	
1998	6	3	
1999	5	3	
2000	5	3	
2001	0	11	
2002	7	1	
2003	7	1	
2004	3	5	
2005	0	8	
2006	4	4	
2007	5	3	
Total	64	53	117

The data for the above observations spans from February 2, 1994 to December 31, 2007. “Maintained Target FFR” refers to an FOMC meeting date where policy was not adjusted. “Adjusted Target FFR” refers to any date where policy was changed (i.e. during an FOMC meeting or conference call date).

¹¹ The mid yield was used for each measure.

¹² The Bank of America Merrill Lynch U.S. Corporate Master Effective Yield and the Bank of America Merrill Lynch U.S. High Yield Effective Yield were used to track the performance of investment grade and below investment grade U.S. corporate debt, respectively. These data sets were only available starting December 31, 1996.

V. Results

The results of the regression analysis are broken down into subcategories. A summary of each regression's output can be found in Table 2 and all regression coefficients are interpreted at the 5% level. As previously mentioned, the observations concerned for this research include dates when the FOMC met and changed policy, dates when the FOMC met and did not change policy, and dates when policy was changed outside of an FOMC meeting (e.g. during a conference call). After excluding the five outliers, this produces a total of 112 observations. Since the two Bank of America Merrill Lynch corporate bond indices only begin on 31 December 1996, these regressions contain 97 observations instead of 112.

It is not uncommon for the R^2 to be low when dealing with changes in financial data. It is difficult to accurately predict day-to-day changes in asset prices or yields given that they are a function of a complex variety of factors. Additionally, the results may be skewed based on variables that have been omitted from the regression analysis but have a material effect on asset prices or yields. However, it is impractical for the purposes of this paper to try and include *every* variable that could potentially impact the results. The goal of this research is also more narrowly focused; it only seeks to examine how financial markets react to changes in monetary policy specifically.

The coefficient of the unexpected component was significant for almost every data series. Policy that is not anticipated by financial markets should create a more volatile movement in asset returns on the day the policy is announced. On the other hand, the coefficient of the expected component was insignificant for almost every data series, and when the coefficient was significant, its value was very close to zero. An event-study breakdown of days where the absolute difference between the fed funds futures implied FFR and the actual FFR is greatest lends insight into the disconnect between Fed policy and market expectations. This will be discussed in Section VI.

Current best practice in central banking views a high level of monetary policy predictability as desirable.¹³ Communication impacts financial market pricing, which has implications for both the real economy and for the Fed's credibility to impact real economic outcomes in the future. The analysis in this paper predominantly concerns short-term predictability, referring to the ability of the public and financial markets to anticipate day-to-day monetary policy decisions. If policy changes are well telegraphed by a central bank and markets are prepared for these changes, the difference between the fed funds futures implied FFR and the actual FFR should be small, and markets should exhibit little volatility strictly associated with this action.

S&P 500

I first regressed the unexpected and expected components against daily continuous S&P 500 returns. The results of the regression analysis revealed both the unexpected and the expected coefficients were significant at the 5% level, although the expected coefficient

¹³ See Blattner et al. (2008).

was fairly close to zero at -0.00725 . The unexpected coefficient was greater in magnitude and had the expected negative sign, revealing that a one percentage point surprise rate change would result in a -0.02 percentage point change in the one-day S&P 500 continuous returns. The R^2 of the model was 0.0815 , which indicates 8.15% of the variation in the S&P 500 returns can be associated with news about monetary policy actions. Compared with the regression results for Treasury securities, corporate bond indices, and currency exchange rates, the S&P 500 output contains one of the highest t-statistics. In linear regression models, t-statistics are useful for making inferences about the regression coefficients, essentially testing the hypothesis that the true value of the coefficient is non-zero. This indicates the S&P 500 seems to be highly sensitive to monetary policy news relative to other financial assets or indices. See Table 2 for the regression output.

2-YEAR, 5-YEAR, AND 10-YEAR TREASURY NOTES AND 30-YEAR TREASURY BOND

The results of how Treasury yields react to changes in unexpected and expected policy actions vary by maturity, which is consistent with the literature. More specifically, as the maturity of the Treasury security increases, the magnitude of the unexpected coefficient decreases. For example, if there is a one percentage point surprise rate cut, the one-day percent change in the 2-year Treasury note's yield will change by 0.10634 percentage points in response. However, the one-day percent change in the 5-year, 10-year, and 30-year yield is only 0.05364 , 0.03099 , and 0.01425 percentage points, respectively. The R^2 for each maturity exhibits a similar trend. Over 22% of the variation in the day-to-day 2-year yield can be attributed to monetary policy actions. However, only 13% , 9% , and 4% of the variation in the 5-year, 10-year, and 30-year yields, respectively, can be attributed to monetary policy actions. See Table 2 for the regression outputs.

The difference in the coefficient value between the 2-year and 5-year Treasury notes is the largest, while the difference in the coefficient value between the 10-year Treasury note and 30-year Treasury bond is the smallest. The maturity of the Treasury security influences, to some at least some degree, how significantly the day-to-day percent change in yield moves in response to unexpected changes in monetary policy. The unexpected coefficient for all maturities is positive and significant, with the exception of that for the 30-year Treasury. A positive unexpected coefficient suggests that the prices of Treasuries fall in response to unanticipated policy news, pushing up their yields as a result. The expected coefficient is insignificant for all maturities, which suggests that Treasury yields do not exhibit significant day-to-day movement in response to predictable changes in monetary policy.

These results confirm an important characteristic about how Treasury yields of different maturities react to Fed target rate changes. The general idea is that central banks have better control over short rates, whereas long rates are impacted by a multitude of factors. In the short end of the yield curve, a rate cut is most directly associated with a lower short-term interest rate. But at the long end, the results are not as robust because expansionary policy could cause changes to inflation expectations or liquidity premia, which exert additional influence on long-term rates.

TABLE 2

Regression Results

Regressor	S&P 500	2 Year	5 Year	10 Year	30 Year	IG	Below IG	USD Index	Yen / USD	USD / GBP
Intercept	0.002 (1.97)	0.000 (0.04)	-0.001 (-0.36)	-0.001 (-0.95)	-0.001 (-1.35)	-0.000 (-0.01)	0.002 (1.61)	-0.004(- 1.02)	0.000 (0.69)	0.001 (1.22)
Expected Change	-0.007* (-2.78)	0.003 (0.36)	0.002 (0.24)	0.000 (0.05)	-0.000 (-0.08)	-0.001 (-0.39)	0.009* (2.45)	-0.002 (-1.51)	-0.002 (-1.00)	0.001 (0.73)
Unexpected Change	-0.022* (-3.06)	0.106* (2.54)	0.054* (2.62)	0.031* (2.32)	0.014 (1.35)	0.027* (4.37)	0.005 (0.72)	0.004 (1.02)	0.000 (0.04)	0.000 (0.06)
R-squared	0.082	0.228	0.126	0.090	0.037	0.107	0.056	0.032	0.009	0.005

This is the regression output table for each of the associated regressors. Parentheses contain t-statistics and asterisks denote statistical significance for the 0.05 level.

INVESTMENT GRADE, BELOW INVESTMENT GRADE CORPORATE BOND INDICES

The regressions involving the two Bank of America Merrill Lynch bond indices – representing investment grade and below investment grade rated corporate debt publically issued in the U.S. domestic market – provided slightly less predictable results. The investment grade coefficient for the unexpected component was significant and was associated with a very high t-statistic, as anticipated. More specifically, a surprise rate change of one percentage point will cause the one-day percent change in the investment grade index to move by 0.02672 percentage points. However, the unexpected and expected coefficients for the below investment grade bond index came back insignificant and significant, respectively, which is slightly counterintuitive. See Table 2 for the regression outputs.

Comparing these results to the existing literature does not reduce the ambiguity. Kon-tonikas, Maio, and Zekaite (2016) investigated the impact of monetary policy shocks on the excess returns of U.S. investment grade corporate bonds from 1989 to 2013. They find a negative and significant response of excess returns on corporate bonds to unanticipated FFR target changes, drawing similar conclusions to the results obtained from my regression analysis. A key determinant of the variability in the current unexpected excess returns, the authors explain, is the variance of the revisions in expectations about future bond risk premia, while the effects of monetary policy shocks on the expectations of future inflation and real interest rates are relatively small by comparison. This makes the below investment grade regression results all the more surprising because one would think that riskier companies would react more strongly to changing policy news and be more sensitive to changing economic conditions. One possible explanation is that risk premia adjustments for below investment grade bonds exhibit greater unpredictability and are affected by more influential factors than monetary policy alone, such as firm-specific conditions.

U.S. DOLLAR INDEX, YEN/USD EXCHANGE RATE, USD/POUND EXCHANGE RATE

The unexpected coefficients came back insignificant and close to zero for the U.S. dollar index, yen/USD exchange rate, and USD/pound regressions.¹⁴ These specific exchange rates are examined because they are three of the top four most commonly traded currencies on a global scale. The results indicate there is minimal impact of unanticipated policy actions on currency markets. See Table 2 for the regression output.

While the value of currencies over the medium-term tend to move with interest rates, day-to-day movements in their value following monetary policy action may have a more subdued effect. Indeed, studies show that over short horizons, exchange rates can follow many paths that do not correspond to the predictors from recent monetary policy actions.¹⁵ This is especially true given that currency indices and exchange rates are a factor of a wide

¹⁴ While the U.S. dollar index incorporates the Japanese yen and Pound sterling into its calculations, the index reflects a weighted geometric mean of the dollar's value relative to a multitude of other currencies as well. Regressing it separately was intended to help draw additional conclusions and hopefully provide greater insight.

¹⁵ For example, see Scheld and Allardice (1994) as well as Grilli and Roubini (1995).

variety of global influences. The endogeneity of interest rates with respect to exchange rates and investor expectations also make it extremely difficult to use statistical analysis to identify the impact of monetary policy on the exchange rate. Moreover, policy *divergence*, not individual policy actions, ultimately determines relative exchange rate movements. It should be noted, however, that in response to the Fed's quantitative easing (QE) announcements there was certainly intraday movement of the dollar against other major currencies, though this impact diminished with each subsequent QE program.

VI. Event-Study Analysis and Implications for Monetary Policy

The results of this study provide meaningful implications for monetary policy and highlight the significance of central bank communication.

Communication is an important and powerful part of any central bank's toolkit since it has the ability to move financial markets, to enhance the predictability of monetary policy decisions, and potentially to help achieve a central bank's macroeconomic objectives. The ability of policymakers to influence market expectations is argued to be of equal or greater importance than the actual practice of controlling overnight interest rates. Decomposing monetary policy actions into those that markets anticipate versus those that markets do not anticipate helps us understand how past policy maneuvers could be improved to better manage expectations.

The following dates are associated with the highest absolute value of the unexpected components: 20 December 1994, 26 September 1995, 24 September 1996, 25 June 2003, and 18 September 2007.¹⁶ In a standard event-study manner, re-examining what occurred on these dates lends insight into what may have caused this disconnect between market expectations and Fed policy action.

On 20 December 1994, the FOMC met but decided to maintain the target FFR at 5.5%. The FOMC had raised its target FFR six times that year prior to its December meeting, with the magnitude of the tightening increasing from 25 basis points (in February, March, and April) to 50 basis points (in May and August) to 75 basis points (in November). A closer inspection at the minutes reveal that nonfarm payroll "rose sharply" in November after an "appreciable expansion in October," and the civilian unemployment rate declined to 5.6%. Markets likely expected this tightening to continue, especially given the 75 basis point increase the meeting before and the continuing improvement in the labor market. Indeed, one voting member dissented because he favored an immediate policy tightening action. The FOMC "did not include in the [November meeting] directive a presumption about likely further adjustments to policy," which may explain some of the associated ambiguity. Mexico's government also enacted a surprise devaluation of the peso the same day of the December FOMC meeting, which prompted currency swap discussions between the Bank of Mexico and the Fed and likely blindsided financial market participants.

¹⁶ The dates 21 April 2001 and 17 September 2001 also exhibited very high unexpected component values, but since these observations were excluded from the regression analysis, they were not discussed in this section.

The 25 September 1995 observation again featured an FOMC meeting but no formal policy action. The market's reaction here is a bit more surprising. Even though the FOMC changed the direction of policy in July with a 25 basis point decrease, they held rates steady at their meeting in August. Also similar to the December 1994 decision, the September 1995 decision was accompanied by increases in nonfarm payroll employment, a drop in the civilian unemployment rate, and other positive economic indicators such as large gains in industrial production. There was no guidance to the future path of interest rates or policy maneuvers as well, though this meeting featured no dissents from voting members.

Only one policy rate change occurred in 1996, and that was on January 31st where the FOMC voted to decrease the target FFR by 25 basis points. When the FOMC met on 24 September 1996, they voted to maintain the FFR at its existing level of 5.25%, although one participant dissented on the basis that a more restrictive policy was required to preempt future inflation. The *New York Times* noted that "the market remained deeply divided over what course the Fed would take... [as] since the last Federal Open Market Committee meeting in August... there has been evidence to support both a rate increase and no increase."¹⁷ Evidently, Reuters put out a report the week prior to the meeting that documented eight of the 12 regional Fed presidents favoring a rate increase. This explains why the lack of policy action may have caught markets by surprise.

Unlike the previous three observations, the FOMC meeting of 25 June 2003 was associated with a policy action in the form of a 25 basis point rate cut, the first FFR target change since November 2002. The Committee cited the lack of sustainable economic growth and minor concerns about falling inflation as the main justifications for the policy change, with one member dissenting and calling for a 50 basis reduction in the target FFR instead. *CNN Money* reported that Fed officials had "built up market expectations for a rate cut for weeks," but were unsure about its magnitude. A Reuters poll of major banks that conduct business directly with the Fed indicated a majority expected a 50 to 75 basis point cut in the target FFR. The *Wall Street Journal* reported, though, that policymakers worried about the repercussions of taking the FFR so low.

Finally, the 50 basis point target rate cut during the 18 September 2007 meeting was both the first since June 2006 and featured a change in the direction of policy. The Committee cited tightening credit conditions, which had the potential to intensify the housing correction and to restrain economic growth more generally, as the reason for this policy action. Red flags in the subprime mortgage market, specifically, prompted Chairman Bernanke to lead a greater-than-expected policy action to "help forestall some of the adverse effects on the broader economy." Similar to the previous observation, the Fed was widely expected to cut rates, but markets were unsure whether about the magnitude. A few major financial institutions such as Morgan Stanley and Bear Stearns anticipated a 25 basis point move in conjunction with a more aggressive move on the discount rate, while others like Goldman Sachs and Deutsche Bank expected a 50 basis point easing. These financial institution estimations of future policy action are publically available and likely had some bearing on the expectations of other market participants.

¹⁷ See <http://www.nytimes.com/1996/09/24/business/bond-prices-are-higher-on-thin-day.html>

The event-study results indicate that observations associated with a disconnect between market expectations and Fed action result in a higher unexpected component and less favorable financial market movements. This disconnect is exacerbated by conflicting public views of those who influence policy, such as opinionated speeches by policymakers that reach differing conclusions about the economy or the future path of interest rates. Aligning market expectations with future monetary policy decisions is no easy task; there will always be at least some degree of uncertainty surrounding future monetary policy actions. Former chairman Alan Greenspan – in an April 2016 panel discussion with Janet Yellen, Ben Bernanke, and Paul Volcker – admitted that monetary policy is essentially economic forecasting, and policymakers' ability to forecast is significantly limited. Even when Fed officials successfully communicate their intentions to act, markets may still experience short-term volatility if the policy action's magnitude is less than or greater than anticipated.

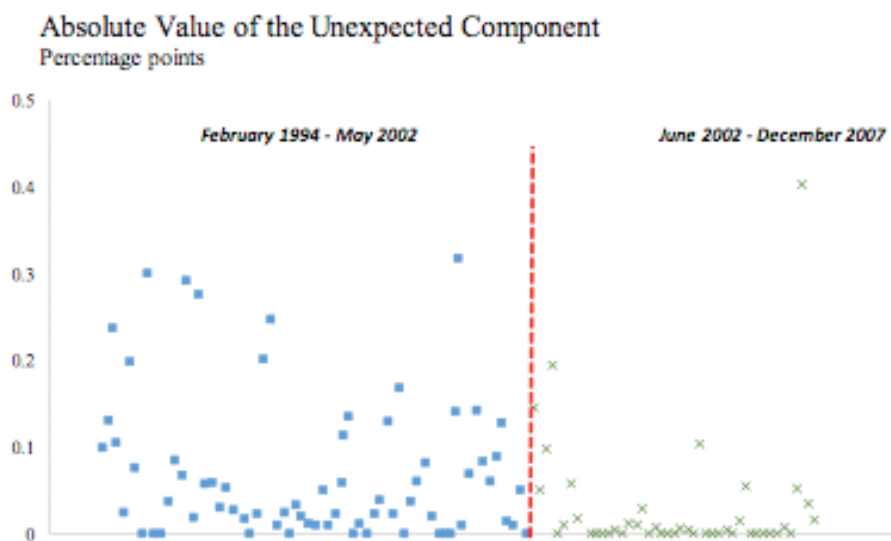
The amount of information released by the Fed for financial market interpretation has increased substantially over the sample period used in this event study. For example, the word count of FOMC statements increased from around 150 in the mid-1990s to around 300 in 2006, 500 in 2010, and 850 in 2015. The number of FOMC member speeches has increased substantially as well, rising from around 140 per year in the late 1990s to over 200 by 2007. The FOMC has also provided more insight regarding risks to their forecasts and the future path of interest rates.

The increased use and detail of forward guidance and other measures of transparency should theoretically reduce the unexpected component magnitude through aligning market expectations with Fed intentions and increasing the efficiency of monetary transmission channels. I calculated the standard deviation of the unexpected component over the sample period to compare these results to those of Bernanke and Kuttner (2005), who calculated a standard deviation of 10.4 from May 1989 to January 1994 and of 9.5 from February 1994 to December 2002. My calculations over the February 1994 to December 2007 sample reveal a standard deviation of 3.6. Additionally, Figure 1 displays how the absolute value of the unexpected component diminishes over time. By comparing a sample period encompassed within Bernanke and Kuttner (2005) to a more recent sample period, we observe how the disconnect between financial market expectations and Fed policy action has diminished substantially over time, attributable to increased use of forward guidance and the predictable, measured pace of the 2004-2006 tightening cycle.

I then isolated Treasury securities to examine the effects of more updated data and determine if similar conclusions could be drawn. I re-ran regressions for the 2-year, 5-year, and 10-year Treasury notes and 30-year Treasury bond, using the February 1994 to May 2002 data for the first regression and June 2002 to December 2007 data for the second. The reaction of the more recent period is greater in magnitude for Treasuries of shorter maturities; since these securities are typically most sensitive to incoming monetary policy news, they bear greater importance for the regression interpretation. The results imply that an unexpected change in monetary policy over the more recent period is less likely (Figure 2). In other words, because the Fed was communicating more over this period and there was greater predictability of monetary policy actions, when there was an unanticipated policy

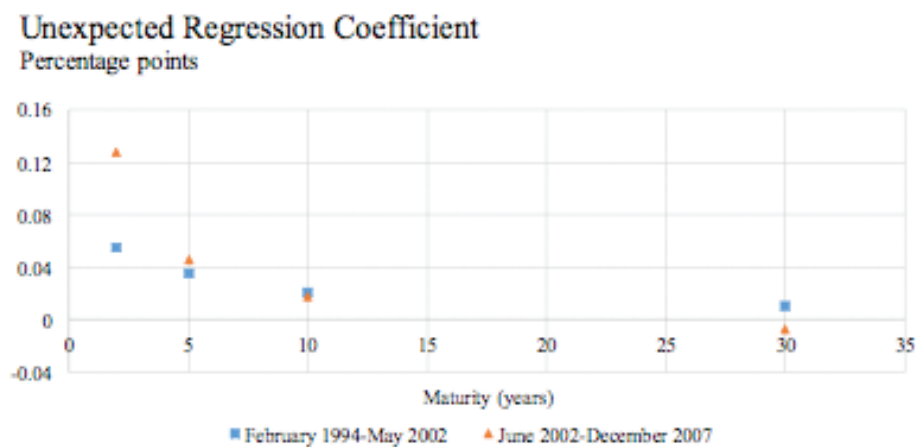
change, it generated a bigger surprise reaction than during the previous period. Overall, these assessments are consistent with the notion that Fed communication has generally improved over time to help better align market expectations with Fed policy and reduce uncertainty around FOMC meetings.

FIGURE 1



The data points represent the absolute value of the unexpected component for February 1994 to May 2002 as well as June 2002 to December 2007. The samples are divided as shown because the Bernanke and Kuttner (2005) study ends at May 2002 and serves as a useful comparison for more updated data. As the graph illustrates, the difference between what markets expect and what the Fed actually does diminishes when comparing the two data sets side-by-side. Reasons include increased and more explicit use of forward guidance as well as the predictable, measured pace of the 2004-2006 tightening cycle.

FIGURE 2



This graph plots the unexpected regression coefficients for Treasury securities of different maturities. The square data points represent these coefficients for the February 1994 – May 2002 sample, while the triangle data points represent these coefficients for the June 2002 – December 2007 sample. The more recent sample period yields a stronger financial market reaction to unexpected monetary policy actions for Treasuries of shorter maturities, which are typically most sensitive to incoming monetary policy news.

VII. Conclusion

This event-study analysis supports how central bank communication can influence expectations-driven uncertainty and improve financial market outcomes. The study uses daily fed funds futures contract data spanning February 1994 to December 2007 to proxy market-based expectations of future Fed policy decisions. These expectations, decomposed into an anticipated and unanticipated component, are regressed against S&P 500 continuous returns, Treasury securities of varying maturities, investment and below investment grade corporate bond indices, and exchange rates of highly-traded currencies.

The results are consistent with the literature, suggesting monetary policy that is unexpected will have a greater impact on short-term financial market outcomes. The degree of this impact depends on a variety of factors. From a financial asset standpoint, this includes the maturity of the Treasury security and level of corporate bond riskiness. From a communication standpoint, this includes speeches by Fed officials and reports by major financial institutions and news publications, all of which can work to align or divide market and Fed expectations. An important takeaway from this paper is that Fed communication has generally improved over time, which has helped reduce both uncertainty and immediate undesired reactions in financial markets.

References

- Balduzzi, P., Elton, E. J., & Green, T. C.** (2001). "Economic News and Bond Prices: Evidence from the U.S. Treasury Market." *Journal of Financial and Quantitative Analysis*, 36(4), 523-543.
- Ball, R., & Brown, P.** (1968). *Journal of Accounting Research*, Autumn, 159-178.
- Bernanke, Ben S** (2013). "Communication and Monetary Policy." Herbert Stein Memorial Lecture. National Economists Club Annual Dinner, Washington, D.C. 19 Nov. 2013. Board of Governors of the Federal Reserve System.
- Bernanke, Ben S** (2004). "Fedspeak." At the Meetings of the American Economic Association, San Diego, California. 3 January 2004. Board of Governors of the Federal Reserve System.
- Bernanke, B. S., & Kuttner, K. N.** (2005). "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance*, 60(3), 1221-1257.
- Blattner, T. S., Catenaro, M., Ehrmann, M., Strauch, R., & Turunen, J.** (2008). "The Predictability of Monetary Policy." *ECB Occasional Paper No. 83*.
- Blinder, A. S., Ehrmann, M., Fratzscher, M., Haan, J. D., & Jansen, D.** (2008). "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." *NBER Working Paper*, 13932.
- Board of Governors of the Federal Reserve System** (1994). FOMC Statement, press release, December 20, 1994.
- Board of Governors of the Federal Reserve System** (1995). FOMC Statement, press release, September 26, 1995.
- Born, B., Ehrmann, M., & Fratzscher, M.** (2011). "Central Bank Communication on Financial Stability." *ECB Working Paper Series*, 1332(April).
- Carlson, John B., Ben Craig, Patrick Higgins, and William Melick.** (2006). "FOMC Communications and the Predictability of Near-Term Policy Decisions," *Federal Reserve Bank of Cleveland, Economic Commentary*, June.
- Chan, L. K., Jegadeesh, N., & Lakonishok, J.** (1996). *Momentum Strategies. Journal of Finance*, 51(5), 1681-1713.
- Cook, T., & Hahn, T.** (1989). "The Effect of Changes in the Federal Funds Rate Target on Market Interest Rates in the 1970s." *Journal of Monetary Economics*, 24(9), 331-351.
- Demiralp, Selva and Oscar Jordà** (2004). "The Response of Term Rates to Fed Announcements," *Journal of Money, Credit and Banking*, 36(3): 387-405.
- Evans, Charles.** (1998). "Real-time Taylor Rules and the Federal Funds Futures Market." *Federal Reserve Bank of Chicago, Economic Perspectives*, 2, 44-55.
- Fama, E.** (1965). "The Behavior of Stock Market Prices." *Journal of Business*, 38, 34-105.

- Fama, E.** (1969). "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, 25(2), 383-417.
- Fama, E.** (1998). "Market efficiency, long-term returns, and behavioral finance." *Journal of Financial Economics*, 49, 283-306.
- Fama, E., Fisher, L., Jensen, M. C., & Roll, R.** (1969). "The Adjustment of Stock Prices to New Information." *International Economic Review*, 10.
- Ferrero, G., and A. Secchi.** (2007). "The Announcement of Future Policy Intentions." Bank of Italy working paper.
- Geraty, Michael,** (2000). "How to Calculate the Odds of a Change in the Fed Funds Rate," Bianco Research, L.L.C., January.
- Grilli, Vittorio and Nouriel Roubini.** (1995). "Liquidity and Exchange Rates: Puzzling Evidence From the G-7 Countries." New York University: Salomon Brothers Working Paper: S/95/31, July 1995.
- Gürkaynak, Refet S., Sack, Brian P., & Swanson, Eric T.** (2002). "Market-Based Measures of Monetary Policy Expectations." Manuscript, Board of Governors of the Federal Reserve System.
- Gürkaynak, Refet S, Sack, Brian P., & Swanson, Eric T.** (2005). "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models." *The American Economic Review*, 95(1): 425-436.
- Keasler, T. R., & Goff, D. C.** (2007). "Using Fed Funds Futures to Predict a Federal Reserve Rate Hike." *Journal of Economics and Finance Education*, 6(2), 9-14.
- Kontonikas, A., Maio, P., & Zekaite, Z.** (2016). "Monetary Policy and Corporate Bond Returns." Working paper, University of Glasgow.
- Kreuger, Joel and Kenneth Kuttner.** (1996). "The fed funds futures rate as a predictor of Federal Reserve Policy." *The Journal of Futures Markets*, 16(December), 865-79.
- Kuttner, Kenneth N.** (2001). "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics*, 47(3), 523-24.
- Lo, A. W., & MacKinlay, A. C.** (1999). *A Non-Random Walk Down Wall Street*. Princeton, NJ: Princeton University Press.
- Lucca, D. O., & Moench, E.** (2011). "The Pre-FOMC Announcement Drift." *Federal Reserve Bank of New York Staff Reports*, 512.
- Malkiel, B. G.** (2003). "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives*, 17(1), 59-82.
- Mishkin, F. S.** (1978). "Efficient-Markets Theory: Implications for Monetary Policy." *Brookings Papers on Economic Activity*, 3, 707-768.

- Piazzesi, Monika and Swanson, Eric T.** (2008). "Futures prices as risk-adjusted forecasts of monetary policy." *Journal of Monetary Economics*, 55, 677-691.
- Poole, W.** (1999). *Synching, not sinking, the markets*. Speech, August 6, 1999, presented in Council for Business Economics, Philadelphia.
- Rafferty, M., & Tomljanovich, M.** (2002). "Central bank transparency and market efficiency: An econometric analysis." *Journal of Economics and Finance*, 8, 487-510.
- Robertson, John C., and Daniel L. Thornton.** (1997). "Using Federal Funds Futures Rates to Predict Federal Reserve Actions." *The Review*, November/December, 45-53.
- Roley, V. V., & Sellon, G. H., Jr.** (1995). "Monetary Policy Actions and Long-Term Interest Rates." *Economic Review*, 80(Fourth Quarter), 73-89.
- Rudebusch, G. D., & Williams, J. C.** (2006). "Revealing the Secrets of the Temple: The Value of Publishing Central Bank Interest Rate Projections." *Federal Reserve Bank of San Francisco Working Paper Series*.
- Sack, Brian** (2004). "Extracting the Expected Path of Monetary Policy From Futures Rates." *Journal of Futures Markets*, 24, 733-54.
- Scheld, K. A., & Allardice, D.** (1994). "The dollar and the federal funds rate." *Chicago Fed Letter*, 84.
- Scholes, Myron** (1969). "A Test of the Competitive Hypothesis: The Market for New Issues and Secondary Offerings." Unpublished PH.D. thesis, Graduate School of Business, University of Chicago, 1969.
- Woodford, M.** (2005). "Central Bank Communication and Policy Effectiveness." *NBER Working Paper No. 11898*.
- Woodford, Michael.** (2001). "Monetary Policy in the Information Economy." In: *Economic Policy for the Information Economy*. Kansas City: Federal Reserve Bank of Kansas City, 297-370.
- Wynne, M. A.** (2013). "A Short History of FOMC Communication." *Economic Letter*, Federal Reserve Bank of Dallas, 8(8).

PATIENT-PROVIDER DISCUSSION ABOUT NUTRITION DURING ROUTINE VISITS: FREQUENCY, QUALITY, AND OUTCOMES

By Kerriann Fitzgerald & Danielle Blanch-Hartigan*

With over two-thirds of the adult population in the US overweight or obese, discussion about nutrition between patients and providers during routine visits may be a point of intervention. However, a strong evidence-base is needed for how often these discussions occur, the quality and characteristics of these discussions, the relationship with patient weight, and how these discussions influence patient satisfaction, perceptions of patient-centered communication, and behavioral intention. This study used an online Mechanical Turk survey of 339 participants who had seen a doctor for a routine visit in the past year. Only 16.2% of participants had discussed diet and nutrition in detail with their doctor at their last routine visit. 45.1% said they had discussed these topics a little, and 123 (36.3%) had not discussed these topics with their doctor at all. The majority of these discussions lasted less than 5 minutes. Discussion about nutrition-related topics was associated with higher overall satisfaction with communication and more patient-centered communication. Doctors were more likely to have these conversations with patients with greater body mass index. Increased nutrition-related discussion between doctors and patients at routine checkups was associated with an increased intent to change behavior and eating habits following the conversation. The positive association between doctor-patient communication about nutrition and increased intention to change nutrition-related behavior, patient satisfaction, and perceptions of patient-centered communication suggests that doctors may want to increase not only the frequency but the quality of their discussions about nutrition.

Keywords: Nutrition; doctor-patient communication; patient-centered care; patient satisfaction; obesity.

I. Introduction

There is a growing obesity epidemic in the United States. According to the Centers for Disease Control and Prevention (CDC), 70.7% of the population in the United States is overweight or obese, with a BMI higher than 25.0 (CDC, 2016). The increasing obesity epidemic in the country is related to a number of health conditions including heart disease, stroke, type 2 diabetes, and cancers. Additionally, the associated medical costs resulting from obesity-related complications was approximately \$147 billion in 2008 (CDC, 2015).

Patient-provider communication may be one touchpoint for bending the curve in the obesity epidemic. Patient-centered communication in particular is key to the quality and

* The study was supported through an honors research fellowship grant to Kerriann Fitzgerald by the United Technologies Corporation. Blanch-Hartigan: Assistant Professor, Bentley University. dhartigan@bentley.edu.

effectiveness of discussions between patients and their healthcare providers. Patient-centered communication is defined by the Institute of Medicine as “healthcare that establishes a partnership among practitioners, patients, and their families (when appropriate) to ensure that decisions respect patients’ wants, needs, and preferences and that patients have the education and support they need to make decisions and participate in their own care (Institute of Medicine, 2001).” This includes six key elements: 1. Fostering a healing patient-provider relationship through building rapport and trust; 2. Exchanging clinical information and understanding patients’ representations of that information; 3. Responding to patients’ emotional needs; 4. Helping patients manage uncertainty; 5. Involving patients in the decision-making process; and 6. Enabling patient self-management through supporting patient autonomy and providing appropriate resources (Epstein & Street, 2007). The impact of patient-centered care on a patient’s satisfaction is well established. Research demonstrates that a more patient-centered communication style is related to more patient satisfaction and better patient outcomes including more appropriate diagnosis and treatment, increased patient adherence, and even reduced malpractice and lower costs (for reviews see Stewart, 1995; Kelley, Kraft-Todd, Schapira, Kossowsky, & Riess, 2014).

Although we have extensive information about the impact of patient-centered communication in clinical interactions in general, less work has looked specifically at the quality and patient-centeredness of communication about nutrition. Previous research has demonstrated that discussions about weight and nutrition are not ubiquitous. In a study of PCPs, 8.9% PCPs provided 52% of all weight counseling and 58% of PCPs did not discuss weight at all with their patients (Kraschnewski et al., 2015). Studies have explored the impact of routine nutrition counseling completed by a primary care physician on patients’ long-term health decisions. Although the overall effect sizes of these studies are low to moderate, they highlight the role of routine care in communication about nutrition (Newland, 2003). Additional research has been conducted regarding whose responsibility it is to counsel patients on nutrition (Kolasa, 2010). Some health care providers believe that it is, in fact, a duty of a primary care doctor to include nutrition counseling within a routine checkup, while others find that a patient seeking such information must be referred to a dietician or nutritionist (Kolasa, 2010). The patient perceptions of the quality of this communication is less understood.

The present study used an online survey research design to explore the frequency and quality of patient-provider communication about nutrition during routine care. We examined the relationship between nutrition communication and patient satisfaction. In addition, we assessed whether having conversations about nutrition impacts a patient’s dietary and nutrition-related habits following the routine checkup. We also assessed the impact of patient overweight status on nutrition discussions.

II. Methods

All study materials and procedures were reviewed by Bentley University Institutional Review Board. The study was funded by an honors research fellowship grant to Kerriann Fitzgerald by the United Technologies Corporation. This study distributed a survey to participants via Mechanical Turk, an online platform offered by Amazon. Each participant was paid a small sum, ranging from \$0.50 to \$2.50 through Mechanical Turk. The survey took approximately 10 to 15 minutes to complete. Participants who were less than 18 years of age, who did not consent to the study, or who had not visited their doctor for a routine checkup within the past year were excluded from the study. Survey items were adapted from the Health Information National Trends Survey (HINTS).

Participants were asked how they perceived the level of care received at their most recent routine checkup, defined as a general physical exam, not an exam for a specific injury, illness, or condition. Participants were asked six questions corresponding to the six functions of patient-centered communication: *“How often did the doctor you saw for your last routine checkup do each of the following?”*

1. Give you the chance to ask all the health-related questions you had? (exchanging information)
2. Give the attention you needed to your feelings and emotions? (responding to emotions)
3. Involve you in decisions about your health care as much as you wanted? (making decisions)
4. Make sure you understood the things you needed to do to take care of your health? (enabling self-management)
5. Help you deal with feelings of uncertainty about your health or healthcare? (managing uncertainty)
6. At your last routine check-up, how often did you feel you could rely on your doctor to take care of your health care needs? (fostering healing relationships)

Response options were always, usually, sometimes or never. Participants were also asked to rate the overall communication with their doctor at their last routine check-up as excellent, very good, good, fair or poor.

The survey then asked participants to indicate which topics were discussed at their last routine check-up with their primary care provider. Response options were “no, did not discuss”, “yes, discussed a little”, and “yes, discussed in detail.” Topics were: weight or weight loss; fitness, exercise or physical activity; and food, nutrition, or diet. A follow-up question was then presented to those who had responded that they had not discussed nutrition-related topics at their last routine checkup. This question asked if this topic had ever been discussed between participants and their doctor. If participants responded that they had discussed food, nutrition, or diet, they were then asked to report on the duration of the discussion (less than 1 min, 1-4 min, 5-9 min, or 10 min or more), who brought up the topic of nutrition first (doctor, patient, someone else, not sure/don’t remember), and the

participants attitudes towards nutrition discussion on a 5-pt likert scale from strongly agree to strongly disagree:

- I think my doctors should talk about nutrition or diet with all their patients during routine checkups
- I want my doctor to talk about nutrition or diet with me during routine checkups
- I want to be the one to bring up nutrition or diet with my doctors during routine checkups
- I do not want to talk about nutrition or diet with my doctor during routine checkups
- Another healthcare provider, i.e. nurse, nutritionist, dietician, specialist, should be responsible for talking with patients about nutrition or diet

For participants who had discussed food, diet or nutrition at their last routine visit, the survey asked which specific topics relating to nutrition patients had discussed with their doctors at their last routine checkup. This question was only presented to those who had discussed this topic at their last routine visit because of concerns about recall.

The questionnaire then asked participants how they viewed their own health and wellness and prompting them to rate how often they try to increase or decrease consumption of certain food groups. Additionally, this question asked participants about weight maintenance and whether they have intentionally tried to change or maintain their weight in the past year. To assess behavior change intentions, participants were asked if their most recent discussion with their doctor about nutrition or diet impacted their daytoday nutrition or eating habits (0 = not at all, 1 = a little, 2 = some, 3 = a lot).

Participants were then asked whether they had ever sought out nutrition information from any source and ranked their choices of the most reliable resources for such information. Finally, the participants were asked their overall health status, whether they had a personal or family history of cancer, and whether they had received a diagnosis from a list of other medical conditions such as diabetes or a heart condition. Participants reported height and weight used to calculate BMI, categorized into underweight (BMI < 18.5), normal weight (BMI 18.5-24.9), overweight (BMI 25.0 -29.9), obese (BMI 30.0 – 39.9), and very obese (BMI 40 and above). Additionally, participants were asked their gender (male, female, prefer not to answer), age, race, ethnicity, education, employment status, native status, marital status, and income. For a full list of survey items, please contact the authors.

ANALYSES

Descriptive statistics were used to analyze patient characteristics and the frequency of nutrition-related discussions. The relationship between nutrition-related discussions and patient-centered communication was analyzed using a series of ANOVAs. All analyses were conducted in SPSS v. 24.

III. Results

PARTICIPANT CHARACTERISTICS

339 participants reported having seen a doctor for a routine check-up in the previous year and were thus included in the analysis (Table 1). Participants included 159 (46.9%) males and 158 (46.6%) females. Additionally, the sample included 3 (.9%) participants who preferred not to disclose their gender and 19 (5.6%) participants who failed to complete the question regarding gender. The majority of participants for the survey were white (N = 271, 79.9%), born in the United States (N = 308, 90.9%), were either married (N = 133, 39.2%) or single (N = 120, 35.4%), and the vast majority of participants (74%) were employed. The mean age of participants was 37.20 years of age with a standard deviation of 11.77 years. The majority had completed college (N = 148, 43.7%) and most had household incomes of \$20,000 to \$34,999 (N = 68, 20.1%) or \$50,000 to \$74,999 (N = 70, 20.6%).

TABLE 1

Gender	Male	159 (46.9)
	Female	158 (46.6)
	Prefer not to answer	3 (0.9)
	Missing	19 (5.6)
	Age (mean)	37.20 (SD=11.77)
BMI	Not Overweight/Obese (BMI < 25)	168 (49.6)
	Overweight/Obese (BMI >= 25)	152 (44.8)
	Missing	19 (5.6)
Education	Less than 8 years	0 (0)
	8 through 11 years	1 (0.3)
	12 years or completed high school	28 (8.3)
	Post high school training other than college	15 (4.4)
	Some College	101 (29.8)
	College graduate	148 (43.7)
	Postgraduate	27 (8.0)
	Missing	19 (5.6)
Annual Household Income	\$0 to \$9,999	11 (3.2)
	\$10,000 to \$14,999	15 (4.4)
	\$15,000 to \$19,999	13 (3.8)
	\$20,000 to \$34,999	68 (20.1)
	\$35,000 to \$49,999	57 (16.8)

	\$50,000 to \$74,999	70 (20.6)
	\$75,000 to 99,999	48 (14.2)
	\$100,000 to 199,999	31 (9.1)
	\$200,000 or more	3 (0.9)
	Prefer not to answer	4 (1.2)
	Missing	(1.2)
Race/ethnicity	White	271 (79.9)
	Non-white	68 (20.1)
Marital Status	Married	133 (39.2)
	Living as married	30 (8.8)
	Divorced	31 (9.1)
	Widowed	5 (1.5)
	Separated	1 (0.3)
	Single, never married	120 (35.4)
	Missing	19 (5.6)
Born in the United States	Yes	308 (90.9)
	No	12 (3.5)
	Missing	19 (5.6)
Current Occupational Status	Employed	251 (74.0)
	Unemployed	16 (4.7)
	Homemaker	21 (6.2)
	Student	7 (2.1)
	Retired	7 (2.1)
	Disabled	7 (2.1)
	Other	11 (3.2)
	Missing	19 (5.6)

NUTRITION DISCUSSION FREQUENCY

In response to whether a doctor had discussed anything related to food, nutrition, or diet with them at their last routine checkup, 55 (16.2%) participants stated that they had discussed this topic in detail, 153 (45.1%) said they had discussed this topic a little, and 123 (36.3%) had not discussed this topic with their doctor at their last routine checkup. For those that had not discussed this topic at their last routine checkup, 50 (14.7%) replied that they had discussed this topic but not at their last checkup and 73 (21.5%) stated that they had never discussed this topic with their doctor.

Of the sample of participants who had a discussion regarding nutrition with their doctor, 131 participants (38.6%) stated that their doctor had brought up the topic first, 108 (31.9%) participants began the conversation themselves, 3 (.9%) participants had the conversation started by a caregiver, family member, friend or nurse at the appointment with them, and 12 (3.5%) are not sure or do not remember who began the nutrition conversation when it was discussed.

The majority of participants who had discussed nutrition-related topics at some point in the past had a conversation that lasted between 1 to 4 minutes (153 participants or 45.1% of the sample that had discussed nutrition-related topics). 71 participants (20.9%) had had conversations with their doctor lasting a duration of 5 to 9 minutes total. 16 participants (4.7%) had conversations lasting 10 minutes or more on the topics and 14 participants (4.1%) discussed the topics for less than 1 minute.

The 208 participants who had discussed nutrition at their last routine check-up were asked to report which specific topics were discussed. Increasing healthy food choices, including fruits, vegetables, and whole grains, was discussed with 117 participants (56.3%). Decreasing unhealthy food choices such as reducing intake of processed foods, saturated fats, salt, high fructose corn syrup, etc. was discussed with 105 participants (50.5%). Recommended daily intake and portion sizes based on the food pyramid or other dietary guidelines were discussed with 36 participants (17.3%). 72 participants (34.6%) discussed nutrition or diets relating to specific health conditions, such as a low-salt diet for high blood pressure. 26 participants (12.5%) discussed diets for allergies or gluten-free diets with their doctor. 59 participants (28.4%) discussed weight-loss or weight control. 19 participants (9.1%) discussed specific diet plans such as Weight Watchers or low-carb diets. 16 participants (7.7%) stated that they had discussed other topics relating to nutrition with their doctor. Such open-ended responses included discussions of foods that might cause constipation, foods related to a low protein intake, good proteins for vegetarians, and low fat or fat free options during pregnancy.

NUTRITION DISCUSSION AND PATIENT-CENTERED COMMUNICATION

The relationship between discussion of nutrition and participant reports of patient-centered communication was significant for almost all of the patient-centered communication items. Participants who had discussed nutrition in detail with their doctors (mean = 4.33) or a little (mean = 4.30) were more likely to report that their doctor gave them time to ask all health-related questions ($F = 2.93$, $p = .093$) than those who had not discussed nutrition with their doctor (mean = 4.08). Participants were more likely to report that their doctor had given the attention they needed to their feelings and emotions ($F = 4.45$, $p = .012$) if they had discussed nutrition in detail (mean = 4.22), or discussed nutrition a little (mean = 3.93), compared to those who had not had such discussions with their doctor (mean = 3.70).

Participants who had discussed nutrition in detail (mean = 4.44) were more likely to report their doctor involved them in decisions about their health care as much as they wanted than those who had discussed a little (mean = 4.14) or not at all (mean = 3.96, $F = 4.80$, $p = .009$). Participants who reported more nutrition discussions also felt their doctor made sure they understood the things they needed to do to take care of their health ($F = 11.78$, $p < .001$), with those who had discussed nutrition in detail (mean = 4.65) and those who had discussed it a little (mean = 4.33) higher than and those who had not discussed at all (mean = 4.00). Reporting that their doctor explained things in a way they could understand

followed the same significant pattern ($F=3.70$, $p=.026$) with patients who had discussed nutrition in detail (mean = 4.64) or a little (mean = 4.46) reporting higher patient-centered communication than those who had not discussed nutrition (mean = 4.00).

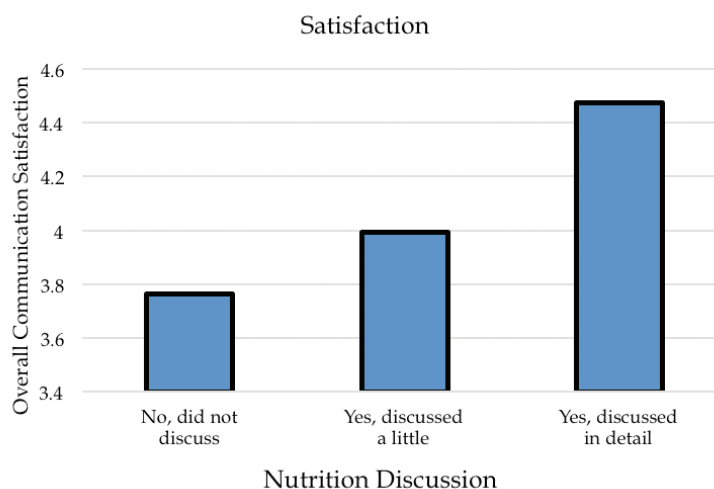
Finally, participants who had more nutrition-related discussion reported they could rely on their doctor to take care of their health care needs ($F=13.02$ and $p<.001$).

The relationships between nutrition communication and patient-centered communication did not significantly differ when comparing male vs. female patients or male vs. female doctors.

NUTRITION DISCUSSION AND PATIENT SATISFACTION

In addition to more patient-centered communication, patients who reported discussions with their doctor about nutrition at their last routine checkup also reported higher levels of overall satisfaction with the communication with their doctor at that visit ($F = 11.15$, $p < .001$, Figure 1). This also did not vary by patient or provider gender.

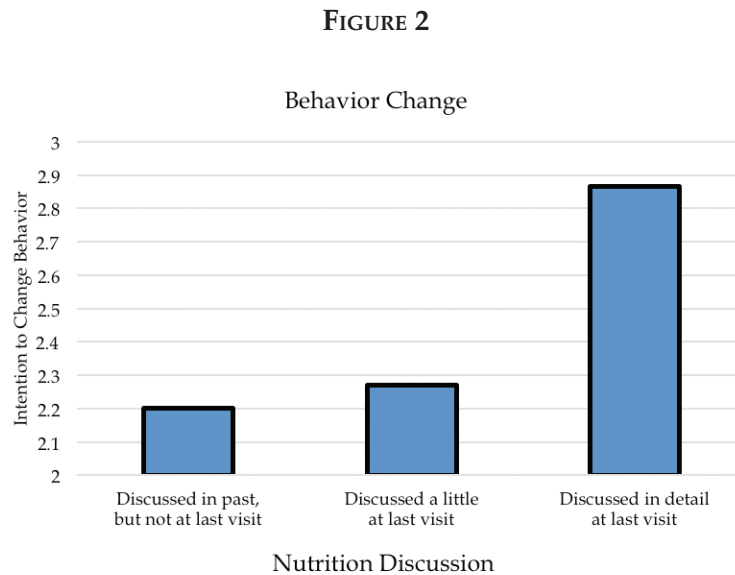
FIGURE 1



Relationship between nutrition-related discussion and overall satisfaction with doctor's communication during the last routine checkup.

NUTRITION DISCUSSION AND BEHAVIORAL CHANGE INTENTIONS

There was a significant relationship between the level of nutrition-related discussion a participant had with their doctor and the degree to which that participant reported attempting to change his or her day-to-day nutrition or eating habits over the previous year ($F = 11.78$, $p < .001$, Figure 2). Participants who had discussed nutrition in detail at the last visit were more likely to report behavior change intentions, followed by those who had discussed nutrition a little, or in the past.



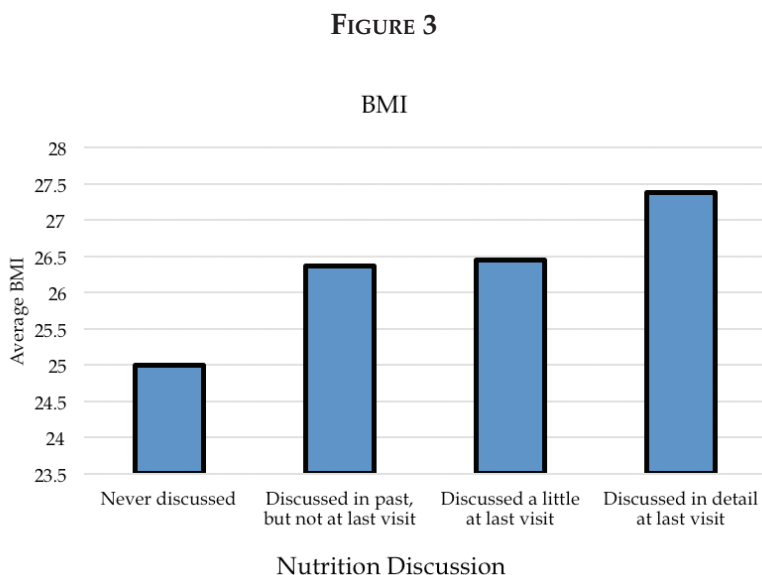
Relationship between nutrition-related discussion and intention to change behavior.

BMI AS A PREDICTOR OF NUTRITION DISCUSSION

BMI was significantly related to discussions about nutrition during the last routine checkup ($F=3.74$ and $p=.011$, Figure 3). Participants with higher BMIs were more likely to have discussed nutrition in detail at their last visit than those with lower BMIs. Those who reported never discussing nutrition with a doctor at a routine checkup had the lowest BMIs.

IV. Discussion

This study explored the frequency, quality and characteristics, and relationship to patient satisfaction and perceptions of patient-centered communication of nutrition-related discussions during routine check-ups between patients and primary care doctors. Nutrition-related discussions occur in detail less than 20% of the time and the majority last for less than 5 minutes.



Relationship between nutrition-related discussion and participant body mass index (BMI)

However, this research also demonstrates that when these discussions do occur, particularly when they occur in detail during a routine visit, they had benefits for the patient and the patient-provider communication. Nutrition-related discussions were associated with increased patient perceptions of the quality of the communication both overall and for all 6 functions of patient-centered communication. In addition, patients who reported having these discussions in more detail also reported more intention to change nutrition-related behaviors as a result of having the conversation with their doctor. In 2003, the US Preventive Service Task Force concluded that there is insufficient evidence to recommend nutrition-related counseling in primary care settings (Newland, 2003). It is important to continue studies like this to build the evidence-base for the efficacy of nutrition-related discussions in a routine care context.

This study has some key limitations. This was an online, convenience sample and may not be generalizable to the general population. In addition, although there was a strong relationship between nutrition-related discussion and the communication, satisfaction, and behavioral intention, a causal relationship cannot be determined from this cross-sectional design. Patients who have better relationships with their doctors and perceive better communication may be more likely to report having these discussions.

We are also relying on participant recall and perceptions of these discussions which are inherently subjective. What is regarded as discussing a topic in detail for one patient might not be the same for another patient. Future research should combine patient reports with objective coding of nutrition-related discussion frequency and quality. Studying the effect of interventions to increase communication about nutrition on patient perceptions can also support a causal relationship. Despite these limitations and the lack of causal conclusions, the positive relationships indicate a potential benefit of nutrition-related discussions between patients and doctors at routine checkups.

Interestingly, gender did not change the relationship between nutrition-related discussions and perceptions of patient-provider communication. Although female doctors often engage in a more patient-centered communication style (Hall, 1984), the question of gender differences in communication about nutrition or the patient-centered communication between providers and overweight or obese patients had not been extensively studied.

The present study also suggests that these nutrition-related discussions are more likely to occur and in more detail with patients who have higher BMIs. This is in line with previous research in an adolescent population (Klein, et al., 2006). The relationship between these discussions and quality of communication in patients with varying BMIs is important to understand given a small amount of research has demonstrated a negative relationship between patient-centered communication and obesity. The communication between providers and obese patients in these studies shows lower levels of rapport and includes less patient education about key health issues. Communication with obese patients may be less patient-centered because healthcare providers hold similar weight-based stigma as the general population. A vignette study experimentally manipulating the patient as obese or not demonstrated that providers expected the visit with the obese patient to be more of a waste of their time and that they would be less positive towards the patient and that the patient would annoy them more.

This study suggests that it is not just the presence or absence of nutrition-related communication but how much detail the patient perceives. The study suggests a positive association between how in-depth the nutrition-related discussion was between the doctor and patient and the related level of satisfaction a patient has with the doctor's communication overall. This study has important implications for routine clinical care and medical education. Providers should consider not only the quantity but quality of their discussions about nutrition. It may also lessen provider concerns that having these discussions will harm their relationship with patients or lead to negative patient perceptions of their care.

References

- Centers for Disease Control and Prevention** (2015). "Adult Obesity Facts." *Centers for Disease Control and Prevention*, 21 Sept. 2015 (<http://www.cdc.gov/obesity/data/adult.html>)
- Centers for Disease Control and Prevention** (2016). "Obesity and Overweight." *Centers for Disease Control and Prevention*, 27 Apr. 2016. (<http://www.cdc.gov/nchs/fastats/obesity-overweight.htm>)
- Epstein R.M., and R.L. Street** (2007). "Patient communication in cancer care: Promoting healing and reducing suffering." *NIH Publication No. 07-6225*. Bethesda, MD: National Cancer Institute.
- Hall, J. A.** (1984). *Nonverbal Sex Differences: Communication Accuracy and Expressive Style*. Baltimore: Johns Hopkins University Press.
- Institute of Medicine** (2001). "Envisioning the National Healthcare Quality Report."
- Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J., & Riess, H.** (2014). "The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials." *PloS one*, 9(4), e94207.
- Klein, J. D., Postle, C. K., Kreipe, R. E., Smith, S. M., McIntosh, S., Spada, J., & Ossip-Klein, D.** (2006). "Do physicians discuss needed diet and nutrition health topics with adolescents?" *Journal of Adolescent Health*, 38(5), 608-e1.
- Kolasa, Kathryn M., and Katherine Rickett** (2010). "Barriers to Providing Nutrition Counseling Cited by Physicians." *Nutrition in Clinical Practice* 25.5 (2010): pp. 502-09.
- Kraschnewski, J. L., Sciamanna, C. N., Pollak, K. I., Stuckey, H. L., & Sherwood, N. E.** (2013). "The epidemiology of weight counseling for adults in the United States: a case of positive deviance." *International Journal of Obesity*, 37(5), pp. 751-753.
- Newland, Jamesetta A.** (2003). "U.S. Preventive Services Task Force: Behavioral Counseling in Primary Care to Promote a Healthy Diet: Recommendations and Rationale". *The American Journal of Nursing* 103.8. pp. 81-95.
- Stewart, M. A.** (1995). "Effective physician-patient communication and health outcomes: a review." *Canadian Medical Association Journal*, 152(9), pp. 1423-1433

ARTICLES

Editor's Introduction

Aaron L. Jackson

Doing Well and Doing Good: Performance of ESG Integration Approaches

George Iliopoulos and Otgontsetseg Erhemjams

Southwest Airlines and the Impact of Low-Cost Carriers on Airline Ticket Prices

Sidney Field

The Impact of Appearance in Evaluating Political Candidates

Julia T. Paradis

The Short-Term Effect of Monetary Policy on Financial Markets

Brian Levine

Patient-Provider Discussion about Nutrition during Routine Visits: Frequency, Quality, and Outcomes

Kerriann Fitzgerald and Danielle Blanch-Hartigan

Bentley University Honors Program

Waltham, MA, USA

fusio@bentley.edu



BENTLEY
UNIVERSITY