ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 7 2016

Aims and Scope of the *Academy of Economics and Finance Journal*:

The *Academy of Economics and Finance Journal (AEFJ)* is the refereed publication of selected research presented at the annual meetings of the Academy of Economics and Finance. All economics and finance research presented at the meetings is automatically eligible for submission to and consideration by the journal. The journal and the AEF deal with the intersections between economics and finance, but special topics unique to each field are likewise suitable for presentation and later publication. The *AEFJ* is published annually. The submission of work to the journal will imply that it contains original work, unpublished elsewhere, that is not under review, nor a later candidate for review, at another journal. The submitted paper will become the property of the journal once accepted, and not until the paper has been rejected by the journal will it be available to the authors for submission to another journal.

Volume 6:

Volume 7 of the *AEFJ* is sponsored by the Academy of Economics and Finance and the Department of Economics and Finance at the University of North Carolina Wilmington. The papers in this volume were presented at the Fifty-third Annual Meeting of the AEF in Pensacola Beach, FL. from February 11-13, 2016. The program was arranged by Dr. Mary Kassis, University of West Georgia.

ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 7

Table of Contents

2016

Peer-to-Peer Equity Investments in Germany. A Note on Successful Company Characteristics

Parthepan Balasubramaniam, Florian Kiesel, and Dirk Schiereck, Technische Universitaet Darmstadt, Germany

Abstract

Peer-to-peer investing is often considered to be a mass alternative to business angel investing. This paper introduces peer-to-peer investing as a new financing instrument for start-ups in the traditionally bank-based financial system of Germany and identifies characteristics that influence the funding success of peer-to-peer investing projects. Based on a comprehensive picture of the peer-to-peer investing platforms in Germany, we analyze 126 funded projects from the three largest German platforms Seedmatch, Companisto and Innovestment. Our results show that larger projects are more successful in funding while the smallest platform is less promising.

Introduction

Moenninghoff and Wieandt (2013) illustrate that the peer-to-peer equity investing markets in Europe are less developed than the peer-to-peer lending markets despite the emergence of the first investing platforms in 2006. Overall, peer-to-peer equity investing is still a hardly used financing source and by far smaller than the according hype in the media might expect. The market volume reached only a low-double digit amount early this decade and appears tiny in absolute terms compared to the overall investment volume managed globally by private equity and venture capital funds. But the compound monthly growth rate of 10% since the beginning of 2010 indicates the rapid growth of this innovative form of financial intermediation over the last years (Moenninghoff and Wieandt, 2013, p. 6-7).

In Germany, not only the IPO and equity capital markets are comparably small compared to the overall economic strength of its economy, but also the peer-to-peer equity investing markets. One often claimed reason for this weakness lies in the less developed German venture capital (VC) industry which might restrict funding sources for young growth companies. With the upcoming of electronic peer-to-peer financing platforms, there are new alternative funding sources which address the VC gap. However, if a leading economy is confronted with a less developed VC community, it is unclear how this institutional background allows peer-to-peer investing platforms to generate sufficient capital supply for funding purposes of young ventures. Here our analyses start.

We take a focus on peer-to-peer investing platforms in Germany which were established since 2010. The market is still hardly analyzed by empirical research on firm characteristics that can explain the success of funding campaigns on these platforms. Our results allow a deeper understanding of the effectiveness of peer-to-peer investing as an alternative financing source for start-ups, and to analyze the success factors of German peer-to-peer investment projects.

The rest of the paper is organized as follows: The following chapter provides a more comprehensive description of peer-to-peer investing, in particular the German peer-to-peer investing market. We then analyze successful peer-to-peer investing projects empirically, examining the dependence of certain attributes. Finally, we briefly summarize our main results and offer our conclusion.

The German Peer-to-Peer Investing Platforms and Instruments

In Germany, peer-to-peer investing and trading platforms are highly regulated in the case of pure equity financing. Platforms, which allow equity financing via stock offerings, need a bank license for their businesses. Consequently, in an attempt to avoid strict regulation, most platforms in Germany decided to offer only hybrid financing instruments which are equity-like but not that strictly regulated. As a result, the German peer-to-peer investing market is dominated by platforms which claim to place equity but really bring together sponsors and founders/entrepreneurs for hardly fungible hybrid assets. These assets include a variety of configuration options. They all have in common being subordinated to traditional debt capital and primary to pure equity in terms of repayment. The existing flexibility in these contracts is the main advantage of these security designs which may also explain that in practice the peer-to-peer investing in Germany exists almost exclusively in hybrid forms of financing, where these are limited to silent partnerships, participation rights and subordinated shareholder loans, and only in very few transactions as stock equity investments. Table 1 gives an overview of the four forms of financing and security designs.

Table 1: Characteristics of the four forms of per-to-peer investing financing

	Equity Investment	Silent Partnership	Participation Right	Subordinated Shareholder Loan
Type	Equity	Mezzanine	Mezzanine	Mezzanine
Liability	No	No	No	No
Profit sharing	Yes	Yes	Yes	Yes
Loss sharing	Yes	Yes	Yes	No
Share in the assets	Yes	Yes (atypical)	No	No
Share in the liquidation proceeds	Yes	Yes	Optional	Yes
Voting rights	Yes	Optional (atypical)	No	No
Control rights	Yes	Yes	No	Yes
Prospectus requirement	Yes	Yes (above 100,000 EUR)	Yes	No

The peer-to-peer investing market is increasingly growing in Germany. It started from a very low base in 2011 at a volume of 450,000 EUR, with enormous growth in 2013 to a financing volume of 15 million EUR, which is an increase of around 250 per cent over 2012. The first half of 2014 amounted to 8.3 million EUR. Table 2 provides an overview of all German peer-to-peer investments which operate from Germany. Based on these figures, it can be assumed that the platforms with the highest funding volumes provide the best infrastructure and represent a benchmark for the industry. In particular, Seedmatch, Companisto and Innovestment, meet these selection criteria. Therefore, our further analysis focuses on the most successful platforms in Germany. Seedmatch went online on October 14, 2009, and presented in August 2011 their first financed companies. It had 85 successful financing projects and 25.7 million EUR in financing volume by the end of 2015. This platform only supports mezzanine capital in the form of a subordinated shareholder loan. All financing investments before November 29, 2012, were originally financed by another form of mezzanine capital – the typical silent partnership.

Table 2: Characteristics of the most active German peer-to-peer investing platforms (December 31, 2015)

Platforms	Start	Completed funding	Completed funding volumes	Financing types	Minimum investment	Holding period
Seedmatch	2011	85	25.7 mn EUR	Subordinated shareholder loans	250 EUR	5 years
Innovestment	2011	30	2.6 mn EUR	Silent partnership	500 EUR	No specific
Companisto	2012	55	26.4 mn EUR	Subordinated shareholder loans	5 EUR	8 years
Deutsche Mikroinvest	2012	50	18.2 mn EUR	Variable	250 EUR	Variable
Fundsters	2012	10	1 mn EUR	Indirect silent partnership	1 EUR	5 years
Bankless24	2012	8	0.6 mn EUR	Participation rights	100 EUR	5 to 7 years
United Equity	2012	4	0.1 mn EUR	Participation rights and silent partnership	100 EUR	5 to 10 years
Bergfürst	2011	2	3.8 mn EUR	Variable	10 EUR	No specific
Direct Startups	2012	Not stated	Not stated	Not stated	2.50 – 50 EUR	Variable

Companisto had 55 successful fundings and 26.4 million EUR funding volume until the end of 2015. Since the beginning of 2013, Companisto only supports mezzanine capital in the form of a subordinated shareholder loan. All financing before the year 2013 was originally financed by another form of mezzanine capital – the atypical silent partnership. Innovestment had 30 successful fundings and 2.6 million EUR financing volume by the end of 2015. This platform only supported mezzanine capital in the form of an atypical silent partnership until second half of 2015. The new participation model is in the form of a subordinated shareholder loan in a special purpose vehicle, which then invests the whole amount as an equity investment into

the start-up. For the analysis, only the investment types until end of 2013 were considered. Innovestment started almost at the same time as Seedmatch and first established itself as a number two in the crowdfunding market, but without growth as that of the market leader Seedmatch. Innovestment obviously has not succeeded in inspiring start-ups and investors in the same way as Seedmatch and Companisto. A particular disadvantage is sticking to the 100,000 EUR limit in relation to the maximum funding volume for the start-ups, and the fact that Companisto for a long time kept a minimum investment of 1,000 EUR for investors. Meanwhile, Innovestment has lowered the minimum amount for some projects to 500 EUR. Table 3 provides an overview of the characteristics of the three peer-to-peer investing platforms.

Table 3: Characteristics of the investment platforms Seedmatch, Companisto and Innovestment

Company	**Seedmatch**	**Companisto**	**Innovestment**
Start	2011	2012	2011
Financing objects	B2C, Cleantech, Social Business	No limitations	No limitations
Investment types	Subordinated shareholder loans (silent partnership until 2013)	Subordinated shareholder loans (silent partnership until 2013)	Atypical silent partnership
Funding limit	Unlimited	Unlimited	100,00 EUR
Minimum amount	250 EUR	5 EUR	500 EUR
Minimum holding period	Normally 5 years	Normally 8 years	None
Conditions for investors	Share on profit +increase in enterprise value	Share on profit +increase in enterprise value	Share on profit +increase in enterprise value
Fees for start-ups	5 to 10 % agency fee	10 % agency fee	1% Emission fee + 5 % agency fee

Empirical Analysis

We collected all peer-to-peer investing projects at Seedmatch, Innovestment, and Companisto starting with the first financed projects in 2011 up to the most current finalized projects in 2014. Peer-to-peer investing projects still in the fundraising process were removed due to the limited comparability of completed and uncompleted projects. Starting with 144 peer-to-peer investing projects initially, 126 remain in the sample. Table 4 presents the peer-to-peer investing projects, the respective platform and the key statistics of the project. The average enterprise value is about 2.25 million EUR and the average funding threshold is about 240,000 EUR, which is only around 10% of the enterprise value. The mean number of funding investors for each project is approximately 324, the mean funding for each investor is therefore around 744 Euro per investment project. The minimum holding period for an investment is 5.7 years in average, if a holding period is required.

In order to analyze the determinants of successful peer-to-peer investing we focus on variables from prior studies (e.g. Agrawal et al., 2015; Hornuf and Schwienbacher, 2014) which engage with similar analyses and mostly use linear regression models in the following form:

$$v_i = \alpha_i + \beta_1 \cdot x_{1i} + \cdots + \beta_m \cdot x_{mi} + u_i \,, \qquad (1)$$

where i = 1, ..., N are the peer-to-peer investing projects, v_i is the natural logarithm of the funds raised in the first model and funds raised over funding limit, in the second model. These two variables define our measure of funding success. α_i is the company-specific regression constant, x_{ji} (j = 1, ..., m) are the explanatory characteristics and u_i denotes the corresponding error terms. Standard errors are corrected for heteroskedasticity and as a robustness of our results we include firm and year fixed effects in our model. Equation (1) specifies without further restriction an ordinary least square (OLS) regression model.

Table 4: Descriptive statistics of the three platforms

Year	Platform	Investments	Sum Funding	Mean Firm Size	Mean No. of Investors
2011	Total	5	449,250	969,740	108.00
	- *Innovestment*	1	100,000	1,403,200	25.00
	- *Seedmatch*	4	349,250	861,375	128.75
2012	Total	42	3,802,444	1,332,558	168.90
	- *Companisto*	7	630,525	1,122,857	439.29
	- *Innovestment*	13	971,919	796,622	28.77
	- *Seedmatch*	22	2,200,000	1,472,364	165.68
2013	Total	51	11,811,470	2,127,983	319,40
	- *Companisto*	17	3,435,325	1,387,647	600.59
	- *Innovestment*	11	1,053,895	N/A	23.18
	- *Seedmatch*	23	7,322,250	2,868,318	334.43
2014	Total	28	11,294,368	3,181,720	536.36
	- *Companisto*	7	2,638,118	1,707,143	1048.43
	- *Innovestment*	3	290,000	N/A	27.00
	- *Seedmatch*	18	8,366,250	3,755,167	422.11
Total		126	30,357,532	2,251,241	323.88

As a measure for the peer-to-peer funding success, two variables will be analyzed independently. The first variable *TOTAL_INVESTMENT* is defined as the natural logarithm of the total funds raised in EUR. However, the upper limit on which a peer-to-peer investing campaign is stopped prevents free market forces from developing a price and this could influence the analysis. Therefore, we additionally analyze the variable *SUCCESS_RATIO*, which is the ratio of funds raised over the funding limit. Other variables, such as the time it takes to reach the funding threshold or limit, might also be a good measurement. However, this data is not publicly available on the crowdinvesting plat-form, and therefore not included in this analysis.

The number of founders is relevant according to findings of Evers (2012) due to the fact that the success rate of start-up peer-to-peer investing is higher when a team initiates a project rather than an individual. The variable *NO_FOUNDERS* is an integer and was retrieved from the description on the company's peer-to-peer investing platform. The enterprise value was chosen as a parameter, because it determines the company's capital requirements and the share that an investor gets in a company. The variable *ENTER_VALUE* is the natural logarithm of the enterprise value in EUR that was announced on the peer-to-peer investing platform prior to the start of the fundraising process.

The stage in the corporate lifecycle can also be another important part of an investor's decision, whether to invest in a company. Our models consist of two dummy variables *STARTUP* and *EXPANSION*. The company specific information was retrieved from the online platforms. We also include a dummy to control for the legal form of the firm that needs new capital. Some legal forms have a higher capital reserve than others; for example, a *Gesellschaft mit beschraenkter Haftung* (GmbH), which is similar to a private limited company, has a higher capital reserve than an entrepreneurial company, namely *Unternehmergesellschaft* (UG). The legal forms represented in our sample are only GmbH and UG. This characteristic is captured by the dummy variable *LF_GMBH* that is defined as 1, if the legal form is a GmbH, otherwise the variable is 0. In terms of experience, the age at the time of the peer-to-peer investing campaign of the start-up might determine the successful funding. The variable *AGE* is retrieved from the peer-to-peer investing platforms and calculated as the deviation between the year of the establishment and the year of the peer-to-peer investing.

In addition, the actual financing instrument may a factor which is important for a successful peer-to-peer investing. Different types of financing instruments have different risks as well as payback rates. The type of financing instrument comes along with different benefits as explained in the prior section. This characteristic is captured in the variable *SILENT* defined as 1, if the investment is financed via a silent partnership, and 0 for the type of subordinated shareholder loans. The financing instrument participation right was not used in our data sample. Finally, we control for platform effects, expecting that the smallest platform is less attractive, because of its limited visibility. *COMPANISTO* and *INNOVESTMENT* each defined as 1, if the firm raising money through peer-to-peer investing used the platform Companisto or Innovestment as their platform.

Table 5 presents the descriptive statistics with the variables used in the subsequent analysis. Due to missing observations, the number of observations dropped from 126 to 101. Two additional facts can be read from the table. First, the average team size is two and the maturity of the companies is 1.6 years in average. In contrast, the holding period with 5.7 years in average seems to be long. It shows that investors are willing to participate in companies with less than two years maturity even though they get a small share in the company and have to hold this investment for at least five years. The top three industries in our sample are information and communication industry (NACE Code J) with 32.3 per cent, wholesale and retail trade (NACE

Code G) with 23.6 per cent and manufacturing industry (NACE Code C) with 22.8 per cent. The common legal form for companies raising money through peer-to-peer investing is a GmbH with 86 per cent of the whole sample.

Table 5: Characteristics of crowdfunded projects in Germany: summary statistics of the control variables

Variable	Mean	SD	Median	Min	Max
NO_FOUNDERS	2.00	0.97	2.00	1.00	6.00
ENTER_VALUE	14.32	0.65	14.32	13.12	16.34
STARTUP	0.70	0.46	1.00	0.00	1.00
EXPANSION	0.06	0.23	0.00	0.00	1.00
LF_GMBH	0.86	0.35	1.00	0.00	1.00
AGE	1.62	1.30	1.00	0.00	6.00
COMPANISTO	0.24	0.43	0.00	0.00	1.00
INNOVESTMENT	0.22	0.42	0.00	0.00	1.00
SILENT	0.48	0.50	0.00	0.00	1.00

Before further analyzing those variables with a regression model, a simple correlation analysis is performed. In Table 6, the correlation between each independent variable can be seen. Most of the independent variables are not correlated with each other. We only find some correlations between variables. As expected, the firm size is positively correlated with the expansion dummy, the age of the company, and the legal form of the company. Larger firms mostly have the legal form of GmbH, are longer in the market, and are in the expansion phase. In addition, we find that the variable of *COMPANISTO* has a negative correlation to the firm size, indicating that larger firms do not use this platform in particular. In order to avoid multicollinearity in our OLS regression analysis, we analyze the variance inflation factors (VIF) to quantify the severity of multicollinearity between each variable in the analysis.

Table 6: Coefficient matrix

	Variable	1	2	3	4	5	6	7	8	9
1	FOUNDERS	1.000								
2	VALUE	-0.194*	1.000							
3	START_UP	0.118	0.116	1.000						
4	EXPANSION	-0.156	0.217**	-0.312***	1.000					
5	LEGAL_FORM	-0.043	0.255**	0.241**	0.081	1.000				
6	AGE	-0.153	0.457***	0.107	0.350***	0.273***	1.000			
7	COMPANISTO	0.390***	-0.268***	0.185*	-0.021	0.010	-0.090	1.000		
8	INNOVESTMENT	0.006	-0.185*	0.021	0.219**	0.081	-0.012	-0.132	1.000	
9	SILENT	0.002	-0.400***	-0.315***	-0.049	-0.230**	-0.281***	-0.179*	0.267***	1.000

This table provides the correlation coefficient matrix of main independent variables. The sample includes 126 crowdinvesting projects from the three peer-to-peer investing platforms Innovestment, Seedmatch, and Companisto. ***, **, * denotes statistical significance at the 1%, 5%, and 10% level, respectively.

Table 7 shows the results for the total funds raised. The firm value has a highly significant impact on the total funds raised. That is not surprising, since larger companies require higher funding. All investments before the beginning of 2013 were originally restricted to a maximum of 100,000 EUR by the previously practiced atypical silent partnership. The regression analysis shows that this type of partnership also results in lower funds. Contrary to expectations, the number of founders, the stage in the corporate lifecycle, legal form and the age of the company do not appear to determine the sum of total funds raised, as their coefficients generally lack significance. In addition, we do not find any significant differences for the platforms. Adding year and industry fixed effects improves the overall explanatory power of the model minimally.

Table 7: Results of the regression analysis for total funds raised

Sample		**Model 1**			**Model 2**	
	Coefficient	**t-Value**	**VIF**	**Coefficient**	**t-Value**	**VIF**
β_0	2.807	1.554	-	2.702	1.483	-
FOUNDERS	-0.033	-0.478	1.257	-0.016	-0.195	1.546
VALUE	0.672***	5.295	1.779	0.668***	5.361	2.099
START_UP	-0.059	-0.430	1.447	-0.093	-0.727	1.638
EXPANSION	0.142	0.474	1.560	-0.017	-0.052	1.732
LEGAL_FORM	-0.070	-0.479	1.193	-0.154	-1.035	1.338
AGE	-0.017	-0.326	1.477	-0.029	-0.520	1.604
COMPANISTO	0.027	0.156	1.558	-0.010	-0.028	1.720
INNOVESTMENT	0.138	0.797	1.317	0.108	0.483	1.541
SILENT	-0.600***	-5.400	1.544	-0.655***	-3.592	2.524
Industry Effects	NO			YES		
Year Effects	NO			YES		
n	101			101		
Adj. R^2	51.47%			51.95%		
F-value	20.99***			12.65***		

This table shows the results of the OLS regression for the 101 investing projects on Seedmatch, Innovestment, and Companisto. The dependent variable is the natural logarithm of the total funds raised in EUR. Standard errors are corrected for heteroskedasticity and associated t-statistics are given. ***, **, * denotes statistical significance at the 1%, 5%, and 10% level, respectively.

A peer-to-peer investing project is only assessed as successfully funded if at least the predefined threshold set by the firm, which wants to raise money, has been reached. Furthermore, the greater the ratio of funds raised over the funding limit, the more successful is the funding itself. Table 8 provides the regression analysis with the success ratio of investments as the dependent variable.

We again find that the company size influences, as the coefficient for *VALUE* ceases to be significant. Larger firms are able to raise funds over the announced funding limit. This may indicate that larger firms have a smaller probability to default and investors rely more heavily on these firms. This is nevertheless somewhat surprising as even larger start-ups still do not have an established business model, and therefore the firm value is difficult to predict precisely. Moreover, the firm value announced on the platform is a subjective estimation of the firm owners and therefore not free of potential biases. In addition, we find that the success ratio is smaller for firms using the platform of Innovestment. Innovestment is the smallest of the three platforms in our sample. 28 of the 126 funding projects in our investigation are raised with Innovestment with an average funding of 86,279 EUR. The mean funding of Companisto and Seedmatch combined is 254,507 EUR. This result suggests that larger platforms increase the probability of a successful funding. For a successful funding, the financing instrument matters as well. The variable of silent partnership is highly significant and positive, indicating that the successful funding increases for projects offering a silent partnership. This may also be explained by the limitations of the maximum funding to a limit of 100,000 EUR. This supports our prior results of the total funds raised. In addition, industry and year fixed effects slightly improve the overall explanatory power of the model.

Overall, our results indicate that the company size has a strong positive influence on the funds raised and the success of the investing projects. In addition, we find that the platform has at least a weak significant influence on the success of funding. Platforms with larger quantity of large funding projects determine the success of funding in a similar manner than the determinants of the firms raising capital.

Table 8: Results of the regression analysis for the success ratio of investment

Sample	**Model 1**			**Model 2**		
	Coefficient	**t-Value**	**VIF**	**Coefficient**	**t-Value**	**VIF**
β_0	0.023	0.051	-	-0.275	-0.561	-
FOUNDERS	-0.012	-0.486	1.257	-0.024	-0.917	1.546
VALUE	0.060*	1.873	1.779	0.076**	2.262	2.099
START_UP	0.025	0.551	1.447	-0.001	-0.031	1.638
EXPANSION	0.076	1.270	1.560	0.067	0.724	1.732
LEGAL_FORM	0.001	0.018	1.193	-0.030	-0.554	1.338
AGE	-0.008	-0.494	1.477	-0.015	-0.919	1.604
COMPANISTO	-0.028	-0.524	1.558	-0.028	-0.507	1.720
INNOVESTMENT	-0.093**	0.042	1.317	-0.038	-0.489	1.541
SILENT	0.142***	0.039	1.544	0.170***	3.136	2.524
Industry Effects	NO			YES		
Year Effects	NO			YES		
n	101			101		
Adj. R^2	6.57%			12.92%		
F-value	3.86***			3.29***		

This table shows the results of the OLS regression for the 101 investing projects on Seedmatch, Innovestment, and Companisto. The dependent variable is the ratio of funds raised over the funding limit. Standard errors are corrected for heteroskedasticity and associated t-statistics are given. ***, **, * denotes statistical significance at the 1%, 5%, and 10% level, respectively.

Summary and Conclusion

Peer-to-peer investing is going to become a popular method of financing start-ups and new business ideas. We introduce peer-to-peer investing as a new financing instrument for start-ups in the traditionally bank-based financial system of Germany and identify characteristics which influence the funding success of peer-to-peer investing projects. Based on a comprehensive picture of the peer-to-peer investing platforms in Germany, we analyze 126 funded projects from the three largest German platforms Seedmatch, Companisto and Innovestment. Our results show that larger projects are more successful in funding while the smallest platform Innovestment offers a less promising service.

References

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2015. "Crowdfunding: Geography, social networks, and the timing of investment decisions." *Journal of Economics & Management Strategy*, *24*(2): 253-274.

Evers, Mart, Carlos Lourenço, and Paul Beije. 2012. "Main drivers of crowdfunding success: a conceptual framework and empirical analysis." Erasmus Universiteit.

Hornuf, Lars, and Armin Schwienbacher . 2014. "The Emergence of Crowdinvesting in Europe." *Available at SSRN 2481994*.

Moenninghoff, Sebastian C., and Axel Wieandt. 2013. "The future of peer-to-peer finance." *Zeitschrift für betriebswirtschaftliche Forschung, 65*: 466-487.

Currency ETF Tracking Error

Robert B. Burney, Coastal Carolina University

Abstract

The emergence of currency exchange traded funds (ETFs) has provided an alternative vehicle for both speculation and hedging in the currency markets. Because currency ETFs trade like equities and have no relevant expiration date, they represent an intriguing alternative for managing certain types of foreign exchange risk. However, how well a particular currency ETF tracks the associated currency is important

One of the most common metrics for ETF performance is tracking error. This metric, along with tracking difference has been used extensively in many settings to describe ETF performance. However, to date a relatively limited amount of research has addressed these same issues for currency ETF specifically. This paper examines the tracking error and tracking difference to assess the performance of several widely traded currency ETFs.

Currency ETFs

In recent years numerous currency ETFs (Currency ETFs) have been introduced. These include ETFs which cover most of the major currencies and an increasing number of second tier currencies. Variations include both long and short position ETFs, and more recently, double and triple long and short varieties. Table 1 presents the sample of currency ETFs addressed in this study.

Currency ETFs have at least three different objectives. These are 1) to track the price of the underlying currency, 2) to track the daily price change (return) of the underlying currency, or 3) to track the performance of the underlying currency and the associated foreign money market. The basic currency ETFs involve foreign currency denominated bank accounts. The inverse or leveraged Currency ETFs also use derivatives to attempt to meet their objectives (Lachini and Michel, 2011)

To varying degrees, a currency ETF substitutes for a long or short position in the target currency. Because the currency ETFs are traded on a share basis, an investor can take any conceivable position depending on the number of shares purchased. While most applications of both speculation and hedging in foreign currencies can be accomplished with outright currency trades and traditional derivatives, the ease of trading currency ETFs is attractive to less experience investors and those who are only occasionally concerned with foreign exchange risk. Also, there is no set denomination per share of such Currency ETFs. Some Currency ETFs are quoted in multiples of the underlying currency, while others are based on an arbitrary notional principal. Recent per share values ranged from $15.87 to $132.72 for the Currency ETFs presented in Table 1.

The minimum number of shares which an individual investor could trade depends on the individual brokerage housing the account. This conceivably could be as few as one share, and should not be confused with the inter-institutional "creation units" of much larger magnitude (Abner, 2010). Commissions on currency ETF trades are also subject to wide variation, with typical commissions at discount brokerages below $10 per trade. Some brokerages also offer commission-free trading on select ETFs.

The characteristics of currency ETFs makes them potentially attractive for the management of foreign currency risk for market participants who face only very small or occasional foreign currency exposures and those who face extremely long-term foreign currency exposures. For such market participants, existing liquid market derivative securities simply do not match the transaction scale or maturity. Table 2 summarizes this situation for the dominant U.S. market derivatives for the euro. In this case, the smallest notional principal amount involves 10,000 euros. A smaller size transaction would force the market participant to, in effect; take on a residual exposure of opposite nature to the initial exposure.

While services for small scale market participants do exist among the retail foreign exchange dealers, these arrangements frequently have a high cost structure with account details which may create more difficulties for the potential hedger (full margin calls, etc.). Also, the smaller FX dealer based derivative contracts or minor electronic exchanges suffer from illiquidity which may negatively impact pricing.

With respect to maturity, the available exchange traded contracts tend to have relatively short maturities when compared to certain long-term foreign exchange exposures. Some OTC derivatives do have longer maturities, but these would be expected to have low liquidity. Because the currency ETFs have no effective maturity date, they could conceivable be used for longer-term hedging.

Table 1: Currency ETFs in this Study

Symbol	Name*	Objective	Objective Multiple	Multiple of	Priced in
FXY	CURRENCY SHS.JAPE.YEN TST.	long	1	price	USD
FXB	CURRENCYSHARES BRIT.PND. STERLING TST.	long	1	price	USD
FXC	CURRENCYSHARES CDN. DOLLARS TST.	long	1	price	USD
FXCH	CURRENCYSHARES CHINESE RENMINBI TRUST	long	1	price	USD
FXE	CURRENCYSHARES EURO TST.	long	1	price	USD
FXSG	CURRENCYSHARES SING. DOLLAR TST.	long	1	price	USD
FXS	CURRENCYSHARES SWEDISH KRONA TST.	long	1	price	USD
FXF	CURRENCYSHARES SWISS FRANC TST.	long	1	price	USD
FXA	CURRRENCYSHARES AUST. DOLLAR TRUST	long	1	price	USD
EUFX	PROSHARES SHORT EURO	short	-1	1 day return	USD
ULE	PROSHARES ULTRA EURO	long	2	1 day return	USD
YCL	PROSHARES ULTRA YEN	long	2	1 day return	USD
CROC	PROSHARES ULTRASHORT AUST.DOLLAR	short	-2	1 day return	USD
EUO	PROSHARES ULTRASHORT EURO	short	-2	1 day return	USD
YCS	PROSHARES ULTRASHORT YEN	short	-2	1 day return	USD
BZF	WISDOMTREE BRZLN.RL.FD.	long	1	MM+Price**	USD
CYB	WISDOMTREE CHS.YUAN FD.	long	1	MM+Price	USD
ICN	WISDOMTREE INDIAN RUPEE FD.	long	1	MM+Price	USD

* as given in the data source, ** foreign currency money market return

Table 2: Example Derivative Contract Sizes

PHLX:	
Euro Options	10,000 euros
CME:	
Euro Futures*	125,000 euros
E-mini Futures	62,5000 euros
E-micro Futures	12,5000 euros

*CME Options are limited to larger contracts.

Tracking Error and Tracking Difference

Tracking error is generally defined as the annualized standard deviation of the return differences between a fund and its benchmark. In the case of currency ETFs, the benchmark is the price of the underlying currency expressed in the same terms as the ETF itself is priced. That is, an Australian dollar ETF priced in U.S. dollars would have the benchmark of the AUDUSD (U.S. dollar direct) exchange rate.

Tracking difference is defined as the annualized difference between a fund's actual return and its benchmark return. In the case of currency ETFs, this would be the difference between the return on the fund and the return on the underlying currency during the specified time period.

Tracking error essentially describes the variability of the tracking difference. Thus, tracking error can be used develop confidence intervals for predicted tracking error given an assumption of the distribution of the tracking errors. For example, assuming a normal distribution, an ETF with a tracking error of 1% would be expected to exhibit tracking differences of +/- 1% from its average tracking difference 68% of the time (Johnson et al., 2013).

Factors Influencing Tracking Error and Difference

The literature has examined multiple factors which might logically be expected to impact ETF performance as measured by tracking error and tracking difference. These include the total expense ratio (TER) as a proxy for management involvement, trading volume as a proxy for liquidity, fund total assets as a proxy for management resources, and nature of the markets involved (emerging or developed). In addition, the literature has examined whether tracking error is itself related to tracking difference. While other types of ETFs have been examined in multiple studies, currency ETFs are to date largely unexamined.

Data

Data for this study was drawn from the Reuters DataStream data sets. Two years of daily data from January 2014 through January 2016 was selected. This resulted in an overall number of useable observations of 526. The ETF price was the closing price in New York. The exchange rates used were the WM/Reuters rates which are the London 16:00 fixing. Depending on the time of year and associated stage of Daylight Savings Time and British Summer Time, the timing difference between the ETF prices and daily exchange rates is four to five hours.

All of the currency ETFs in the study are traded on the New York markets and are priced in U.S. dollars. The exchange rates reported by WM/Reuters are in the Interbank Market format with all but the euro, British pound, and Australian dollar in U.S. indirect format. To match the currency denomination of the ETFs, the exchange rates in the dataset in U.S. indirect terms were restated to U.S. direct terms.

The WM/Reuters dataset provides bid, offer, and mid-range quotations. For this study the mid-range quotes were used. Throughout the paper any reference to exchange rates and exchange rate returns refers to the mid-range quotes.

Table 3 presents the correlations between the twelve currencies and the prices of the associated currency ETFs. Note that the number of ETFs reported differs among the currencies. In the sample, the euro-dollar and euro-yen currency pairs have the largest number of ETFS. In the table, currencies are described using their international trading symbol quotations while the funds are described using their trading symbols.

The correlations presented in Table 3 are those between the individual fund prices and associated exchange rates over the entire two year period. Of the eighteen funds in the study, all show a very high correlation with the under lying currency except for the Wisdom Tree Chinese yuan and Indian Rupee funds (WTCYB and WTINR). It should be noted that both of these ETFs have a money market component in their stated objectives. However, the Wisdom Tree Brazilian Real fund reports this same general approach, but evidenced a much higher correlation with the underlying currency.

The data is divided into four approximately equal time periods of approximately six months in length. The resulting estimates are then annualized.

Daily returns are calculated using logarithmic returns such that the one day return is $R= \ln(P_t/P_{t-1})$. The annualized tracking error for the period is thus as follows with n=number of observations per year (250).

$$Tracking\ Error = std(R_{ETF} - R_{FX\ Rate})\sqrt{n} \tag{1}$$

Tracking differences are the annualized daily return difference. This construct can be measured in various ways. For this study we use the sum of the logarithm daily returns for the total period return. This return is then annualized by multiplying by the factor (250/N) where N is the total number of observations in the sub period.

$$Tracking\ Difference \sum_{i=1}^{N}(R_{ETF} - R_{FX\ Rate})_i\ (\frac{250}{N}) \tag{2}$$

The data for TER, volume, and total assets is drawn from the Schwab research data. In this case TER is the currently reported value, while volume is the recent 10 day average, and size is the most recently reported value of total assets.

Hypotheses

This study examines four hypotheses concerning tracking error (TE) and tracking difference (TD). Each of these is tested in separate stages of the analysis. The four hypotheses are as follow.

Hypothesis One: Higher Total Expense Ratio (TER) ETFs will have lower TEs and TDs
Hypothesis Two: Larger ETFs will have lower TEs and TDs
Hypothesis Three: Lower volume ETFs will have Higher TEs and TDs
Hypothesis Four: Funds with higher TEs will also have TDs

Hypotheses ne and two are based on the assumption that higher TER and larger fund size should be associated with greater managerial resources. This in turn could be expected to improve fund performance leading to both lower TEs and TDs. Hypothesis Three is based on the assumption that lower trading volumes will be associated with inefficiencies and lower effectiveness in attaining fund objectives. Finally, Hypothesis Four is based on the assumption that poor performance will be evidenced in both the variability and accuracy with which the fund meets its objectives.

Table 3: Currency and CETF Correlations

Currency	*ETFs*		*Currency*	*ETFs*			
AUDUSD	CSFXA	PSCROC	EURUSD	CSFXE	PSEUFX	PSULE	PSEUO
	0.99965*	-0.99303		0.99964	-0.99701	0.99904	-0.99474
	<.0001**	<.0001		<.0001	<.0001	<.0001	<.0001
	523***	523		523	523	523	523
BRLUSD	WTBZF		GBPUSD	CSFXB			
	0.99176			0.99904			
	<.0001			<.0001			
	523			523			
CADUSD	CSFXC		INRUSD	WTICN			
	0.99954			0.48556			
	<.0001			<.0001			
	523			523			
CADUSD	CSFXC		JPYUSD	CSFXY	PSYCS	PSYCL	
	0.99954			0.99957	-0.99695	0.99893	
	<.0001			<.0001	<.0001	<.0001	
	523			523	523	523	
CHFUSD	CSFXF		SEKUSD	CSFXS			
	0.99587			0.99944			
	<.0001			<.0001			
	523			523			
CNYUSD	CSFXCH	WTCYB	SGDUSD	CSFXSG			
	0.94729	0.83406		0.99037			
	<.0001	<.0001		<.0001			
	523	523		520			

*****correlation, **significance, ***observations (pattern repeats throughout table)

Methodology

To test Hypotheses One through Hypothesis Three, a multiple regression will be run between the funds' TEs and TERs, volumes, and sizes. A second regression will be run between the funds' TDs and TERs, volumes, and sizes. The data will be divided into subsets of data each of covering approximately a six month period, with separate regressions being run for each of the subperiods. Estimates will be stated on an annualized basis.

To test Hypothesis Four, a separate set of regressions will be run relating the funds' TEs and TDs. The regression will be run for each of the four time periods. Although the time periods are each of around six months in length, the resulting estimates will be stated on a per annum basis.

Results

Table four presents a summary of the results of the time period specific regressions of TE and TD against TER, volume, and size. These four time periods are the four approximately six month period in the 2015-2016 study period. In none of the

eight regressions is any parameter estimate significant. So, our analysis provides no support for Hypotheses One through Three. That is, no significant relationship is found between TE or TD and Total Expense Ratio, Volume, or Total Assets. None of the estimated coefficients is significant for any of these explanatory variables

Table 4: Regression of TE and TD against TER, Volume and Size

Period 1: TE1									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.01879	R-Square	0.2634
						Dependent Mean	0.04287	Adj R-Sq	0.1055
Intercept	1	0.05683	0.01177	4.83	0.0003	Coeff Var	43.83933		
TER	1	-0.00961	0.01746	-0.55	0.5907				
AVGVOL	1	4.55E-09	3.42E-08	0.13	0.8962				
TOTALASSETS	1	-7.7E-05	5.7E-05	-1.34	0.2				
Period 1: TD1									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.03386	R-Square	0.1176
						Dependent Mean	0.01605	Adj R-Sq	-0.0715
Intercept	1	0.03333	0.02121	1.57	0.1385	Coeff Var	211.0341		
TER	1	-0.01408	0.03146	-0.45	0.6614				
AVGVOL	1	9.36E-09	6.17E-08	0.15	0.8815				
TOTALASSETS	1	-8.8E-05	0.000103	-0.85	0.4083				
Period 2: TE2									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.02738	R-Square	0.0799
						Dependent Mean	0.06395	Adj R-Sq	-0.1172
Intercept	1	0.07566	0.01715	4.41	0.0006	Coeff Var	42.80756		
TER	1	-0.01786	0.02543	-0.7	0.4939				
AVGVOL	1	-3.54E-08	4.99E-08	-0.71	0.4897				
TOTALASSETS	1	2.63E-05	8.3E-05	0.32	0.756				
Period 2: TD2									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.07859	R-Square	0.0436
						Dependent Mean	-0.00315	Adj R-Sq	-0.1613
Intercept	1	-0.039	0.04923	-0.79	0.4415	Coeff Var	-2495.37		
TER	1	0.04966	0.07301	0.68	0.5075				
AVGVOL	1	-2.39E-08	1.43E-07	-0.17	0.8699				
Period 3: TE3									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.03329	R-Square	0.1262
						Dependent Mean	0.07163	Adj R-Sq	-0.061
Intercept	1	0.09542	0.02085	4.58	0.0004	Coeff Var	46.47305		
TER	1	-0.02736	0.03092	-0.88	0.3912				
AVGVOL	1	2.47E-08	6.07E-08	0.41	0.6902				
TOTALASSETS	1	-9.1E-05	0.000101	-0.91	0.3805				
Period 3: TD3									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.05034	R-Square	0.0215
						Dependent Mean	0.00687	Adj R-Sq	-0.1882
Intercept	1	0.00225	0.03153	0.07	0.9442	Coeff Var	732.941		
TER	1	0.01507	0.04676	0.32	0.752				
AVGVOL	1	-8.19E-09	9.17E-08	-0.09	0.9301				
TOTALASSETS	1	-2.9E-05	0.000153	-0.19	0.8				
Period 4: TE4									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.03985	R-Square	0.137
						Dependent Mean	0.07829	Adj R-Sq	-0.048
Intercept	1	0.10564	0.02496	4.23	0.0008	Coeff Var	50.90187		
TER	1	-0.02851	0.03702	-0.77	0.454				
AVGVOL	1	3.83E-08	7.26E-08	0.53	0.606				
TOTALASSETS	1	-0.00013	0.000121	-1.08	0.2986				
Period 4: TD4									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.03914	R-Square	0.013
						Dependent Mean	0.00676	Adj R-Sq	-0.1985
Intercept	1	0.00741	0.02452	0.3	0.7669	Coeff Var	579.028		
TER	1	0.00323	0.03636	0.09	0.9305				
AVGVOL	1	-1.23E-08	7.13E-08	-0.17	0.8652				
TOTALASSETS	1	-9.7E-06	0.000119	-0.08	0.936				

Table Five presents a summary of the results of the time period regressions of TE against TD. In this case we find a significant relationship in period one and in period four. This suggests that poor performance in one criterion is associated with poor performance in the other. So, limited support is found for Hypothesis Four. This also leads us to believe that time period definition may be more critical than originally anticipated.

Table 5: Regression of TE and TD by Time Period

Time Period 1: Model TE1=TD1									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.01574	R-Square	0.4099
						Dependent Mean	0.04287	Adj R-Sq	0.373
Intercept	1	0.03663	0.00415	8.82	<.0001	Coeff Var	36.70304		
TD1	1	0.38892	0.11666	3.33	0.0042				
Time Period 2: Model TE2=TD2									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.02661	R-Square	0.0065
						Dependent Mean	0.06395	Adj R-Sq	-0.0555
Intercept	1	0.06404	0.00628	10.2	<.0001	Coeff Var	41.60898		
TD2	1	0.02874	0.08849	0.32	0.7496				
Time Period 3: Model TE3=TD3									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.03119	R-Square	0.1234
						Dependent Mean	0.07163	Adj R-Sq	0.0686
Intercept	1	0.06994	0.00744	9.4	<.0001	Coeff Var	43.54221		
TD3	1	0.24582	0.1638	1.5	0.1529				
Time Period 4: Model TE4=TD4									
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Root MSE	0.02721	R-Square	0.5403
						Dependent Mean	0.07829	Adj R-Sq	0.5116
Intercept	1	0.07288	0.00653	11.16	<.0001	Coeff Var	34.75099		
TD4	1	0.8003	0.18455	4.34	0.0005				

Summary and Future Research

This paper has investigated factors potentially important in understanding currency ETF tracking error and tracking difference. The analysis found no significant relationships between the fund characteristics studied and the magnitudes of either TE or TD. However, the analysis did show a significant relationship between an individual fund's TE and TD in two of the four subperiods studied.

In future research additional details and issues may be examined. First, the length and timing of the subperiods may be playing a significant role – particularly since this study covers a time period during which various global monetary authorities were making significant changes in policy. It may be useful to specifically incorporate monetary policy events into the study. Second, it is noted that the TER, volume, and fund size data are all from a single most recently available data source. It may well be that using then current data for these characteristics for each of the subperiods could change the results. Third, this study used the mid-range exchange rate data to represent the exchange rate. Since these funds represent both short and long ETFs, it may be that using the implied bid or offer rate instead of the simple mid-range rate would change the results. Fourth, the cumulative daily return definition of TD may be less effective than alternative point to point definitions of TD. Finally, this study concentrated on the ETFs tracking errors and tracking differences. It would also be interesting to assess the practical usefulness of these currency ETFs in foreign exchange risk hedging applications.

References

Abner, David. 2010. *The ETF Handbook*, John Wiley & Sons, Hoboken, New Jersey.

Burney, Robert B. 2012. AHedging with Currency ETFs: The Implications of Return Dynamics.@ *Journal of Accounting and Business, Vol. 5, No.1, 25-35.*

Cain, Alexandra. 2011. ASmall Exporters Getting Savvy About Currency Risk.@ *Sydney Morning News*, July 27.

Cheng, Minder and Ananth Madhaven. 2009. "Dynamics of Leveraged and Inverse ETFs." Barclays Global Investors, October.

DeFusco, R.A., Ivanov, S.I., Karels, G.V., 2011. The exchange traded funds' pricing deviation: analysis and forecasts. *Journal of Economics and Finance* 35(2), 181-197.

De Rossi, Giuliano. 2012. "Measuring the Tracking Error of Exchange Traded Funds: An Unobserved Components Approach." *UBS Investment Research*. June

Hill, Joanne and George Foster. 2009. "Understanding Returns of Leveraged and Inverse Funds." *Journal of Indexes*, September/October.

Hill, Joanne and Solomon Teller. 2010. "Hedging with Inverse ETFs: A Primer." *Journal of Indexes*, November/December, 18-24.

Goodboy, David. 2008.. AFOREX spot trades vs. Currency Futures." *TradingMarkets.com*, May 21.

Hudacheck, Dennis. 2011. ACurrency-Hedged ETFs Lower Volatility." *Index Universe***,** November 16.

Johnson, Ben, Hortense Bioy, Alastair Kellett, and Lee Davidson. 2013. "On The Right Track: Measuring Tracking Efficiency in ETFs." *Morningstar ETF Research*, February.

Lachini, Michael. 2011. "Leveraged and Inverse ETFs: Not Right for Everyone." *Schwab.com*, October 20.

Lachini, Michael and Tatjana Michel. 2011. "Currency ETFs: The Facts and Current Dollar Outlook." *Schwab.com*, October 31.

Merz, Thomas. 2015. "Tracking Risk of Exchange Traded Funds Revisited: A Multivariate Regression Approach." SML Working Paper No. 4. Zurich University of Applied Sciences. January.

Rompotis, G.G. 2010. Investigating the Tracking Error of Exchange Traded Funds, *International Journal of Management Research* 1, 23-42.

Schubert, Leo. 2011. "Hedge Ratios for Short and Leveraged ETFs." *Revista Atlántica de Economia*, Volumen 1, 1-33.

Shin, Sangheon and Gokce Soydmir. 2010. "Exchange-traded Funds: Persistence in Tracking Errors and Information Dissemination." *Journal of Multinational Financial Management*. Vol. 20, Issue4-5, pp. 214-234.

Yates, Tristan. 2007. "The Case Against Leveraged ETFs." *Seeking Alpha.com*, May 17.

Zweig, Jason. 2009. "How Managing Risk with EFTs Can Backfire." *Wall Street Journal*, February 26.

2009. AHedge Against Exchange Rate Risk With Currency ETFs." *Investopedia*, August 12.

2016. "Currency ETF List." *ETFdb.com*. April 29.

DeFusco, R.A., Ivanov, S.I., Karels, G.V., 2011. The Exchange Traded Funds' Pricing Deviation: Analysis and Forecasts. *Journal of Economics and Finance* 35(2), 181-197

The Effect of U.S. Official Reserve Flows on the Japanese Yen-U.S. Dollar Exchange Rate in a Business Cycle

Pablo A. Garcia-Fuentes and Yoshi Fukasawa, Midwestern State University

Abstract

This paper assesses the effect of U.S. official reserves flows as proxy for U.S. central bank intervention on the Japanese yen-U.S. dollar real exchange rate. It covers the period 1974:1-2014:4. The results suggest that Japanese central bank interventions have been aimed at depreciating the yen against the U.S. dollar and that U.S. central bank interventions have been aimed at appreciating the yen against the U.S. dollar. Thus, U.S. official reserves flows can be used as a proxy for U.S. central bank interventions. In addition, it shows that Japanese business cycles affect the real exchange rate.

Introduction

Globalization has made the economies of the world more interrelated. Expansions or recessions in a developed economy affect other economies because of the link through international trade and/or international finance. So, it is likely that policies or any other factors that affect, for example, the Japanese exchange rate also affect the flows of international trade and international finance between the this economy and its trade partners. This has motivated a lot of research that focuses on exchange rate determination. Costa's (2005) survey describes the different methods of exchange rate determination most used in the literature. It identifies behavioral equilibrium exchange rate methods as the ones that explain exchange rate behavior based on fundamental determinants chosen by researchers. These methods include studies that focus on, for example, the relationships between the real exchange rate and productivity differentials (Lee, Nziramasanga, and Ahn, 2002; Miyakoski, 2003), interest rate differentials (Messe and Rogoff, 1988; Edison and Pauls, 1993; Baxter, 1994); the terms of trade (Amano and Van Norden, 1995; Karfakis snd Phipps, 1999; De Gregorio and Wolf, 1994), the real oil price (Chen and Chen, 2007; Huang and Guo, 2007; Lizardo and Mollick, 2010), the real gold price (Kakkar and Yan, 2014), reserve differentials (Tsen, 2011 and 2014), or official reserves (Taylor, 1982; Obsteld, 1983; Kearney and McDonald, 1986; Gartner, 1987 and 1991; Takagi, 1991; Neumann and von Hagen, 1993; Almenkinders, 1996; Szakmary and Mathur, 1997).

Regarding the relationship between official reserves and the real exchange rate, changes in official reserves have been used as a proxy for central bank interventions in the foreign exchange market. However, Neely (2000) argues that official reserves are an imperfect proxy for central bank interventions and that there is not a strong correlation between these two variables. As shown above, even official foreign reserves may not be the best proxy for central bank interventions, they have been included as an explanatory factor in behavioral models of exchange rate determination. Tsen (2011) finds that reserve differential is an important determinant of the real exchange rate between the Japanese yen and the U.S. dollar.

This study uses behavioral models of exchange rate determination between the Japanese yen and the U.S. dollar. First, it assesses the effect of reserve differentials on the real yen-U.S. dollar exchange rate while controlling for the effects of Japanese business cycles and other fundamental factors of exchange rate determination. Second, it assesses the effect of U.S. official reserves flows as a proxy for U.S. central bank interventions while accounting for the effects of Japanese business cycles. We are not aware of U.S. official reserves flows being used in earlier research, so this is the contribution of our research to this literature. The results of the first estimation suggest that Japanese central bank interventions have been aimed at depreciating the yen against the U.S. dollar. In addition, the business cycles dummy variables capture the behavior of the Japanese central bank interventions during the Japanese recessions. The results of the second estimation suggest that Japanese central bank interventions have been aimed at depreciating the yen against the U.S. dollar and, interestingly, that U.S. central bank interventions, proxied by U.S. official reserves flows, have been aimed at appreciating the yen against the U.S. dollar. These results are in line with earlier literature (Ito, 2002; Humpage, 1988). Thus, U.S. official reserves flows can be used to proxy for U.S. central bank interventions.

The rest of the paper is organized as follows. Section two is a review of the literature. Section three describes the data and the methodology. Section four presents the discussion of the results. Section five presents the conclusion and suggestions for further research.

Literature Review

Costa (2005) describes and compares different methods of exchange rate determination. She describes the purchasing power parity method, and the methods that determine time-varying equilibrium exchange rates. The latter group is divided into the structural methods, which include the fundamental equilibrium exchange rate (FEER) methods, and the direct methods, which include behavioral equilibrium exchange rate (BEER) methods, the permanent equilibrium exchange rate (PEER) methods, and the natural rate of exchange (NATREX) methods. BEER methods explain exchange rate behavior based on fundamental determinants chosen by researchers. The inclusion of exchange rate determinants in these models is usually based on the Balassa (1964) hypothesis and the Samuelson (1964) hypothesis, as well as on the Frenkel and Mussa (1984) interest rate parity theory or asset model.

The literature on the determinants of exchange rate behavior is very vast. Lee, Nziramasanga, and Ahn (2002), for example, argue that productivity differential can explain some of the variations in the long-run real exchange rate between New Zealand and Australia. Interest rate effects on exchange rate determination are examined in Messe and Rogoff (1988), Edison and Pauls (1993), and Baxter (1994). Eichenbaum and Evans (1995) report that a contractionary shock to U.S. monetary policy leads to persistent increases in U.S interest rates and prolonged appreciations in the U.S. dollar exchange rates. Bacchetta and Van Wincoop (2010) explained this phenomenon by formalizing Froot and Thaler's (1990) intuition of investors' infrequent portfolio decisions. Commodity prices have also been included in models of exchange rate determination. Chen and Chen (2007), Huang and Guo (2007) and Lizardo and Mollick (2010) postulated that the real oil price could influence the real exchange rate, while Kakkar and Yan (2014) examined the real price of gold as a factor affecting the real exchange rate.

The literature on exchange rate determination has also focused on the relationship between central bank interventions and exchange rates. However, the availability of data on central bank interventions has been a problem. To overcome this problem, some studies have used news reports on central bank interventions (Peiers, 1997; Goodhart and Hesse, 1993). Other research has used data on official reserves (Taylor, 1982; Obsteld, 1983; Kearney and McDonald, 1986; Gartner, 1987 and 1991; Takagi, 1991; Neumann and von Hagen, 1993; Almenkinders, 1996; Szakmary and Mathur, 1997). However, Neely (2000) argues that official reserves are an imperfect proxy for central bank interventions because reserves can change when the central bank intervenes in the foreign exchange market as well as when the government carries out a debt payment in a foreign currency. Neely's (2000) study reports a positive, but not a strong correlation (0.423) between changes in U.S. official reserves and U.S. central bank interventions (market and customer interventions). Furthermore, Suardi and Chang (2012) report correlation asymmetries in purchases of U.S. dollars in the case of the United States and suggest caution on the use of changes of U.S. official reserves when computing an exchange market pressure index. Even official reserves may not be the best proxy for central bank interventions, they have been included as explanatory factor in behavioral models of exchange rate determination.

The effectiveness of interventions by the central bank is said to be nil to significant. Some research suggest little impact of interventions on the Japanese yen-U.S. dollar exchange rate (Jurgensen, 1983; Edison, 1993; Humpage, 2003). On the other hand, Dominguez and Frankel (1993) showed some effects of interventions on the level, volatility, and the risk premium of exchange rate. They emphasized the importance of a signaling effect of interventions on exchange rate expectations. The most recent survey on this literature is done by Taylor (2001) and Remana and Samiei (2000). But, there are no definitive conclusions on the effectiveness of government interventions on the exchange rate. Recently, Tsen (2011) conducted an analysis of exchange rate determination between the currency of Asian economies (Japan, Korea, and Hong Kong) and the U.S. dollar. He found that terms of trade, real oil price, and reserve differential were important determinants of the real exchange rate between the Japanese yen and the U.S. dollar.

Another important issue in the literature on exchange rate is the relationship between business cycles and exchange rate. One example is the correlation among exchange rates, money supply and output in the case of Japan (Chada and Prasad, 1997; Hamori, 1998; Kim, 2000; Hamori and Hamori, 2000). Nadenichek (2000) develops a two-country model to assess the Japan-U.S. trade imbalance and argues that changes in the real exchange rate are due to shifts in the use of resources, subsidies, or productivity levels between the traded and non-traded sectors. Nadenichek's (2000) findings show that depreciation of the real U.S. dollar-yen exchange rate is due to permanent productivity innovations and that exchange rate depreciation has been the most important factor on the U.S. trade deficit with Japan.

Our research uses behavioral models of exchange rate determination. First, it assesses the effect of reserve differentials on the real Japanese yen-U.S. dollar exchange rate while controlling for the effects of Japanese business cycles and other fundamental factors of exchange rate determination. Second, it uses another behavioral model to assess the effect of U.S. official reserves flows as proxy for U.S. central bank interventions while accounting for the effects of Japanese business cycles. We are not aware of this variable being used in earlier research, so this is the contribution of our research to this literature.

Data and Methodology

Data

This study uses quarterly data that covers the period from 1974:1 to 2014:4. The data for most of the variables is obtained from the International Financial Statistics, October, 2015; otherwise, the source is specified. The real exchange rate is defined as $RER_t = NER \times \frac{CPI_{Jpn}}{CPI_{US}}$, where NER is the nominal exchange rate between the Japanese yen and the U.S. dollar, CPI_{Jpn} is the Japanese consumer price index, and CPI_{US} is the U.S. consumer price index. Thus, increases in the value of the real exchange rate represent depreciation of the Japanese yen against the U.S. dollar. The real interest rate differential is defined as $IRD_t = RIR_{Jpn,t} - RIR_{US,t}$, where $RIR_{Jpn,t}$ is Japanese real money market rate or money market rate minus inflation and $RIR_{US,t}$ is the real U.S. federal funds rate, or the federal funds rate minus inflation. Productivity differential is defined as $PD_t = \frac{GDP_{Jpn}}{E_{Jpn}} - \frac{GDP_{US}}{E_{US}}$, where GDP_{Jpn} is Japanese GDP volume, E_{Jpn} is Japanese employment index, GDP_{US} is U.S. GDP volume, and E_{US} is U.S. employment index. The real world oil price is defined as $OP_t = \frac{P_t}{CPI_{Jpn,t}}$, where P_t is the 3-spot oil price index. Reserve differential is defined as $RD_t = \frac{R_{Jpn}}{GDP_{Jpn}} - \frac{R_{US}}{GDP_{US}}$, where R_{Jpn} is Japanese reserves given by the sum of total reserves and gold, GDP_{Jpn} is Japanese GDP in millions of U.S. dollars, R_{US} is U.S. reserves given by the sum of total reserves and gold, and GDP_{US} is U.S. GDP in millions of U.S. dollars. Japanese foreign exchange intervention is defined as $\frac{I_{Jpn,t}}{GDP_{Jpn}}$, where $I_{Jpn,t}$ is Japanese reserves as defined above. United States foreign exchange intervention is defined as $\frac{I_{USt}}{GDP_{US}}$, where $I_{US,t}$ is U.S. official reserves flows (see Pugel, 2007, p. 368-369) as share of GDP. U.S. official reserves flows is defined as the sum of U.S. official reserve assets (line 41), U.S. government assets, other than official reserve assets (line 46), and foreign official assets in the United States (line 56), and it is balance of payment data obtained from the Bureau of Economic Analysis. All dollar figures and indices are in 2010 dollars.

Model and Econometric approach

Given that the main objective of this research is to assess the effect of U.S. official reserve flows on the real exchange rate between the Japanese yen and the U.S. dollar, Equation (2) below is the model of interest. However, we define a model of exchange rate determination between the Japanese yen and the U.S. dollar to assess the effect of reserve differential on the real exchange rate as it has been done in some past studies. This model is defined as

$$LnRER_t = \beta_0 + \beta_1 Trend + \beta_2 IRD_t + \beta_3 PD_t + \beta_4 LnOP_t + \beta_5 RD_t + \delta_i DuRec_j + \varepsilon_t \quad (1)$$

where RER_t is real exchange rate between Japan and the United States, $Trend$ is time trend, IRD_t is interest rate differential between Japan and the United States, PD_t is productivity differential between Japan and the United States, OP_t is real world oil price, RD_t is reserve differential between Japan and the United States, $DuRec_j$ represents dummy variables as a proxy for Japanese recessions ($j = 1,..n$), and ε_{it} is the error term. Ln is the natural logarithm operator. Similar specifications have been used in Tsen (2011, 2014).

The description of the relationships between the real exchange rate and the explanatory variables follows. The interest rate differential is expected to have a negative effect on the real exchange rate (Chen and Chen, 2007). Regarding the relationship between productivity differentials and the real exchange rate, Balassa (1964) and Samuelson (1964) suggest that the higher productivity of traded goods relative to non-traded goods causes appreciation of the real exchange rate. Negative effects of productivity differential are reported in Choudhri and Khan (2005) and Gou (2010), so a negative effect of productivity differential on real exchange rate is expected. It has been found that the oil price accounts for a large amount of the variation of the terms trade (Backus and Crucini, 2000), which suggests that the oil price is an important determinant of the terms of trade. In addition, Tsen (2011) reports a negative correlation (-0.9530) between terms of trade and oil price and finds a negative and significant effect of terms of trade on the real exchange rate for the Japanese yen and the U.S. dollar. Thus, we expect the price of oil to have a negative effect on the real exchange rate. Ergert, Lommatzsch, and Lahreche-Revil (2006) argue that net foreign assets can have either a negative or positive effect on the real exchange rate. However, they find that net foreign assets have a negative effect on the real exchange rate for a sample of OECD countries. Also, Tsen (2011) reports a positive effect of

reserve differentials on the exchange rate between the Japanese yen and the U.S. dollar, so we expect the effect of reserve differentials on the real exchange rate to be positive.

It is important to consider that the level of official international reserves is also affected by official foreign exchange interventions by countries' central banks. By selling or buying foreign exchange, a central bank can affect its official international reserves which thereby affect the exchange rate (Krugman, Obstfeld, and Melitz, 2012, p. 312-313). Ito (2002) examines Japanese foreign exchange interventions over the decade of the 1990s and uses data on Japanese daily interventions to develop dummy variables such that the value of the dummy on the day of the intervention is the amount of the intervention and zero otherwise. He reports that there were some interventions aimed at stopping the yen to appreciate against the U.S. dollar, but there were other interventions aimed at stopping the yen from depreciating too much against the U.S. dollar. He also argues that these interventions were a reaction of Japanese monetary authorities to short-run changes in the exchange rate as well as to deviations of the exchange rate from its long-run levels. He estimates a GARCH model of exchange rate determination that includes Japanese interventions and U.S. interventions as explanatory variables, so a positive estimate on Japanese interventions suggests a depreciation of the yen against the U.S. dollar, but a negative coefficient on U.S. interventions suggests an appreciation of the yen against the U.S. dollar. Humpage (1988) also reports that a sharp depreciation of the dollar between August 1984 and August 1987 happened during high U.S. central bank foreign exchange intervention activity. Humpage (1988) uses data from internal documents on U.S. interventions to develop dummy variables to proxy for U.S. interventions such that the value of the dummy variable is one to represent U.S. interventions and zero otherwise. Chaboud and Humpage (2003) use official Japanese intervention data to analyze Japanese foreign exchange interventions over the period 1992-2002. They report that the effectiveness of Japanese interventions is determined by the frequency and size of the transactions, and that interventions were important to forecast short-term depreciation of the yen after June 1995. Our research differs from these studies because it uses U.S. central bank official reserves flows as a proxy for U.S. central bank foreign exchange interventions. This is a different approach to proxy for U.S. central bank interventions and we are not aware of any other research using this proxy. In addition, this research controls for the effect of business cycles by including dummy variables to represent Japanese recessions. Therefore, the next model of exchange rate determination is our model of interest. That is,

$$LnRER_t = \beta_0 + \beta_1 Trend + \beta_2 IRD_t + \beta_3 PD_t + \beta_4 LnOP_t + \beta_5 I_{Jpn,t} + \beta_5 I_{US,t} + \delta_i DuRec_j + \varepsilon_t \qquad (2)$$

where $I_{Jpn,t}$ is Japanese foreign exchange interventions measured by Japanese holdings of foreign exchange reserves plus gold, $I_{US,t}$ is U.S. foreign exchange interventions measured by U.S. official reserves flows, and the other variables are defined as in Equation (1).

The econometric approach begins with the tests of stationarity for each of the series included in Equations (1 and 2). We conduct augmented Dickey Fuller tests (ADF), Andrews and Zivot (1992) tests that account for one structural break in the series, and Clemente, Montañes, and Reyes (1998) tests that account for one and two structural breaks in the series. Then, given Clemente, Montañes, and Reyes' (1998) unit root tests that control for two structural breaks in the series, we conduct bounds tests to determine whether there is cointegration among the series based on the methodology developed by Pesaran and Shin (1995) and Pesaran, Shin, and Smith (2001). This method allows for identifying cointegration among series that are either *I*(0), or *I*(1), or a mix of both. Thus, Equation (1) is transformed into an autoregressive distributed lag unrestricted error correction model (ARDL-UECM-3). That is,

$$\begin{aligned}\Delta LnRER_t = \alpha + \sum_{k=1}^{n} \Gamma_1 \Delta LnRER_{t-1} + \sum_{k=0}^{n} \Gamma_2 \Delta IRD_{t-k} + \sum_{k=0}^{n} \Gamma_3 \Delta PD_{t-k} + \sum_{k=0}^{n} \Gamma_4 \Delta LnOP_{t-k} + \\ \sum_{k=0}^{n} \Gamma_5 \Delta LnRD_{t-k} + \beta_2 IRD_{t-1} + \beta_3 PD_{t-1} + \beta_4 LnOP_{t-1} + \beta_5 LnRD_{t-1} + v_t\end{aligned} \qquad (3)$$

where α is a drift component, Δ is the first difference operator, v_t is the error term. Note that the beta coefficients in Equation (3) represent the beta coefficients in Equation (1). The gamma coefficients represent short-run effects, while the beta coefficients represent long-run effects.

Equation (3) is estimated by least squares to test the hypothesis that the beta coefficients are zero. Or, $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ against the alternative that $H_1: \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5 \neq 0$ using the *F*-test which is compared against the critical values from Pesaran, Shin, and Smith (2001). Thus, if the *F*-statistics associated with equation (3) is greater than the upper bound critical value, the null hypothesis of no cointegration is rejected; if the *F*-statistic is smaller than the lower bound critical value, the null hypothesis of no cointegration is not rejected; but if the *F*-statistic is in between the two bound critical values the test of hypothesis is inconclusive.

Equation (2) is transformed into an autoregressive distributed lag unrestricted error correction model (ARDL-UECM-4).

$$\Delta LnRER_t = \alpha + \sum_{k=1}^{n} \Gamma_1 \Delta LnRER_{t-1} + \sum_{k=0}^{n} \Gamma_2 \Delta IRD_{t-k} + \sum_{k=0}^{n} \Gamma_3 \Delta PD_{t-k} + \sum_{k=0}^{n} \Gamma_4 \Delta LnOP_{t-k} + \sum_{k=0}^{n} \Gamma_5 \Delta I_{jpn,t-k} + \sum_{k=0}^{n} \Gamma_6 \Delta I_{US,t-k} + \beta_2 IRD_{t-1} + \beta_3 PD_{t-1} + \beta_4 LnOP_{t-1} + \beta_5 I_{Jpn,t-1} + \beta_6 I_{US,t-1} + w_t \quad (4)$$

where α is a drift component, Δ is the first difference operator, w_t is the error term. The beta coefficients in Equation (4) represent the beta coefficients in Equation (2). The gamma coefficients represent short-run effects, while the beta coefficients represent long-run effects.

Equation (4) is estimated by least squares to test the hypothesis that the beta coefficients are zero. Or, $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ against the alternative that $H_1: \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5 \neq \beta_6 \neq 0$ using the *F*-test which is compared against the critical values supplied in Pesaran, Shin, and Smith (2001). Thus, if the *F*-statistics associated with equation (4) is greater than the upper bound critical value, the null hypothesis of no cointegration is rejected; if the *F*-statistic is smaller than the lower bound critical value, the null hypothesis of no cointegration is not rejected; but if the *F*-statistic is in between the two bound critical values the test of hypothesis is inconclusive.

Estimation of Long-Run Effects: Stock-Watson Dynamic OLS (DOLS)

The DOLS method developed by Saikkonen (1991) and Stock-Watson's (1993) is used to estimate the long-run relationships between the real exchange rate and the right-hand side variables in Equations 1 and 2. Ferreira and Harrison (2012) argue that, based on Monte Carlo simulations, the DOLS method has proven to be better than other methods that estimate long-run parameters such as those proposed by Engle and Granger (1987), Johansen (1988), and Phillips and Hansen (1990). The DOLS method not only allows for including variables of different integration order, but also it is asymptotically equivalent to Johansen's maximum likelihood method. This method has been used in several studies (e.g., Ferreira and Harrison, 2012; Herzer and Nowak-Lehnmann, 2006). More recently, Tsen (2014) uses a DOLS to estimate the real exchange rate determination between the Malaysian Ringgit and the U.S. dollar.

Empirical Results

Correlation Analysis

We begin with a correlation analysis between the variable of interest and the usual proxies for U.S. central bank interventions used in the literature. The correlation between U.S. official reserves flows and U.S. central bank daily interventions in the foreign exchange market is positive and highly significant (0.3218). So, if there is a balance of payment deficit, both U.S. official reserves flows and U.S. central bank daily interventions increase. The correlation between U.S. official reserves flows and U.S. reserves level is negative and highly significant (-0.2794). Then, given a balance of payment deficit, U.S. official reserves flows increase and U.S. reserves level decreases. The correlation between U.S. central bank daily interventions in the foreign exchange market and U.S. reserves level is negative and insignificant (-0.1206). Then, given a balance of payment deficit, U.S. central bank daily interventions increase and U.S. reserves level decreases. Therefore, based on these three relations, given a balance of payment deficit, the U.S. government intervenes in the foreign exchange market by selling foreign exchange (a positive daily intervention), which increases U.S. official reserves flows and decreases the level of official reserves. On the contrary, given a balance of payment surplus, the U.S. government intervenes in the foreign exchange market by purchasing foreign exchange (a negative daily intervention), which decreases U.S. official reserves flows and increases the level of official reserves.

Unit Root Tests

The unit root tests are based on Clemente, Montañes, and Reyes' (1998) unit root tests that control for two structural breaks in the series. Therefore, real exchange rate, interest rate differential, productivity differential, real oil price, reserve differential, and Japanese reserves are integrated of order one or *I*(1), while U.S. official reserve flows are integrated of order zero or *I*(0). The next step is to identify whether the series are cointegrated.

Cointegration Test

Given that the series included in Equation (2) have different orders of integration, the Johannsen's (1995) cointegration test cannot be applied. So, we conduct bounds tests to determine whether the series are cointegrated based on the methodology developed by Pesaran and Shin (1995) and Pesaran, Shin, and Smith (2001). The first step is to identify the lag order of the ARDL-UECM. Based on the Akaike and Schwarz's Bayesian information criteria, only one lag is included in the estimation of the ARDL-UECMs for Equations (3 and 4). Table 1 shows the results for the bounds test of cointegration. Note that the *F*-statistic values associated with Equations (3 and 4) are below the *I*(0) critical bounds values. Thus, for Equation (3), the null hypothesis $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ cannot be rejected, so there is no cointegration between the real exchange rate and interest rate differentials, productivity differentials, the real world oil price, and reserve differentials. This also suggests that interest rate differentials, productivity differentials, the real world oil price, and reserve differentials do not have a long-run impact on the real exchange rate. In the case of Equation (4), the null hypothesis $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ cannot be rejected, so there is no cointegration between the real exchange rate and interest rate differentials, productivity differentials, the real world oil price, Japanese intervention, and U.S. intervention. Thus, the next step is to estimate DOLS models of Equations (1 and 2).

Table 1: Bounds Test for Cointegration

Model	***LAG***	***F-statistic***	***10% Critical Bounds***	
			I(0)	***I(1)***
ARDL-UECM-3	1	1.60	2.75	3.79
ARDL-UECM-4	1	1.23	2.75	3.79

Note: The bounds critical values are from Table CI(v): Unrestricted intercept and unrestricted trend in Pesaran, Shin, and Smith (2001). ARDL-UECM-3 includes reserve differential. ARDL-UECM-4 includes Japanese and U.S. foreign exchange interventions.

Dynamic OLS Estimations

The DOLS method allows for estimating the long-run parameters. So, Equation (1) is transformed into Equation (5).

$$LnRER_t = \theta + \beta_1 Trend + \beta_2 IRD_t + \beta_3 PD_t + \beta_4 LnOP_t + \beta_5 LnRD_t + \sum_{k=-n}^{k=n} \Phi_1 \Delta IRD_{t-k} + \sum_{k=-n}^{k=n} \Phi_2 \Delta PD_{t-k} + \sum_{k=-n}^{k=n} \Phi_3 \Delta LnOP_{t-k} + \sum_{k=-n}^{k=n} \Phi_4 \Delta LnRD_{t-k} + \delta_i DuRec_j + \omega_t \quad (5)$$

where $\beta_2, \beta_3, \beta_4$, and β_5, are the long-run parameters, Φ_1, Φ_2, Φ_3, and Φ_4 represent coefficients of the lead and lags differences of the *I*(1) series and are considered as nuisance parameters. These parameters contribute to control for endogeneity, autocorrelation, and nonnormality of the residuals, as well as to consistently estimate the long-run parameters (Herzer and Nowak-Lehnmann, 2006). Similarly to Equation (3), the lead and lag order is one. δ_i (i = 1,...,8) represents coefficients on nine Japanese recession dummy variables. These recessions occurred over the period from 1974:1 to 2014:4.

The DOLS specification of Equation (2) is given by Equation (6). That is,

$$LnRER_t = \theta + \beta_1 Trend + \beta_2 IRD_t + \beta_3 PD_t + \beta_4 LnOP_t + \beta_5 I_{Jap,t} + \beta_6 I_{US,t} + \sum_{k=-n}^{k=n} \Phi_1 \Delta IRD_{t-k} + \sum_{k=-n}^{k=n} \Phi_2 \Delta PD_{t-k} + \sum_{k=-n}^{k=n} \Phi_3 \Delta LnOP_{t-k} + \sum_{k=-n}^{k=n} \Phi_4 \Delta I_{Jpn,t-k} + \delta_i DuRec_j + \omega_t \quad (6)$$

where $\beta_2, \beta_3, \beta_4, \beta_5$, and β_6 are the long-run parameters, Φ_1, Φ_2 Φ_3, and Φ_4 represent coefficients of the lead and lags differences of the *I*(1) series and are nuisance parameters that contribute to control for endogeneity, autocorrelation, and nonnormality of the residuals, as well as to consistently estimate the long-run parameters (Herzer and Nowak-Lehnmann, 2006). δ_i (i = 1,…., 8) represents coefficients on nine Japanese recession dummy variables. These recessions occurred over the period from 1974:1 to 2014:4.

The results of the estimation of Equation (5) are shown in Table 2. DOLS-5 is estimated as AR(3)-ARCH(1) to control for the effect of autocorrelation and heteroskedasticity. The null hypothesis of normality of the residuals is not rejected. The results show that reserve differential has a positive and significant long-run impact on the real exchange rate. Increases in Japanese reserves relative to U.S. reserves lead to a depreciation of the yen against the U.S. dollar. The increase in Japanese reserves is due to Japanese central bank purchasing of foreign exchange, which also represents negative Japanese daily interventions in the foreign exchange market. This result is in line with previous literature that focus on exchange rate determination between the Japanese yen and the U.S. dollar (Tsen, 2011). But, interest rate differentials, productivity differentials, and the real world oil price has no long-run impact on the real exchange rate.

Table 2: DOLS Estimation of Long-Run Effects on the Real Exchange Rate, *AR*(3)-*ARCH*(1) Model, 1974:1-2014:4

DOLS-5	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$	$\boldsymbol{\beta_4}$	$\boldsymbol{\beta_5}$	$\boldsymbol{\delta_1}$	$\boldsymbol{\delta_6}$	$\boldsymbol{\delta_7}$
	-0.0034	-0.0318	0.0053	0.2222***	5.1316***	-0.0908*	0.0657**
	(0.91)	(0.08)	(0.13)	(8.58)	(16.89)	(1.87)	(2.33)

Note: DOLS-5 is an *AR*(3)-*ARCH*(1) model. *t*-ratios in parentheses are underneath the estimated coefficients. ***, **, and * represent significance at the 1%, 5%, and 10% respectively. Diagnostic tests: R^2 = 0.9433, normality test = 0.5720 (0.7513), *AR*(1) = -1.3108 (0.0001), *AR*(2) = 0.1891(0.3226), *AR*(3) = 0.1895(0.0563), *ARCH*(0) = 0.0010(0.0001), *ARCH*(1) = 0.5014(0.0332). The numbers in parentheses next to the diagnostics statistics are *p*-values. The normality test suggests that the residuals are normal.

Table 2 shows interesting results regarding Japanese business cycles and the real exchange rate. We included the coefficients of the recessions that have significant effects on the real exchange rate (The others are not included to save space). The estimate of δ_1 captures the effect of the first recession (1973:q4 to 1975:q1) that is associated with the first oil shock to the Japanese economy. The price of oil increased from $2.18 in February, 1971 to $5.12 in October, 1973 and to $11.65 in January, 1974 (Câmpean cited in Mihut and Daniel, 2012). Note that this coefficient is positive and highly significant and suggests that the Japanese central bank intervened in the foreign exchange market by purchasing foreign exchange to promote depreciation of the Japanese yen against the U.S. dollar. Komiya and Yasui (1984) argue that over the period 1973-1975 Japan experienced three major problems. The first was high inflation. Wholesale prices increased by 37 percent in February 1974 relative to the 1973 level and consumer prices increased by about 25 percent in November 1974 relative to the 1973 level. The second was a decrease in the productive activity, so manufacturing output decreased by 20 percent by the time the economy reached the trough in March 1975. The third was related to the weakening of the value of the yen exchange rate and the worsening of the balance of payments resulting in current account's deficits. The current account deficit was 4.7 billion dollars in 1974. Thus, the positive coefficient on the first recession suggests that the Japanese central bank intervened in the foreign exchange market by purchasing foreign exchange (negative daily interventions as suggested by the correlation analysis) seeking to depreciate the Japanese yen against the U.S. dollar in order to promote exports and to help fixing the current account deficit.

The estimate of δ_6 captures the effect of the sixth recession (1997:q2-1999:q1). This recession is related to the failure of financial institutions during the late 1990s. That is, the default of the Sanyo Securities in November, 1997 that caused the failure of financial institutions such as Hokkaido Takushoku Bank and Yamaichi Securities. The coefficient for this recession is negative and significant and suggests that the Japanese central bank was promoting the appreciation of the yen against the U.S. dollar. Ito (2002) argues that the fluctuation of the Japanese yen/U.S. dollar exchange rate over this period might be related, in large part, to the failure of financial institutions. He reports that, during this period, the yen was fast depreciating against the U.S. dollar, so the Japanese central bank carried out several interventions aimed at appreciating the yen against the dollar. Because the yen/U.S. dollar exchange rate reached a level over 130, an intervention to appreciate the yen was conducted in December 1997. The largest intervention of the 1990s was conducted on April 9, 1998 when the Japanese central bank purchased 2.8 trillion yen, but lowered the yen/U.S. dollar exchange rate by only one or two dollars. The yen/U.S. dollar exchange rate level was 143 on June 16, 1998, so a joint intervention between the Japanese and the U.S. central banks was conducted seeking to appreciate the yen. Consequently, the Japanese yen/U.S. dollar exchange rate became 120 on October 7, 1998, and reached the 110 level by the end of 1998. This made the Japanese central bank to intervene to promote depreciation of the yen, so the yen/U.S. dollar exchange rate became 112 on January 12, 1999, and reached the 123 level in May 1999. Therefore, even though one expects central bank interventions to promote domestic currency depreciation during a recession, the Japanese central bank conducted interventions to promote appreciation of the yen against the U.S. dollar (Ito, 2002), which is captured by the negative and significant sign of the estimate of δ_6.

The estimate of δ_7 captures the effect of the seventh recession (2000:q4 to 2002:q1). This recession is related to the end of the boom of the IT industry (Hirakata, Sudo, Takei, and Ueda, 2014). The coefficient for this recession is positive and significant and suggests that the Japanese central bank was promoting the depreciation of the yen against the U.S. dollar. Based on Ito's (2002) study, the yen started to appreciate in June 1999, so the Japanese central bank conducted four interventions that moved the yen/U.S. dollar exchange rate from the 118 levels to 122 levels. However, the yen/U.S. dollar exchange rate became

101 on January 3, 2000 and other interventions were conducted to depreciate the yen, so the yen/U.S. dollar exchange rate became 126 by March 2001.

The results of Equation (6), which includes our variable of interest, are shown in Table 3. DOLS-6 is estimated as AR(3) model to control for the effect of autocorrelation. The Lagrange multiplier (LM) tests do not indicate heteroscedasticity problems. The Durbin-Watson statistic for lag 3 is 1.6521, but it is in between the lower and upper critical values (1.554; 1.991) obtained from Savin and White (1977), so the test is inconclusive.

Table 3: DOLS Estimation of Long-Run Effects on the Real Exchange Rate, *AR*(3)-Model, 1974:1-2014:4

DOLS-6	β_2	β_3	β_4	β_5	β_6	δ_1	δ_6	δ_7
	-0.0029	0.1285	0.0129	-1.3006*	5.5218***	4.7743***	-0.0653*	0.0544*
	(0.88)	(0.33)	(0.42)	(1.83)	(8.39)	(24.48)	(2.01)	(1.71)

Note: DOLS-6 is an *AR*(3)-model. *t*-ratios in parentheses are underneath the estimated coefficients. ***, **, and * represent significance at the 1%, 5%, and 10% respectively. Diagnostic tests: R^2 = 0.9549; *AR*(1) = -1.2585 (0.0001), *AR*(2) = 0.1939(19.71), *AR*(3) = 0.1676(0.0736); *DW*(1) = 2.0436, *DW*(2) = 2.0859, *DW*(3) = 1.6521, *DW*(4) = 2.0828; *LM*(1) = 1.2836(0.2572), *LM*(2) = 2.0368(0.3612), *LM*(3) = 2.4510(0.4842), *LM*(4) = 2.4611(0.6516). The numbers in parentheses next to the diagnostics statistics are *p*-values.

The hypothesized relation is confirmed for interest rate differential but is not significant. The productivity differential and the real oil price have unexpected signs but are not significant. The interesting results are about the effect of central banks interventions on the real exchange rate. Japanese foreign exchange interventions have a positive and highly significant effect on the real exchange rate. This suggests that Japanese central bank interventions have a long-run impact on the real exchange rate between the Japanese yen and the U.S. dollar. This result confirms Ito's (2002) findings that Japanese interventions have been aimed at depreciating the yen against the U.S. dollar. On the other hand, U.S. central bank interventions have a negative and significant long-run effect on the real exchange rate. This suggests that U.S. interventions have been aimed at appreciating the Japanese yen against the U.S. dollar (depreciating the dollar against the yen). This result supports Humpage's (1988) findings. The business cycles effects on the real exchange rate are the same as those in Table 2. Therefore, the result in Table 3 confirms past studies' findings and Table 2 results. The most important finding is that U.S. official reserves flows have the expected and significant relation with the real exchange rate. Thus, this suggests that U.S. official reserves can be used as a proxy for U.S. central bank interventions.

To assess the robustness of the results in Table 3, we also estimated an AR(3)-ARCH(1) version of Equation (6). The results are shown in Table 4 and are qualitatively the same. The diagnostics tests do not suggest any problems. Note that Japanese central bank interventions are promoting depreciation of the yen against the dollar, while U.S. interventions are promoting appreciation of the yen against the dollar.

Table 4: DOLS Estimation Of Long-Run Effects On The Real Exchange Rate, *AR*(3)-*ARCH*(1) Model, 1974:1-2014:4

DOLS-6'	β_2	β_3	β_4	β_5	β_6	δ_1	δ_6	δ_7
	-0.0038	0.2047	0.0080	-1.3509*	5.5337***	4.7489***	-0.0665	0.0592
	(0.95)	(0.48)	(0.21)	(1.93)	(7.67)	(24.08)	(1.62)	(0.98)

Note: DOLS-6' is an *AR*(3)-*ARCH*(1) model. *t*-ratios in parentheses are underneath the estimated coefficients. ***, **, and * represent significance at the 1%, 5%, and 10% respectively. Diagnostic tests: R^2 = 0.9546, normality test = 1.7416 (0.4186), *AR*(1) = -1.2751(0.0001), *AR*(2) = 0.2289(0.1940), *AR*(3) = 0.1566(0.1107), *ARCH*(0) = 0.0014(0.0001), *ARCH*(1) = 0.1285(0.3886). The numbers in parentheses next to the diagnostics statistics are *p*-values. The normality test suggests that the residuals are normal.

Conclusion

This research has estimated two models of real exchange rate determination between the Japanese yen and the U.S. dollar. The first model assesses the effects of reserve differentials and Japanese business cycles on the real exchange rate. Reserve differential has been one of the usual proxies for central bank interventions in the foreign exchange market used in the literature. To assess these relations, a DOLS model, Equation (5), was estimated. The results suggest that Japanese central bank interventions have been aimed at depreciating the yen against the U.S. dollar. In addition, the business cycles dummy variables capture the behavior of the Japanese central bank interventions during the recessions.

The second model assesses the relationship of interest. Rather than using reserve differential between Japan and the United States, it uses the level of Japanese reserves as a proxy for Japanese central bank interventions and U.S. official reserves flows as a proxy for U.S. central bank interventions. The main objective of this estimation is to assess the effects of the level of Japanese reserves and U.S. official reserves flows on the real exchange rate. To assess this relationship, a DOLS model was used, Equation (6). The results show that Japanese central bank interventions have been aimed at depreciating the yen against

the U.S. dollar and that U.S. central bank interventions aimed at appreciating the yen against the U.S. dollar. These results are in line with earlier literature (Ito, 2002; Humpage, 1988). Therefore, U.S. official reserves flows can be used as another proxy for U.S. central bank interventions in future research. In addition, the business cycles dummy variables capture the behavior of the Japanese central bank interventions as suggested by Equation (5).

Regarding future research, it will be interesting to assess the behavior of U.S. official reserves flows in the exchange rate determination of other major currencies and the U.S. dollar.

References

Amano, R. A., and S. Van Norden. 1995. "Terms of trade and real exchange rates: The Canadian evidence." *Journal of International Money and Finance* 14(1): 83−104.

Andrews, D. and E. Zivot. 1992. "Further evidence on the Great Crash, the oil price shock, and the unit-root hypothesis." *Journal of Business and Economic Statistics* 10:251‑270.

Bacchetta, Philippe, and Eric Van Wincoop. 2010. "Infrequent portfolio decisions: A solution to the forward discount puzzle." *The American Economic Review* 100(3): 870-904.

Backus, D. K., and Mario J. Crucini. 2000. "Oil prices and the terms of trade." *Journal of International Economics* 50(1): 185-213.

Balassa, Bela. 1964. "The purchasing power parity doctrine: A reappraisal." *Journal Political Economy* 72(1): 584-596.

Baxter, Marianne. 1994. "Real exchange rates and real interest differentials: Have we missed the business-cycle relationship?" *Journal of Monetary Economics* 33(1): 5-37.

Chaboud, Alain P., and Owen F. Humpage. 2003. "An analysis of Japanese foreign exchange interventions, 1991 – 2002." Federal Reserve Bank of Cleveland, Working Paper No. 0309.

Chada, B., and E. Prasad. 1997. "Real exchange rate fluctuations and the business cycle: Evidence from Japan." IMF Staff Papers 44(3): 328-355.

Chen, S. S., and H. C. Chen. 2007. "Oil prices and real exchange rates." *Energy Economics* 29(3): 390−404.

Chen, Shiu-Sheng. 2007. "A note on interest rate defense policy and exchange rate volatility." *Economic Modelling* 24(5): 768-777.

Choudhri, Ehsan U., and Mohsin S. Khan. 2005. "Real exchange rates in developing countries: Are Balassa–Samuelson effects present?" *International Monetary Fund Staff Papers* 52(3): 387-409.

Clemente, J., A. Montañes, and M. Reyes. 1998. "Testing for a unit root in variables with a double change in the mean." *Economics Letters* 59: 175‑182.

Costa, Sónia. 2005. "A survey of literature on the equilibrium real exchange rate an application to the euro exchange rate." *Economic Bulletin* 49-64.

De Gregorio, J., and H. Wolf. 1994. "Terms of trade, productivity and real exchange rate." NBER Working Paper no. 4807.

Dominguez, Kathryn M., and Jeffrey A. Frankel. 1993. "Does foreign-exchange intervention matter? The portfolio effect." *The American Economic Review* 83(5): 1356-1369.

Edison, Hali J. 1993. *The effectiveness of central-bank intervention: a survey of the literature after 1982*. Vol. 18. International Finance Section, Department of Economics, Princeton University, 1993.

Edison, Hali J., and B. Dianne Pauls. 1993. "A re-assessment of the relationship between real exchange rates and real interest rates: 1974–1990." *Journal of Monetary Economics* 31(2): 165-187.

Engle, Robert F., and Clive W. J. Granger. 1987. "Cointegration and error correction: Representation, estimation and Testing." *Econometrica: Journal of the Econometric Society* 55(2): 251–76.

Eichenbaum, M., and C. L. Evans. 1995. "Some empirical evidence on the effects of shocks to monetary policy on exchange rates." *Quarterly Journal of Economics* 110: 975–1009.

Egert, B., Lommatzsch, K., and A. Lahreche-Revil. 2006. "Real exchange rates in small open OECD and transition economies: Comparing apples with oranges?" *Journal of Banking and Finance* 30(12): 3393−3406.

Ferreira, G. F.C., and R.W. Harrison. 2012. "From coffee beans to microchips: Export diversification and economic growth in Costa Rica." *Journal of Agricultural and Applied Economics* 44(4): 517-531.

Frenkel, Jacob, and Michael L. Mussa. 1984. "Asset markets, exchange rates and the balance of payments." NBER Working Paper No. 1287.

Froot, Kenneth A. and Richard H. Thaler. 1990). "Anomalies: Foreign ex-change." *Journal of Economic Perspectives* 4(3): 179-192.

Gartner, Manfred. 1987. "Intervention policy under floating exchange rates: An analysis of the Swiss case," *Economica* 439-53.

_____. 1991. "Foreign-exchange markets and central-bank intervention." In *Markets and Politicians: Politicized Economic Choice, Studies in Public Choice*, edited by Arye L. Hillman. Kluwer Academic. 319-38.

Goodhart, Charles A. E., and Thomas Hesse. 1993. "Central bank forex intervention assessed in continuous time," *Journal of International Money and Finance* 368-89.

Guo, Qian. 2010. "The Balassa–Samuelson model of purchasing power parity and Chinese exchange rates." *China Economic Review* 21(2): 334-345.

Hamori, S. 1998. "Money, exchange rates and international business cycle between Japan and the United States." *Applied Economic Letters* 5: 31-35.

Hamori, S., and N. Hamori. 2000. "An empirical analysis of economic fluctuations in Japan, Japan: 1885-1940." *Japan and the World Economy* 12: 11-19.

Herzer, D., and D.F. Nowak-Lehnmann. 2006. "What does export diversification do for growth? An econometric analysis." *Applied Economics* 38(15): 1825–38.

Hirakata, Naohisa, Nao Sudo, Ikuo Takei, and Kozo Ueda. 2014. "Japan's financial crises and lost decades." Federal Reserve Bank of Dallas, Working Paper No. 220.

Huang, Ying, and Feng GUO. 2007. "The role of oil price shocks on China's real exchange rate." *China Economic Review* 18(4): 403-416.

Humpage, Owen. 1988. "Intervention and Dollar's Decline." Federal Reserve Bank of Cleveland. *Economic Review* 24(2): 2-17.

Humpage, Owen. 2003. "Government intervention in the foreign exchange market." Federal Reserve Bank of Cleveland, Working Paper No. 03-15.

Ito, T. 2002. "Is foreign exchange intervention effective? The Japanese experience in the 1990's." NBER Working Paper No. 8914.

Johansen, S. (1988). "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control* 12(2–3): 231–54.

____. 1995. *Likelihood-based Inference in Cointegrated Vector Autoregessive Models*. Oxford University Press, New York.

Jurgensen, Philippe. 1983. "Report of the working group on exchange market intervention," Washington, D.C., U.S. Department of the Treasury.

Karfakis, C., and A. Phipps. 1999. "Modeling the Australian dollar–US dollar exchange rate using cointegration techniques." *Review of International Economics* 7(2): 265−279.

Kim, J. 2000. "The relationship between the monetary regime and output volatility: a multivariate GARCH-M model of the Japanese experience 1919-1996." *Japan and the World Economy* 12: 49-69.

Komiya, Ryutaro, and Kazuo Yasui. 1984. "Japan's macroeconomic performance since the first oil crisis: Review and appraisal." In *Carnegie-Rochester Conference Series on Public Policy* 20: 69-114. North-Holland.

Lee, M., M. Nziramasanga, and S. K. Ahn. 2002. "The real exchange rate: An alternative approach to the PPP puzzle." *Journal of Policy Modeling* 24(6): 533−538.

Lizardo, Radhamés A., and André V. Mollick. 2010. "Oil price fluctuations and US dollar exchange rates." *Energy Economics* 32(2): 399-408.

Meese, Richard, and Kenneth Rogoff. 1988. "Was it real? The exchange rate-interest differential relation over the modern floating-rate period." *The Journal of Finance* 43(4): 933-948.

Mihut, Marius Ioan, and Decean Liviu Daniel. 2012. "First Oil Shock Impact on the Japanese Economy." *Procedia Economics and Finance* 3: 1042-1048.

Miyakoshi, T. 2003. "Real exchange rate determination: Empirical observations from East-Asian countries." *Empirical Economics* 28(1): 173−180.

Nadenichek, Jon. 2000. "The Japan–US trade imbalance: A real business cycle perspective." *Japan and the World Economy* 12(3): 255-271.

Neely, Christopher J. 2000. "Are changes in foreign exchange reserves well correlated with official intervention?" *Review-Federal Reserve Bank Of Saint Louis* 82(5): 17-32.

Neumann, Manfred J.M., and Jurgen von Hagen. 1993. "Germany." In *Monetary Policy in Developed Economies: Handbook of Comparative Economic Policies*, Michele Fratianni and Dominik Salvatore. Greenwood Press. 299-334.

Obstfeld, Maurice. 1983. "Exchange rates, inflation, and the sterilization problem: Germany, 1975-1981," *European Economic Review* 161-89.

Pesaran, M.H., Y. Shin, and R.J. Smith. 2001. "Bounds testing approaches to the analysis of level relationships." *Journal of Applied Econometrics* 16(3): 289–326.

Pesaran, M.H., and Y. Shin. 1995. "Chapter 11: An autogressive distributed lag modeling approach to cointegration analysis." *Econometrics and Economic Theory in 20th Century: The Ragnar Frisch Centennial Symposium*. Cambridge, MA: Cambridge University Press, 1995.

Phillips, P.C.B., and B.E. Hansen. 1990. "Statistical inference in instrumental variables regression with I(1) processes." *The Review of Economic Studies* 57(1): 99–125.

Peiers, Bettina. 1997. "Informed traders, intervention, and price leadership: A deeper view of the microstructure of the foreign exchange market," *Journal of Finance* 1589-614.

Pugel, Thomas. 2007. *International Economics*. 13 ed. The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, 10020.

Kakkar, Vikas, and Isabel Kitming Yan. 2014. "Determinants of real exchange rates: An empirical investigation." BOFIT Discussion Paper No. 1.

Kearney, Colm and Ronald MacDonald. 1986. "Intervention and sterilisation under floating exchange rates: The UK 1973-1983," *European Economic Review* 345-64.

Komiya, Ryutaro, and Kazuo Yasui. 1984. "Japan's macroeconomic performance since the first oil crisis: Review and appraisal." In *Carnegie-Rochester Conference Series on Public Policy* (20): 69-114. North-Holland.

Krugman, P., M. Obsteld, and J. M. Melitz. 2012. *International Economics Theory and Policy* 9th Ed. Pearson Education, Inc. Boston, MA 02116.

Ramaswamy, Ramana and Hossein Samiei. 2000. "The yen-dollar rate: Have interventions mattered?" IMF Working Paper, WP/00/95.

Samuelson, P. A. 1964. "Theoretical notes on trade problems." *Review of Economics and Statistics* 46(2): 145−154.

Saikkonen, P. 1991. "Asymptotically efficient estimation of cointegrating regressions." *Econometric Theory* 7: 1–21.

Stock, J. H. and M. W. Watson. 1993. "A simple estimator of cointegrating vectors in higher order integrated systems." *Econometrica* 61: 783–820.

Savin, Nathan E., and Kenneth J. White. 1977. "The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors." *Econometrica: Journal of the Econometric Society* 1989-1996.

Suardi, Sandy, and Yuanchen Chang. 2012. "Are changes in foreign exchange reserves a good proxy for official intervention?" *Journal of International Financial Markets, Institutions and Money* 22(4): 678-695.

Takagi, Shinji. 1991. "Foreign exchange market intervention and domestic monetary control in Japan, 1973-89." *Japan and the World Economy* 147-80.

Taylor, Dean. 1982. "Official intervention in the foreign exchange market, or, bet against the central bank." *Journal of Political Economy* 356-68.

Taylor, John B. 2001. "The role of the exchange rate in monetary-policy rules." *The American Economic Review* 91(2): 263-267.

Tsen, W. H. 2011. "The real exchange rate determination: An empirical investigation." *International Review of Economics and Finance* 20: 800-811.

Tsen, W. H. 2014. "The real exchange rate determination: An empirical evidence from Malaysia." *The Singapore Economic Review* 59(2): 1450016(1-19).

Economic Growth and Revitalization on Long Island: the Role of the Recreational Fishing and Marine Economy

Sheng Li, Richard Vogel, & Nanda Viswanathan, Farmingdale State College

Abstract

Long Island's economy grew from a set of farm and marine based communities into a diverse service based economy. These traditional industries have for the most part been displaced as a result of increasing development pressures throughout the Island and its coastal areas. Recreational fishing is still a significant factor in local and tourism related recreational activities. This study, part of an ongoing 2-year research project funded by New York Sea Grant, focuses on the role that marine based activity may play in the region's economic activity and policies that may be implemented to reinvigorate the sector. Results show that there are still pockets of Long Island that are dependent on marine-related activities, and that these areas have undergone rapid change in the past decade. Tourism and transportation are the dominating sources contributing to the marine coastal economy.

Introduction

Maritime based industry has played a crucial role in Long Island's coastal economic development. However, potential growth in tourism and recreation, transportation, construction related-commerce business sectors, are positioned to be the driving economic engines (Pomeroy, et al. 2013). While the number of individuals directly working within marine based industries such as commercial and recreational fishing are relatively small in relationship to the greater Long Island economy, they may be able to serve as a vehicle to help fuel the regional tourism sector and to provide substantial support to local economic growth.

Recent studies such as Ma (2014) found the marine sector provided not only the necessary foods and resources but also offer the basic platform for the transportation industry and tourism industries. Zhang (2007)'s study suggested that wildlife recreational activities have significant contributions both on income and employment. Colgan (2013) estimated the total size of the U.S marine economy was 2.68 million employees working in over 140,000 establishments and earning nearly $94 billion in wages in 2007. They contributed over $238 billion to the U.S. GDP. The National Marine Fisheries Service's (NMFS) report in 2014 demonstrates that in 2012, there were over 11 million marine recreational anglers across the U.S. who spent $4.6 billion on fishing trips and $20 billion on durable fishing-related equipment. These expenditures contributed over $58 billion in sales to the U.S. economy, generated $30 billion in value added impacts, and supported over 381,000 job impacts (NMFS, 2014).

On Long Island, the marine economy contributes significantly to the state's economy in tourism, recreational fishing, and other activities. Some traditional industries, like fishing, have been threatened and displaced as a result of increasing development pressures in coastal areas (Carey, 2014). Even though the commercial fishery is still important to the region, tourism (in which recreational fishing-based tourism still plays a role), transportation, and the information technology sector have become significant economic activities in the local economy.

The purpose of this paper is to assess the contributions that marine and related industries provide in the overall Long Island economy as part of an overall evaluation of strategies to revitalize the marine-based economy of the region. As part of that process, we also evaluate the economic trends affecting the marine economy on Long Island. Recreational fishing is the largest component of the maritime sector, and we pay particular attention its interrelationship with tourism as it is possible that recreational fishing may be able to play a significant role in furthering this sector.

Marine Growth and Revitalization on Long Island

Since 2001, Long Island's economy (Nassau County and Suffolk County) has grown from a gross regional product of $105 billion to $170 billion (2014), expanding by 62 percent. Over this same period, New York State's economy grew by 61 percent to an annual gross state product of $1.4 trillion in 2014. , having risen by a comparable 61 percent over the 14-year period. As a share of NYS's GSP, the two counties' economy contribute 12 percent of the gross product.

Table 1 shows the GDP of those two Counties in comparison with the output of the nation and New York State. Before the recessions, economic growth in the Long Island is comparable with the national level and even higher than the state level. Since 2007, the great recession slowed economic growth with national growth and state growth declining to 3.7 percent and 7.7 percent respectively. Long Island's economy continued to increase with a 12 percent from 2007 to 2010. By comparison,

Suffolk grew faster than Nassau in most of years since 2001, and Nassau's growth fell to 1.5 percent during 2010 to 2011. By the end of 2014, Suffolk had surpassed Nassau in GDP by roughly $1.4 billion.

Table 1. A Comparison of GDP Growth by Area, Billion Dollar, 2001-2009

All Industries	*2001*	*2004*	*2007*	*2010*	*2011*	*2014*
US	10,562	12,207	14,391.1	14,859.8	15,406.0	17,232.6
NYS	864.4	954.2	1,120.9	1,207.6	1,230.1	1,395.5
Nassau County	54.2	62.0	69.0	75.3	76.4	84.0
Suffolk County	50.7	59.4	67.2	75.7	77.3	85.4
Nassau County+ Suffolk County	104.8	121.3	136.2	151.0	153.7	169.5
GDP Change in %	*01-04*	*04-07*	*07-10*	*10-11*	*11-14*	*01-14*
US	15.6%	17.9%	3.3%	3.7%	11.9%	63.2%
NYS	10.4%	17.5%	7.7%	1.9%	13.4%	61.4%
Nassau County	14.4%	11.2%	9.2%	1.5%	10.0%	55.1%
Suffolk County	17.2%	13.2%	12.6%	2.1%	10.5%	68.6%
Nassau County+ Suffolk County	15.7%	12.2%	10.9%	1.8%	10.2%	61.7%

Source: U.S. Bureau of Economic Analysis, ENOW, and NOAA Office of Coastal Management

The industrial composition of income in Long Island is presented in Table 2. Total personal income for 2010-2014 was extended from 161 billion dollars to 185 billion dollars. With $14.9 billion of annual income in 2014, the trade sector (both retail and wholesale) is the largest component on Long Island, which rose 14 percent over the period of 2010-2014. Scientific and Technical services ranked the second most important sector increasing slightly during the past four years. Finance & insurance and the construction industry contributed $7.9 billion and $7.1 billion income respectively in 2014. However, income from agriculture and forestry sector, including fisheries based industries is only $0.4 billion. Service industries dominated the Long Island economy while the proportion of traditional sectors, e.g. agriculture and manufacturing have declined.

Marine based and related sectors accounted for 13% of the total GDP and provided more than 45 thousand jobs in 2012 (Table 3). Marine industries including nature species, ports, and waterways support a broad range of sectors, such as shipbuilding, commercial fishing, aquaculture, marine manufacturing, marine engineering, port services, marine terminal operation, marine construction, marine maintenance, marine environmental services, and transportation. Tourism and recreation, construction, and transportation are all ranked highly in the distribution of industry on Long Island. Figure 1 shows the development of marine industries from 2005 to 2012. Jobs added over this period totaled 6,425, or 16% growth, contributing to an increase in GDP of $ 0.5 billion or 25% growth.

Tourism and recreation continue to provide a strong impetus for economic growth in Long Island with consistent growth over time. In 2012, recreational anglers generated $369 million in sales. Tourism and recreation contributed $1.7 billion to the gross product and supported 42,000 jobs (Herfaut et al. 2013). Even during the recession, the output and employment within this sector maintained itself at a stable level. Since 2010, the whole industry has grown, and the sector averaged nearly 13 percent growth in employment and over 13 percent growth in GDP from 2010 to 2012 (Figure 2).

Table 2. Long Island's Economy – Income by Super sector in Billions Dollar, 2010-2013

	Nassau County					*Suffolk County*				
	2010	*2011*	*2012*	*2013*	*2014*	*2010*	*2011*	*2012*	*2013*	*2014*
Total Personal Income	86.1	91.3	96.5	95.7	100.0	74.6	78.1	81.81	81.73	85.26
Trade (Retail+Wholesale)	6.48	6.73	7	7.22	7.51	6.34	6.6	6.96	7.04	7.35
Scientific & Technical	4.27	4.54	4.66	4.86	5.09	3.50	4.00	4.08	4.14	4.37
Finance And Insurance	3.98	3.78	3.82	3.82	3.96	3.83	4.04	4.08	3.85	3.96
Construction	2.72	2.79	2.95	3.03	3.29	2.97	3.06	3.30	3.59	3.85
Information	2.97	2.88	3.13	2.89	2.78	0.97	0.97	1.00	0.98	1.10
Agriculture and Forestry	0.01	-	0.01	0.01	-	0.03	0.03	0.03	0.03	0.03

Source: U.S. Bureau of Economic Analysis

Table 3. Marine Economy by Sectors, Long Island, 2012.

	Nassau County		Suffolk County		Total	
	Employment	*GDP(millions)*	*Employment*	*GDP(millions)*	*Employment*	*GDP(millions)*
All Maine Sectors	15,242	656	30,353	1,610	45,595	2,266 (13% of GDP)
Living Resources	182	10	286	21	468	31
Tourism & Recreation	14,396	568	26,614	1133	41,010	1,700
Construction	92	15	340	35	432	50
Minerals	16	1	27	2	43	3
Transportation	556	62	3,086	420	3,642	482

Note: Data sources; NOAA Office of Coastal Management. Available at: www.coast.noaa.gov/digitalcoast/data/enow/.

The natural resources sector, primarily concentrated in commercial fishing is still popular in Long Island. Commercial fishing industries landed more than 27 million pounds of finfish and shellfish worth $37.6 million in 2012. However, this sector shows a rather high volatility over time due to extreme weather, changes in fish stocks and regulations, as well as the economic recession (Zhang et al. 2015). Fish landings peaked in 2005 at 17 thousand tons and declined to 14 thousand tons in 2011, a 25% reduction (Figure 3).

Marine waters are also an economically important area for transportation including commercial shipping of goods and commodities, and freight entering the country and the state. The transportation sector processed over $482 million output and supporting 3,600 jobs in 2012. Transportation activities increased significantly between 2006 and 2009, although there was a slight reduction in GDP at the beginning of 2010 (Figure 4).

Figure 1. Marine Economy Changes in Long Island NY, 2005-2012

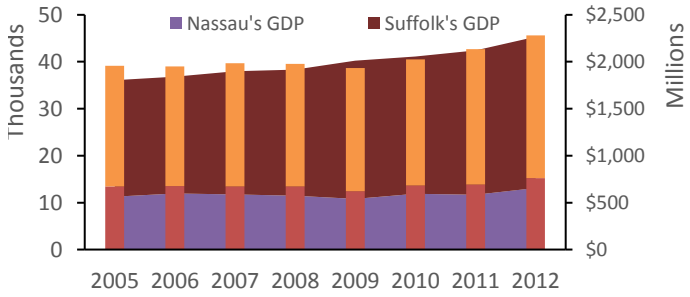


Figure 2. Economic Changes in the Tourism Sector, Long Island NY, 2005-2012

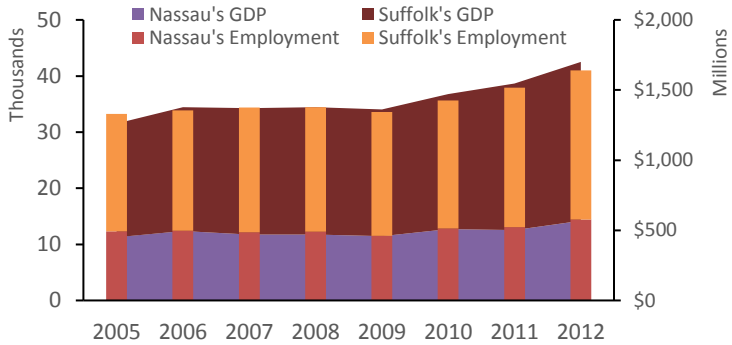


Figure 3. Economic Changes in the Living Resources Sector, Long Island NY, 2005-2012

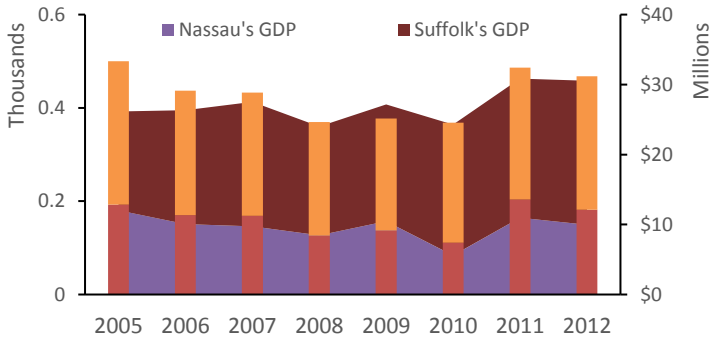
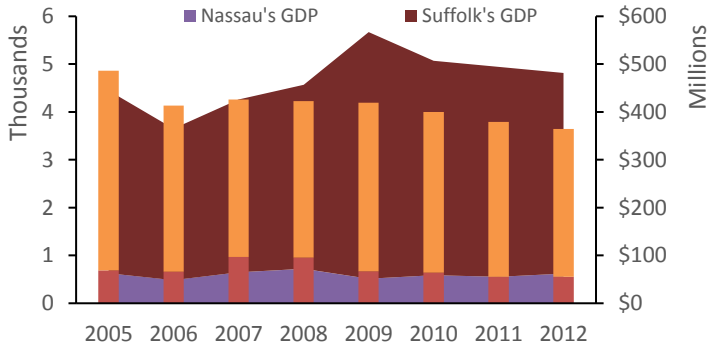


Figure 4. Economic Changes in the Transportation Sector, Long Island NY, 2005-2012



Marine Recreational Fishing: a Driver of Marine Economic Activity

Recreational fishing on Long Island has been thriving since 1998, and the proportion of recreational and sports fishing increased continuously over the last two decades. Recreational fishing generated $316 million dollars in expenditures in New York State as a whole in 2012 (NMSF 2014) and led to economic impacts estimated at $381 million in total output, close to $242 million in value added, $151 million in income, and total employment of 2,959 people. It peaked in 2007 when recreational fishing in the state contributed up to 1 billion in outputs, almost 6,000 people employed, and $0.5 billion of value added.

Table 4. Economic Impacts of Recreational Fishing Expenditures, New York, 2012

	2006	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*
Employment Impacts (1000)	5.37	6.49	5.77	4.57	4.46	3.09	2.96
Output Impacts ($ Millions)	812.27	979.19	875.45	680.46	667.85	398.88	381.30
Value Added Impact s($ Millions)	424.07	511.31	457.20	358.11	350.16	254.73	241.95
Income Impacts ($ Millions)				231.73	227.22	160.03	151.10

Note: Output impacts reflect total dollar sales generated from marine recreational fishing expenditures. Value-added impacts represent the contribution marine recreational fishing makes to gross domestic product. Income impacts represent wages, salaries, benefits, and proprietary income generated from marine recreational fishing. Date source: NOAA Interactive Fisheries Economic Impacts. https://www.st.nmfs.noaa.gov/apex/f?p=160:7:0::NO

Total angler expenditures on marine recreational fishing in New York were $330.3 million in 2011. Trip expenditures were $205.9 million and expenditures on durable goods were $124.4 million. Mean trip expenditures by residents on for-hire fishing trips were $157.83, $59 on private boat trips, and $19.91 for shore trips, compared with $116.37, $38.83, and $44.68 for Non-resident respectively. Recreational fishing also generated $78 million tax revenue, including 40.4 in local tax and 37.4 in federal tax. 70% of the total come from business and households.

Table 5. 2011 Economic Impacts of Recreational Fishing Expenditures (millions of dollars)

Trip Impacts By Fishing Mode:	*Expenditures*
For-Hire	66.3
Private Boat	115.7
Shore	23.9
Total Durable Equipment	124.4
Total State Trip and Durable Equipment	330.3

Table 6. Federal and State Tax Impacts Generated from Marine Recreational Fishing in New York in 2011, (billions)

	Employee Compensation	*Proprietor Income*	*Indirect Business Tax*	*Households*	*Corporations*	*Total*
Local Tax	0.5		27.9	7.8	4.2	40.4
Federal Tax	14.9	1.1	3.4	15.2	3.1	37.7
Total	15.4	1.1	31.3	23.0	7.2	78.1

One critical connection between the coastal and marine-based industries and the larger economy is through the greater tourism sector operating in the region. Nunes et al. (2009) pointed out coastal resources, and ecosystems can contribute significant economic benefits to regional economies especially regarding coastal tourism and recreational services. As the second most popular tourist destination in New York State after New York City, over 9 million overnight visitors come to Long Island annually generating approximately $5.1 billion in spending in 2013 (Matejka, 2014). Overall, tourism supports more than 70,000 jobs or 5.9 percent of jobs on Long Island and is responsible for generating nearly $638 million in state and local tax dollars. Based on a multiplier of 4.3, tourism spending created a $27.4 billion in economic impact on Long Island in 2013.

Table 7 reports the total tourism impact in the Long Island in 2010. Traveler spending in Nassau and Suffolk counties was $4.6 billion, contributed $2.5 billion labor income, supported 70 thousand jobs, and generated $0.56 billion in taxes. The region's tourism economy is roughly evenly split between Suffolk and Nassau Counties.

Table 7. Tourism Economic Impact, Long Island, 2010

	TRAVELER SPEND **(**$Billion**)**	***LABOR INCOME*** ($Billion)	***EMPLOYMENT***	***LOCAL TAXES*** ($Billion)	***STATE TAXES*** ($Billion)
NASSAU	$2.13	$1.18	33,445	$0.13	$0.13
SUFFOLK	$2.47	$1.30	36,675	$0.15	$0.15
TOTAL	$4.60	$2.47	70,120	$0.28	$0.29

Four percent of all labor income on Long Island is generated by tourism, with 4.2 percent in Suffolk County and 3.8 percent in Nassau County. The direct impact was $1.4 billion, and total impact estimated at $2.4 billion. Tourism generates 5.9 percent of all employment on Long Island, with a direct impact of 48 thousand jobs and a total impact of 70 thousand jobs. (Table 8).

Table 8. Economic Impact and Contribution from Tourism in Long Island, 2010

		Direct	*Total (Direct, Indir., Induced)*	*Share (Total)*
Lab. Income ($Billion)	*Nassau*	$0.69	$1.18	3.8%
	Suffolk	$0.76	$1.30	4.2%
	Total	$1.45	$2.47	4.0%
Employment	*Nassau*	23,072	33,445	5.7%
	Suffolk	25,301	36,675	6.1%
	Total	48,373	70,120	5.9%

Discussion and Conclusions

The recreational fishing industry contributes substantially to the Long Island economy both through the creation of employment and output, and through its interrelationship with the rest of the Island's economy. As a driver for related industries, recreational fishing attracted tourists to the region and this in turn supported a number of related industries including

the retail sector, restaurants, hotels and other places of lodging, and related support service industries. Those impacts extend to the greater regional economy, and coastal based activities, which have the potential to provide substantial support to the region's economic growth.

This study focuses on the role that marine based activity may play in the region's economic activities and to assess the contribution of marine and related industries on the overall Long Island economy, and to evaluate strategies to revitalize the marine sector. Despite the significance of the marine economy, it faces significant challenges due to the damages associated with the loss of ecosystem services, water pollution, competition, and a lack of infrastructure investment in the sector. Decreased social resilience and increased vulnerability to both natural and human disasters have been highlighted by some scholars (Bowen, 2003; Douvere 2008; Malone et al. 2010).

This project is still in its early stages though. Initial surveys (150) were conducted late in the tourism season (late August) of 2015, and will again be conducted beginning in the late spring and early summer of this year. We are also in the process of developing a detailed model of the region's economy (CGE) to evaluate the interrelationship between the recreational fishing industry and the greater economy. Once the model is completed we will be able to evaluate better policies that may help to revitalize the marine based industries and sector on Long Island.

Acknowledgement

This paper is a resulting product from project R/CHD-8, entitled Leveraging Long Island's Coastal Heritage for the Future: Integration and Diversification of Long Island's Coastal Industries funded under award 67209 from the National Sea Grant College Program of the U.S. Department of Commerce's National Oceanic and Atmospheric Administration, to the Research Foundation of State University of New York on behalf of New York Sea Grant. The statements, findings, conclusions, views, and recommendations are those of the authors and do not necessarily reflect the views of any of those organizations."

References

Bowen, Robert E., and Cory Riley. 2003. "Socio-economic indicators and integrated coastal management." *Ocean & Coastal Management* 46.3: 299-312.

Colgan, Charles S. 2013. "The ocean economy of the United States: Measurement, distribution, & trends." *Ocean & Coastal Management* 71: 334-343.

Carey-F, Richard. 2014. "The Oceans Economy: Opportunities and Challenges for Small Island Developing States."

Douvere, Fanny. 2008. "The importance of marine spatial planning in advancing ecosystem-based sea use management." *Marine policy* 32.5: 762-771.

Herfaut, Johanna, et al. 2013."The nationwide assessment of marine recreational fishing: A French example." Ocean & Coastal Management 78: 121-131.

Ma, Xing, and Xian Wei Liu. 2014. "Coupled Development and Process of Marine Economy and Marine Environment in China." *Advanced Materials Research*. Vol. 962.

Malone, Thomas, et al. 2010. "Climate change, sustainable development and coastal ocean information needs." *Procedia Environmental Sciences* 1: 324-341.

Majetka, K. 2014. Personal communication with Kristen Matjeeka, Director of Marketing and Communications, Long Island Visitors and Convention Bureau and Sports Commission.

National Marine Fisheries Service. 2014. Fisheries Economics of the United States, 2012. U.S. Dept. Commerce, NOAA Tech. Memo. NMFS-F/SPO-137, 175p. https://www.st.nmfs.noaa.gov/st5/publication/index.html.

Nunes, P.A.L.D., Portela, R., and Rao, N. 2009. "Recreational, Cultural and Aesthetic Services from Estuarine and Coastal Ecosystems." Working Papers 121, Fondazione Eni Enrico Mattei.

Pomeroy, Robert S., Nataliya Plesha, and Umi Muawanah. 2013. Valuing the Coast: Economic Impacts of Connecticut's Maritime Industry. No. 15. University of Connecticut, Department of Agricultural and Resource Economics, Charles J. Zwick Center for Food and Resource Policy.

Regional Impact Evaluation: An initial Assessment of the Economic Impacts of Sandy on New Jersey and New York Commercial and Recreational Fishing Sectors, March 15, 2013. NOAA Fisheries, Office of Science and Technology, NEFSC.

Timmons, M et al. 2004. New York aquaculture industry: status, constraints and opportunity. *www.organicity.org/food/urbaqua/**NewYorkAquaculture**.pdf*

Zhang, J. and Lee, D.J. 2007. "The Effect of Wildlife Recreational Activity on Florida's Economy." *Tourism Economics* 13(1): 87-110.

Zhang, Yaoqi, Sheng Li, and Zhimei Guo. 2015. "The Evolution of the Coastal Economy: The Role of Working Waterfronts in the Alabama Gulf Coast." *Sustainability* 7(4): 4310-4322.

Natural Disasters in Latin America: Disaster Type and the Urban-Rural Income Gap

Madeline Messick, University of Southern Mississippi

Abstract

Natural disasters of varying types (i.e. droughts, earthquakes, floods) have differing impacts on the productive sectors of agriculture, industry, and services; which due to the concentration of certain productive sectors in urban or rural areas (i.e. agriculture is closely associated with rural areas, while industry and services are usually located in more densely populated urban areas), subsequently results in changes to the urban-rural income gap. The results show that earthquakes lead to a decrease in the relative strength of urban incomes when compared to rural while droughts and wildfires increase the gap between rural and urban incomes, leading to a decline in the relative position of rural incomes when compared to urban.

Introduction

This research examines the impact of natural disasters on income inequality in Latin America. It posits that natural disasters of varying types (i.e. droughts, earthquakes, floods) have differing impacts on the productive sectors of agriculture, industry, and services; which due to the concentration of certain productive sectors in urban or rural areas (i.e. agriculture is closely associated with rural areas, while industry and services are usually located in more densely populated urban areas), subsequently results in differing impacts on income inequality in urban or rural areas. The research conducted here uses the urban-rural income gap to examine changes to inequality in Latin American countries as a result of the type of disaster.

Background

According to the National Climate Assessment (Melillo, Richmond and Yohe 2014), extreme weather events will become increasingly common as a result of warming temperatures. Warming temperatures are also expected to result in greater weather volatility. Among the extreme weather events expected to increase in frequency are heat waves, droughts, floods, hurricanes (including storm surges that reach farther inland), and winter storms (Melillo, Richmond and Yohe 2014). Developing countries, with limited resources for preparedness and recovery, are expected to be highly vulnerable to severe natural disasters (Mohapatra, Joseph and Ratha 2012).

Natural disasters are an exogenous shock to the economy of a country. Depending on the severity of the disaster, the shock to the economy may be confined to a local area or have far-reaching impacts across the entire economy. The economic impact of disasters includes direct losses such as damaged or destroyed buildings, crops, or equipment, but also indirect losses resulting from decreased productive capacity due to the displacement or loss of labor or damaged infrastructure (Guha-Sapir, Hargitt and Hoyois 2004). Secondary losses can result from changes to capital availability, government spending, or interest rates.

The combination of increasing natural disasters and the vulnerability of developing countries highlights the need to understand the impact of natural disasters on the economies of developing countries. Increased understanding of the ways that disasters, and type of disaster, affect income inequality and rural to urban migration will play a critical role in addressing the negative externalities associated with disasters.

This research is the first to create a cohesive model of how natural disasters affect rural and urban economic inequality. It is hypothesized that changes to the income gap between urban and rural areas results from the differential impact of disasters on the sectors of agriculture, industry, and services. This impact differs based the type of disaster, as some disasters, such as droughts and wildfires, are predicted to affect rural areas more than urban, while other disasters, such as earthquakes, are predicted to affect urban areas more than rural.

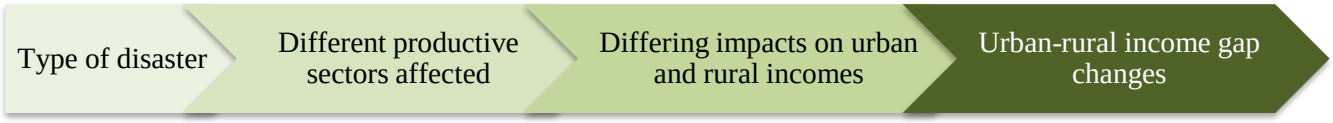


Figure 1. **Hypothesized direction of causality**

Disaster Definitions

The disaster data is from the Emergency Events Database (EM-DAT) from the Center for Research on the Epidemiology of Disasters (CRED). For the purposes of the Emergency Disaster Database defines a disaster is a "situation or event, which overwhelms local capacity, necessitating a request to the national or international level for external assistance" (CRED 2015). The definitions for each disaster type used by the EM-DAT are covered in this section (see Figure 2).

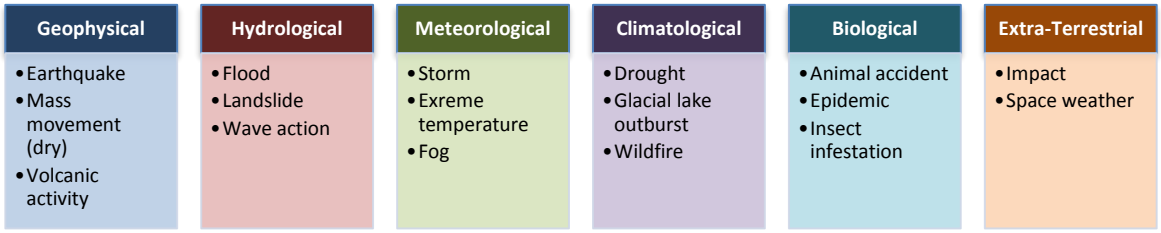


Figure 2. **Disaster Categories as defined by CRED**

Droughts

Drought is defined as "an extended period of unusually low precipitation that produces a shortage of water for people, animals, and plants" (Guha-Sapir, Hoyois and Below 2015, 37). Droughts and famines result in crop and livestock loss, but not damage to infrastructure or buildings. They tend to cover large areas and last multiple years. In fact, the onset of droughts can be difficult to detect. Droughts are often predictable and there are several regional early warning systems in place (Guha-Sapir, Hargitt and Hoyois 2004).

Earthquakes

Earthquakes are a "sudden movement of a block of the Earth's crust along a geological fault and associated ground shaking" (Guha-Sapir, Hoyois and Below 2015, 38). Earthquakes are the least predictable disasters as they strike with minimal or no notice. They also have the highest immediate mortality and structural damage rates, however, they don't affect crops unless landslides are triggered by the earthquake. The risk from earthquakes varies based on the population density, the resistance of buildings and other structures to tremors, the time of the quake (earthquakes that take place when people are sleeping tend to have larger numbers of fatalities), and the intensity of the earthquake (Guha-Sapir, Hargitt and Hoyois 2004). The earthquake sub-group also includes tsunamis.

Floods

Flood is "a general term for the overflow of water from a stream channel onto normally dry land in the floodplain (riverine flooding), higher-than-normal levels along the coast and in lakes or reservoirs (coastal flooding) as well as ponding of water at or near the point where the rain fell (flash floods)" (Guha-Sapir, Hoyois and Below 2015, 38). Floods have the highest ratio of those affected to those killed, meaning while many are affected few are killed. Most of the deaths that do take place are the result of flash floods. The impact on agriculture depends on the timing of the flood. Floods may cover large areas, and can develop slowly or suddenly.

Landslides

Landslide is "the movement of soil or rock controlled by gravity and the speed of the movement usually ranges between slow and rapid, but not very slow" (CRED 2015). While most landslides result from heavy rain or snow or ice melt, dry landslides can happen following earthquakes. Landslides are typically sudden onset disasters.

Storms

Storms include convective storms, extra-tropical storms, and tropical cyclones. Convective storms are "generated by the heating of air and the availability of moist and unstable air masses" and include thunderstorms and tornadoes (CRED 2015). Extra-tropical storms are a "type of low-pressure cyclonic system in the middle and high latitudes (also called mid-latitude cyclone) that primarily gets its energy from the horizontal temperature contrasts (fronts) in the atmosphere" (CRED 2015). When extra-tropical storms take place during winter they can be very damaging (i.e. blizzards). Tropical cyclones are "characterized by a warm-core, non-frontal synoptic-scale cyclone with a low pressure center, spiral rain bands and strong winds" (CRED 2015). They go by various names depending on the region, including hurricane, typhoon, or cyclone.

Windstorms are among the most destructive disasters. They tend to cover large areas and the loss in terms of deaths, injuries, agriculture, and property can be quite large. Mortality is often caused by collapsed buildings while flooding and flying debris account for many injuries (Guha-Sapir, Hargitt and Hoyois 2004).

Volcanoes

Volcanic activity is a "type of volcanic event near an opening/vent in the Earth's surface including volcanic eruptions of lava, ash, hot vapor, gas, and pyroclastic material" (Guha-Sapir, Hoyois and Below 2015, 40). For volcanoes the ratio of people killed to affected is similar to earthquakes. Ash can destroy crops and make it difficult for livestock to find food and water (Guha-Sapir, Hargitt and Hoyois 2004).

Wildfires

Wildfires are defined as "any uncontrolled and non-prescribed combustion or burning of plants in a natural setting such as a forest, grassland, brush land, or tundra which consumes the natural fuels and spreads based on environmental conditions (e.g., wind, topography)" (Guha-Sapir, Hoyois and Below 2015, 40). Wildfires can have natural causes (such as lightening) or may be human caused.

Income Inequality

Income inequality is a measure of the extent to which income is equally distributed. Income inequality is associated with rent-seeking, inefficiency of land utilization, lower savings rates, and an overemphasis on higher education (Todaro and Smith 2009). Todaro and Smith (2009) also speculate that income inequality may lead to self-defeating populist policies as high levels of inequality spur a focus on redistribution of wealth rather than overall economic growth. In addition, high levels of income inequality are linked over the medium-term to lower growth of output (Dabla-Norris, et al. 2015).

The two primary measures of income inequality are the Gini coefficient (Gini 1921) and income share ratios, such as the Kuznets' ratio (Kuznets 1955), that compute the ratio of income pertaining to the upper and lower income percentiles of the population. This research uses a less commonly used measure of inequality, the urban-rural income gap, which focuses specifically on the gap between rural and urban incomes. The income gap is calculated by dividing per capita urban income by per capita rural income. This measure is appropriate for the purposes of this research as it quantifies relative rural and urban incomes which the Gini coefficient and Kuznets' ratio are unable to do. In addition, Young (2013) finds that countries with high levels of overall inequality also have unusually large urban-rural income gaps in living standards.

One concern with the urban-rural income gap, however, is that the estimates may be biased in certain ways. Sicular *et al.* (2007) mention concerns that the imputed rental value of owner-occupied housing is not included as well as the value of public services (such as infrastructure, education, and health care). While including the value of owner-occupied housing and public services increases the gap, the gap decreases when spatial differences in the cost of living are accounted for.

Determinants of Income Inequality

The primary drivers of household income distribution are trade globalization, financial globalization, technical change, macroeconomic policies, labor market policies, wealth inequality, and redistributive fiscal policies such as taxation and transfers (UNDP Bureau for Development Policy 2013). In developing countries, increasing access to education contributes to increasing income shares for the poor and middle class (Dabla-Norris, et al. 2015).

The impact of FDI on inequality is contested among researchers. Proponents of economic liberalization view FDI as an important tool for growth of GDP and subsequent poverty reduction (te Velde 2003), while others see FDI as a means by which industrialized countries extract resources from developing countries and in doing so increase inequality between rich and poor countries. te Velde (2003) finds that FDI brings in new techniques and skills yet also that FDI increases wage differentials in Latin America as a result of increased labor disparity. Growth in FDI leads to an increase in the relative demand for skilled labor in Latin America (Feenstra and Hanson 1997), suggesting that FDI may increase income inequality.

Also contested is the impact of foreign aid on inequality. While some researchers find that foreign aid increases inequality (Herzer and Nunnenkamp 2012), others find no relationship between aid and inequality (Chong, Gradstein and Calderon 2009), or that aid increases inequality in some countries and not others. For example, aid may increase inequality more in democratic countries than autocratic ones (Bjørnskov 2010).

Remittances have been suggested to both increase and decrease inequality, with some suggesting a curvilinear relationship where remittances first increase income inequality in earlier stages when the costs of migration are high and those who migrate are likely to be financially better off (Acosta, et al. 2008). Migration costs tend to decrease over time as migration channels are established, allowing those who are less well off to be able to migrate as well and potentially decreasing income inequality (Koechlin and Leon 2007).

Income Inequality in Latin America

Inequality in Latin America has been falling since the mid-1990s, yet remains high. Most of the decline can be explained by increases in higher education spending, greater foreign direct investment (FDI), and an increase in revenues from taxes (Tsounta and Osueke 2014). Strong GDP growth also appears to have played a role. According to Gasparini *et al.* (2009), increases in international commodities prices decreased overall income inequality in Latin America in the 2000s.

Cornia (2010) investigates whether increased export volumes and improved terms of trade are responsible for declining inequality in Latin America. He points out various mechanisms through which the terms of trade can impact inequality (both positively and negatively), including land and mining rents accruing to owners more than workers, redistribution of tax income by states, and increased availability of foreign exchange. He concludes that the impact of improved terms of trade on reducing inequality in Latin America are moderate.

Cornia (2010) also finds that declining income inequality is related to having a populist or social democratic government, declining educational inequality, a devaluation of the real exchange rate, higher minimum wages, and higher public expenditures. The contribution of remittances by migrants is not significant, while an increase in FDI increases inequality. Educational disparity has the strongest impact on income inequality. Other factors leading to regional inequalities in Latin America include the level of female participation in the labor force, family size, differences in income level by gender, the large informal market, educational discrepancies, occupational status of the head of household, access to public services, and land concentration (Fazio 2005).

There is a general movement of labor in Latin America from high-productivity jobs in manufacturing to lower-productivity jobs in the informal sector or producing commodities (McMillan and Rodrik 2011). A comparison of fiscal redistribution in Western Europe and Latin American finds that the redistributive impact of the fiscal system is comparatively smaller in Latin America compared with the redistributive impact of Western Europe. In addition, when Latin American countries do engage in significant redistribution, they tend to do so through transfers rather than taxes (Goni, Humberto-Lopez and Serven 2011).

Research Question

This research uses the urban-rural income gap to analyze relative changes to rural and urban income in Latin America following natural disasters, i.e. if there is a large disaster that affects industry, does urban income decrease relative to rural income? Or alternatively, if there is an ongoing drought, does rural income decrease relative to urban income? Relative rural and urban incomes are measured using the urban-rural income gap.

Research Question: Does the urban-rural income gap change depending on the type of disaster?

H_{A1}: The urban-rural income gap increases following droughts and wildfires.

H_{A2}: The urban-rural income gap decreases following earthquakes.

H_{A3}: The urban-rural income gap changes following storms, floods, landslides, and volcanic activity (non-directional hypothesis).

Droughts and wildfires are expected to affect rural incomes adversely more than urban incomes and thus increase the income gap, while earthquakes are predicted to affect urban incomes adversely more than rural incomes and thus decrease the income gap. Storms, floods, and landslides may destroy crops, however, they also have the potential to renew depleted soil, and therefore the expected effect on the income gap is unclear. In addition, the impact from volcanic activity is likely to be on whichever community is the closest, regardless of urban or rural.

Methodology

This analysis uses a fixed effects regression analysis with panel data on 18 Latin America countries to test the hypothesis that the urban-rural income gap is affected differently depending on the type of disaster. An equation of the form

$$Y_{it} = \alpha_i + \delta DIS_{it} + \vartheta V^{K}_{it} + \gamma DIS_{it} V^{K}_{it} + \beta X_{it} + u_{it} \quad (1)$$

is modeled, where Y_{it} is the urban-rural income gap for country *i* at time *t*; α_i is an intercept specific to each country; δDIS_{it} is a set of dummy variables indicating the type of disaster; V_{it}^{K} is a term indicating the magnitude of the impact of the disaster where *K* indicates either the number affected, number killed, or damages as a percentage of GDP; $\gamma DIS_{it} V_{it}^{K}$ is the interaction between the type of disaster and the magnitude indicator; βX_{it} is a set of control variables; and u_{it} includes the unobserved country-specific effects, v_i, and the observation-specific error term, e_{it}. The analysis is conducted for the time period from 1980 to 2013. The population consists of all countries in Latin America.

Fixed Effects Estimator

The analysis is conducted using a fixed effects estimator. A fixed effects analysis is chosen over a random effects analysis because of its focus on "the relationship between predictor and outcome variables within an entity" (Torres-Reyna 2010), as opposed to the random effects model where the unobserved effect is believed to be random across the explanatory variables. Because this research is interested in what happens within a country as a result of a natural disaster, the fixed effects estimator with its focus on the relationship within an entity is preferred. The choice of a fixed effects model is confirmed through the use of the Hausman test. A post-estimation test of joint restrictions on the parameters suggests that time fixed effects should be used in addition to country fixed effects. Use of time fixed effects also controls for trends in the data over time.

The regressions are implemented using Driscoll-Kraay standard errors (Driscoll and Kraay 1998). Driscoll-Kraay standard errors are robust to spatial correlation, serial correlation, and heteroskedasticity and perform well with finite samples. When other estimators are used in the presence of cross-sectional dependence (spatial correlation), the standard error estimates are severely downward-biased. Driscoll–Kraay standard errors, on the other hand, "are well calibrated when the regression residuals are cross-sectionally dependent" (Hoechle 2007, 310).

The Driscoll–Kraay standard errors are implemented using the *xtscc* Stata command developed by Hoechle (2007). The *xtscc* estimator is able to handle unbalanced panels with missing data. All the regressions reported here use the default number of lags provided by the software.[1] The *xtscc* command does not allow lagged explanatory variables.

The Data

The dependent variable is the urban-rural income gap, which is a measure of inequality of income in rural and urban areas. The urban-rural income gap is the ratio of urban per capita income to rural per capita income. In the dataset, the income gap value is always above 1, meaning that urban per capita income is always larger than rural per capita income in the sample. An increase in the ratio means that the gap between rural and urban income has become larger.

The urban-rural income for the sample data has declined since 1980, although the decline has not been steady. The gap was at its smallest in 1990 and 1991. After increasing in the early 2000s, the income gap has seen an overall decline since then, indicating rural per capita income is getting closer to urban per capita income.

The explanatory variables include indicator variables for each type of disaster and interaction terms between the type of disaster and the damage as a percentage of GDP (Model 1), the percentage affected (Model 2), or the percentage killed (Model 3). While damages as a percentage of GDP (*percent damages*) is the primary measure of severity for the disasters, analyses are also conducted for the percentage of the population affected, and the percentage of the population killed.

The data on disasters is from the Emergency Events Database (EM-DAT) from the Centre for Research on the Epidemiology of Disasters (CRED). This is the most comprehensive database of natural disasters and is also the one used by the majority of researchers of natural disasters (Cavallo and Noy 2011). In order for a disaster to be included in the database, one or more of the following criteria must be true: ten or more people reported killed, a hundred or more people reported affected, a declaration of a state of emergency, or a call for international assistance (Guha-Sapir, Hoyois and Below 2015).

Control variables used in all models include the first difference in the primary school enrollment rate as well as the growth rates in remittances, FDI, foreign aid, and government expenditures. The growth rates are the change from the previous year divided by lagged GDP (all in current dollars). This provides the growth rate in a variable that is scaled by the size of the economy. The primary school enrollment rate is the first difference. The control variables are chosen because of their potential to impact urban and rural incomes, either through increasing skills levels, creating jobs, changing the wage differential, or providing a transfer of income.

The data for the variables other than the disaster variables are drawn from two sources (see Table 1). Data from the Socio-Economic Database for Latin America and the Caribbean (CEDLAS and the World Bank) is used to calculate the urban-rural income gap. The second source of data is the World Development Indicators (WDI). Remittances, FDI, foreign aid, government expenditures, the primary school enrollment rate, and GDP are all from the WDI.

Table 1: Data description and sources

Variable	Description	Source
Percent rural population	Rural population (% of total population)	WDI[1]
All disaster dummies	Incidence of disaster type	EM-DAT CRED[2]
Urban-rural income gap	Ratio of urban per capita income to rural per capita income	CEDLAS[3]
Remittances	Personal remittances, received	WDI
Foreign aid	Net official development assistance and official aid received	WDI
Government expenditures	General government final consumption expenditure	WDI
FDI	Foreign direct investment, net inflows	WDI

Note: This is an unbalanced panel dataset that covers the time period from 1980-2013 and is comprised of 41 countries.
[1] World Development Indicators
[2] Emergency Events Database from the Centre for Research on the Epidemiology of Disasters
[3] Socio-Economic Database for Latin America and Caribbean (CEDLAS and the World Bank)

The Severity Measures

The severity measures used in this research, and most commonly used by researchers, are the number of deaths, the number of people affected, and the total damages in dollars. There is low correlation among the measures indicating the severity of natural disasters in the EM-DAT (see Table 2). It follows that the outcome of the analysis will also vary based on the measure chosen. For example, Noy (2009) finds that only damages are associated with negative GDP growth and there is no connection between the number affected or the number of deaths and GDP growth.

Table 2: Correlations among the severity measures

	Deaths as a % of total pop	Affected as a % of total pop	Damages as a % of GDP
Deaths as a % of total pop	1		
Affected as a % of total pop	0.2052	1	
Damages as a % of GDP	0.4233	0.3936	1

The choice of which measure to use for analysis varies among researchers. Ebeke and Combes (2013), for example, use the number of persons affected by a disaster as the variable of interest, as they believe that estimates of the number affected are more accurate than estimates of damage. Loayza *et al.* (2012) take a similar approach. Other researchers, however, state that the amount of damages and total deaths are preferred measures over the number affected (Cavallo, Galiani, et al. 2013).

The EM-DAT documentation, in addition, points out the limitations of each measure. Deaths, for example, are often under-reported in the case of drought due to being assigned to other causes such as malnutrition and measles caused by micronutrient deficiency (Guha-Sapir, Hargitt and Hoyois 2004). In addition, deaths are more common with certain types of disasters than others. For example, earthquakes often have a high death toll while the impact of volcanoes is often indirect and not fatal. The number of deaths is more commonly reported than the number of persons affected or the amount of damages, with information being provided on the number of deaths in nearly 90% of disasters.

In two-thirds of disaster reports, the number affected is reported, yet according to documentation supplied by the preparers of the EM-DAT, reports of the number affected tends to be inexact as out-of-date census data may make estimates unreliable or countries may manipulate the number affected for political reasons (Guha-Sapir, Hargitt and Hoyois 2004). The number of persons affected also varies by disaster type, as landslides tend to have a more limited impact than floods or windstorms (Guha-Sapir, Hargitt and Hoyois 2004).

Data on economic losses, however, are not necessarily more reliable, as losses were reported for only 25% of disasters between 2000 and 2003 and rarely exceed more than a third of disasters historically (Guha-Sapir, Hargitt and Hoyois 2004). Economic costs are least likely to be reported for small recurring disasters such as minor droughts, and most likely to be

reported for large disasters, in particular when international aid is requested or needed for insurance valuation. Damages are most likely to be reported for windstorms, followed by earthquakes and floods. In the middle are wildfires, droughts, volcanic eruptions, wave surges, and extreme temperatures, while fewer than 10% of landslides report damages.

Exchange rate shifts can also complicate assessment of economic losses and losses vary greatly between rich and poor countries. Guha-Sapir, Hargitt, and Hoyois (2004) recommend using damages relative to prior year's GDP to standardize losses between rich and poor countries. They also point out that GDP may increase in a disaster year due to investment in reconstruction.

Not all disasters are reported and developing countries in particular may have poorly developed mechanisms for reporting disaster data. More frequent reporting of smaller disasters in recent decades may also have created a time bias in the data. Pinpointing disaster dates can be challenging, as certain types of disasters (droughts for example) may span several months or years. In this case, CRED uses the date recorded by the reporting government.

Countries and Number of Disasters

The sample is the set of countries in Latin America for which a complete set of data is available. The 18 countries included in the analysis are listed in Table 3. The income gap variable is missing for many countries which limits the number of the countries available for analysis.

Table 3: Countries included

Belize	Ecuador	Nicaragua
Bolivia	Guatemala	Peru
Chile	Guyana	Paraguay
Colombia	Honduras	El Salvador
Costa Rica	Jamaica	Uruguay
Dominican Republic	Mexico	Venezuela

The total numbers of disasters included in the models is 243 (see Table 4). Floods are the most common disaster type in both samples. Together, floods and storms make up almost half of the disasters. Wildfires are the least common disaster type, however, volcanoes are a close second.

Table 4: Number of disasters included in the models

Drought	Earthquake	Flood	Landslide	Storm	Volcano	Wildfire
18	27	93	24	54	15	12

Diagnostic Tests

Tests of cross-sectional dependence (spatial correlation), including the Breusch-Pagan statistic for cross-sectional independence and Pesaran's statistic, fail to run due to insufficient common observations across the panels. However, the assumption that the error terms are independent across cross-sections is likely violated as events such as world recessions may cause group-level shocks resulting in correlation in the individual-level fixed effects errors or u_i. The advantage of using estimators with Driscoll-Kraay standard errors is that cross-sectional dependence is automatically controlled for.

Wooldridge's test (2002) for serial correlation in the idiosyncratic errors suggests that serial (temporal) correlation is present in the data. Driscoll-Kraay standard errors are also robust to temporal dependence. The modified Wald statistic for groupwise heteroskedasticity indicates that a robust regression is appropriate. Groupwise heteroskedasticity refers to errors, that while possibly homoskedastic within cross-sections, vary across units (Baum 2001). Driscoll-Kraay standard errors, in addition to being robust to cross-sectional and temporal dependence, are also heteroskedasticity consistent.

The stationarity of the panel data is checked using the Phillips-Perron Fisher-type unit-root test. The Fisher-type tests, such as the Phillips-Perron test, "conduct unit-root tests for each panel individually, and then combine the p-values from these tests to produce an overall test" (StataCorp 2013). The Phillips-Perron unit root tests finds that all variables except for the primary school enrollment rate are stationary. After taking the first difference of the primary school enrollment rate variable it is stationary.

Results

The relationship between disasters and the urban-rural income gap is investigated using three different models. Explanatory variables in Model 1 include disaster indicator dummies for drought, earthquakes, floods, landslides, storms, volcanoes, and wildfires as well as interaction terms between the disaster dummies and damages as a percentage of GDP. It also includes the growth in remittances, FDI, foreign aid, and government expenditures as a percentage of GDP as well as the first difference of the primary school enrollment rate. Model 2 includes the same variables as the first, except that the severity indicator used in the interaction term is the number of persons affected as a percentage of the total population. In Model 3, the severity indicator in the interaction term is the number of deaths as a percentage of the total population.

When interaction terms include a continuous variable, the significance of the term as well as the coefficient varies depending on the value of the continuous variable. Tests of joint significance are used to determine the significance of the variables and the coefficients at different levels of the continuous variable. Because this analysis is primarily interested in the effect of larger disasters, a test of joint significance is conducted using the value of the severity indicator (i.e. percent damages, percent affected, or percent killed) equal to the mean plus one standard deviation. Using the mean for the tests of joint significance instead of the mean plus one standard deviation would have examined only the impact of the "average" disaster. This research, however, is interested in the impact of more extreme events, thus resulting in the choice of the mean plus one standard deviation to account for larger disasters. The results for the tests of joint significance at one standard deviation above the mean for each severity indicator are shown in Table 5.

Table 5: Tests of joint significance at one standard deviation above the mean

	Model 1 **prct damage**[1]		Model 2 **prct affected**[2]		Model 3 **prct deaths**[3]	
Drought	0.451	**	-0.468	**	0.637	
(Std. Err.)	(0.256)	[0.046]	(0.264)	[0.045]	(2.469)	[0.399]
Earthquake	-0.471	***	-0.346	**	-1.095	***
(Std. Err.)	(0.124)	[0.000]	(0.167)	[0.025]	(0.242)	[0.000]
Flood	-0.065		0.026		1.647	
(Std. Err.)	(0.422)	[0.879]	(0.051)	[0.619]	(1.707)	[0.344]
Landslide	-2.066		0.282		-23.799	***
(Std. Err.)	(2.298)	[0.378]	(0.277)	[0.319]	(7.812)	[0.006]
Storm	0.032		0.029		0.037	
(Std. Err.)	(0.050)	[0.531]	(0.051)	[0.574]	(0.042)	[0.380]
Volcano	-0.857		0.042		-4.916	
(Std. Err.)	(0.518)	[0.111]	(0.260)	[0.873]	(4.982)	[0.334]
Wildfire	4.549		-0.290		31.801	***
(Std. Err.)	(3.497)	[0.103]	(0.296)	[0.169]	(4.743)	[0.000]

* p<0.10, ** p<0.05, *** p<0.01

Note: Standard errors are in parentheses and *p* values are in brackets.

[1] Mean value (standard deviation) of damages as a percentage of GDP: 1.377% (9.454)

[2] Mean value (standard deviation) of persons affected as a percentage of the total population: 1.423% (6.725)

[3] Mean value (standard deviation) of deaths as a percentage of the total population: 0.003% (0.061)

Model 1

Model 1 includes each disaster type as an explanatory variable along with the control variables and interaction terms between the type of disaster and damages as a percentage of GDP. The dependent variable is the urban-rural income gap. According to the tests of joint significance shown in Table 5, the coefficient for drought is significant at one standard deviation above the mean (*p* value = 0.046) as well as earthquakes (*p* value = 0.000). The sign is as expected on droughts, as droughts are associated with a greater urban-rural income gap, meaning that urban incomes increase relative to rural in years with droughts. Earthquakes are associated with a decrease in the urban-rural income gap. This is also consistent with expectations, as earthquakes are predicted to decrease the gap between rural and urban incomes due to their greater impact on industry. The coefficient for wildfires (*p* value=0.103) is not statistically significant. See Figure 3 for a graph of the estimates, as well as their

confidence intervals, for the impact of disasters on the income gap at one standard deviation above the mean for damages as a percentage of GDP.

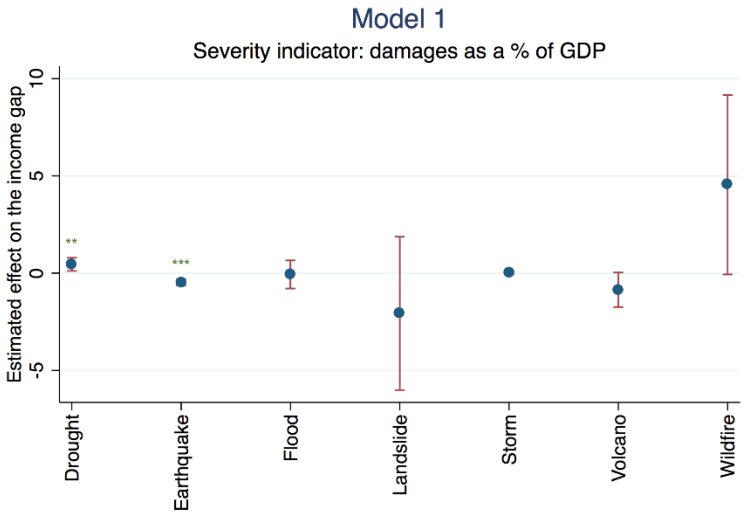


Figure 3. Estimates with 90% confidence intervals for Model 1

Model 2

This model is similar to Model 1 except that it uses the percentage of the population affected as the measure of disaster severity. The tests of joint significance are shown in Table 5 above. See Figure 4 for a graph of the estimates, as well as their confidence intervals, for the impact of disasters on the income gap at one standard deviation above the mean for the percentage of persons affected. The coefficients for droughts (p value = 0.045) and earthquakes are significant (p value = 0.025), however, the sign for droughts is the reverse of what it is for the other two models and counter to expectations.

Noy (2009), upon finding similarly counter-intuitive results when using the percent affected to predict GDP growth, speculates that the impact of the human cost of a disaster on GDP growth may only be visible in long-term growth patterns, while the impact of disaster damages are more immediately visible. This may be the case with the income gap as well. Another possibility is that the percent affected data may be lacking in accuracy. As mentioned earlier, Guha-Sapir, Hargitt, and Hoyois (2004) caution that the data on the number of persons affected may be biased for political reasons or may be based on outdated census information.

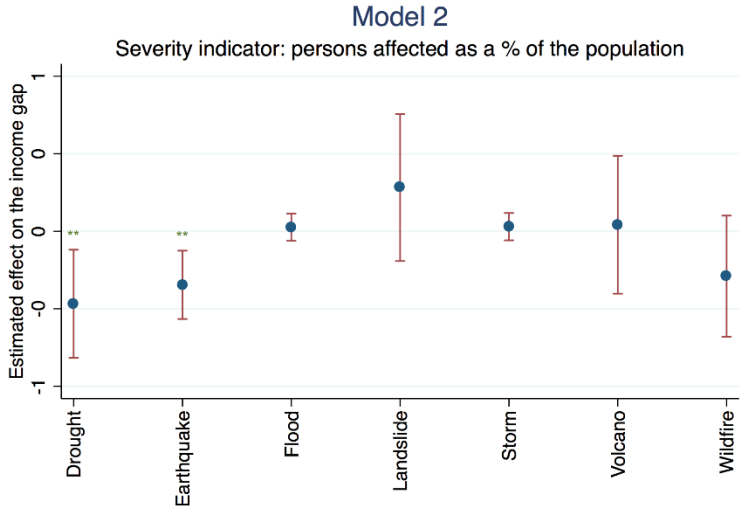


Figure 4. Estimates with 90% confidence intervals for Model 2

Model 3

This model is also similar to Model 1 except that it uses the number of deaths as a percentage of the population as the measure of disaster severity. The tests of joint significance are shown in Table 5 above. See Figure 5 for a graph of the estimates, as well as their confidence intervals, for the impact of disasters on the income gap at one standard deviation above the mean for the percentage of persons killed. The coefficients for earthquakes (*p* value = 0.000), landslides (*p* value = 0.006), and wildfires (*p* value = 0.000) are significantly associated with the urban-rural income gap. The signs of the coefficients for earthquakes and wildfires are as expected, with earthquakes decreasing the urban-rural income gap, indicating that rural incomes are relatively better off when compared to urban, and wildfires conversely increasing the income gap. The signs on earthquakes, landslides, and wildfires are consistent with Model 1.

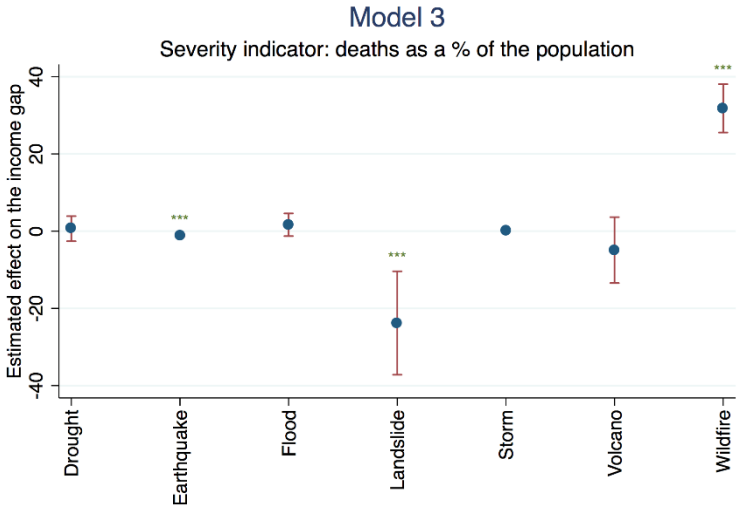


Figure 5. Estimates with 90% confidence intervals for Model 3

Control Variables

The coefficient for the growth in remittances variable is statistically significant across all the models. Remittances, perhaps counter to conventional wisdom, are associated with increases to the urban-rural income gap. A year to year increase in remittances of 1% of GDP increases the income gap by 0.043 to 0.056 points (depending on the model). As remittances tend to flow from urban areas to rural areas it is surprising that they would widen the gap, however, the result is consistent across the models. Foreign aid increases the income gap in all the models, however, the coefficient is not significant in Model 2 and with coefficients of 0.024 to 0.028 the magnitude is less than that of remittances.

The primary school enrollment rate is significantly associated with increases to the urban-rural income gap of approximately 0.03 in all of the models. The coefficient for FDI is only significant in Model 2 where it is -0.028. The sign on the coefficient for FDI is negative in all models, meaning that an increase in FDI decreases the urban-rural income gap. The positive sign on education and the negative sign on FDI are puzzling, as previous studies have found that education decreases inequality and that FDI increases it. The sign on the coefficient for government expenditures in all the models suggests that government expenditures are associated with lower inequality, however, the coefficient is insignificant in these models.

Conclusion

Models 1 and 3 are the most consistent in terms of the signs and significance of the coefficients of the variables. Only one variable, floods, changes sign between the two models and each model has only one variable whose coefficient is not significant in the other model. The difference between the two models shows up primarily in the magnitude of the effect, with all of the disaster types except storms having a stronger effect in the model with the number of deaths as percentage of the population as the severity indicator. This is primarily due to the nature of the severity indicators. The number of deaths from a disaster is typically a smaller percentage of the population than the amount of damages is a percentage of GDP. An increase of deaths by 1% is associated with much more extreme disasters than an increase of damages by 1%, thus the impact on the income gap is larger.

The results of Model 2 are counter to the results of the other models. Earthquakes are the only disaster type to have a significant coefficient with the same sign between all the models. The coefficient on wildfires reverses sign in Model 2. The sign on the coefficient for storms is consistent between the three models but the coefficient is not significant. These results are

in keeping with Noy (2009) who finds that the number affected as a percentage of the population is less effective as a predictor of changes to GDP output. The lack of correspondence between Models 1 and 3 and Model 2 is not surprising given that the correlation between the number of persons affected as a percentage of the population is lower than between the other two measures.

As expected, droughts and wildfires are associated with a decline in the relative position of rural incomes when compared to urban. Also as expected, earthquakes are associated with a decrease in the relative strength of urban incomes when compared to that of rural. There appears to be sufficient evidence to conclude that earthquakes lead to a decrease in the relative strength of urban incomes when compared to rural. There also appears to be some evidence that droughts and wildfires increase the gap between rural and urban incomes, leading to a decline in the relative position of rural incomes when compared to urban.

Acknowledgements

Many thanks to Dr. Shahdad Naghshpour for his guidance and support and to the participants at the annual conference of the Academy of Economics and Finance for their feedback and comments.

Notes

The default lag length, $m(T)$, from Hoechle (2007), is $m(T) = \text{floor}[4(T/100)^{2/9}]$.

References

Acosta, Pablo, Cesar Calderón, Pablo Fajnzylber, and Humberto Lopez. "What is the Impact of International Remittances on Poverty and Inequality in Latin America?" *World Development* 36, no. 1 (2008): 89–114.

Baum, Christopher F. "Residual Diagnostics for Cross-Section Time Series Regression Models." *The Stata Journal* 1, no. 1 (2001): 101-104.

Bjørnskov, Christian. "Do Elites Benefit from Democracy and Foreign Aid in Developing Countries?" *Journal of Development Economics* 92 (2010): 115-124.

Cavallo, Eduardo, and Ilan Noy. "Natural Disasters and the Economy — A Survey." *International Review of Environmental and Resource Economics* 5 (2011): 63–102 .

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. "Catastrophic Natural Disasters and Economic Growth." *The Review of Economics and Statistics* 95, no. 5 (December 2013): 1549-1561.

Chong, Alberto, Mark Gradstein, and Cecilia Calderon. "Can Foreign Aid Reduce Income Inequality and Poverty?" *Public Choice* 140 (2009): 59-84.

Cornia, Giovanni Andrea. "Income Distribution under Latin America's New Left Regimes ." *Journal of Human Development and Capabilities* 11, no. 1 (2010): 85-114.

CEDLAS and The World Bank. "Socio-Economic Database for Latin America and the Caribbean." La Plata, Argentina: Universidad Nacional de La Plata, 2015. http://sedlac.econo.unlp.edu.ar/eng/. Accessed between August and November 2015.

CRED. *Glossary EM-DAT.* 2015. http://www.emdat.be/glossary (accessed April 8, 2015).

Dabla-Norris, Era, Kalpana Kochhar, Frantisek Ricka, Nujin Suphaphiphat, and Evridiki Tsounta. *Causes and Consequences of Income Inequality: A Global Perspective.* IMF Discussion Note, Washington, DC: International Monetary Fund Strategy, Policy, and Review Department, 2015.

Driscoll, John C., and Aart C. Kraay. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80 (1998): 549-560.

Ebeke, Christian, and Jean-Louis Combes. "Do Remittances Dampen the Effect of Natural Disasters on Output Growth Volatility in Developing Countries?" *Applied Economics* 45, no. 16 (2013).

Fazio, Maria Victoria. *Poverty and Inequality in Paraguay: Methodological Issues and a Literature Review.* CEDLAS and The World Bank, 2005.

Feenstra, Robert C., and Gordon Hanson. "Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras." *Journal of International Economics* 42 (1997): 371-393.

Gasparini, Leonardo, Guillermo Cruces, and Leopoldo Tornarolli. *Recent Trends in Income Inequality in Latin America.* Buenos Aires, Argentina: CEDLAS and the World Bank, 2009.

Gini, Corrado. "Measurement of Inequality of Incomes." *Economic Journal* 31 (1921): 124-126.

Goni, Edwin, J. Humberto-Lopez, and Luis Serven. "Fiscal Redistribution and Income Inequality in Latin America." *World Development* 39, no. 9 (2011): 1558-1569.

Guha-Sapir, Debarati, Regina Below, and Philippe Hoyois. "EM-DAT: The CRED/OFDA International Disaster Database." Brussels, Belgium: Université Catholique de Louvain, 2015. www.emdat.be Accessed March 2016.

Guha-Sapir, Debarati, David Hargitt, and Philippe Hoyois. *Thirty Years of Natural Disasters.* Centre for Research on the Epidemiology of Disasters; Presses Universitaires de Louvain, 2004.

Guha-Sapir, Debarati, Philippe Hoyois, and Regina Below. *Annual Disaster Statistical Review 2014: The Numbers and Trends.* Centre for Research on the Epidemiology of Disasters; Presses Universitaires de Louvain, 2015.

Herzer, Dierk, and Peter Nunnenkamp. "The Effect of Foreign Aid on Income Inequality: Evidence from Panel Cointegration." Working Paper, Kiel Institute for the World Economy, 2012.

Hoechle, Daniel. "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence." *The Stata Journal* 7, no. 3 (2007): 281–312.

Koechlin, Valerie, and Gianmarco Leon. "International Remittances and Income Inequality: An Empirical Investigation." *Journal of Economic Policy Reform* 10, no. 2 (2007): 123-141.

Kuznets, Simon. "Economic Growth and Income Inequality." *American Economic Review* March (1955): 1-28.

Loayza, Norman V., Eduardo Olaberría, Jamele Rigolini, and Luc Christiaensen. "Natural Disasters and Growth: Going Beyond the Averages." *World Development* 40, no. 7 (2012): 1317-1336.

McMillan, Margaret, and Dani Rodrik. "Globalization, Structural Change, and Productivity Growth." In *Making Globalization Socially Sustainable*, by Marc Bacetta and Marion Jansen. Geneva: World Trade Organization and International Labour Organization, 2011.

Melillo, Jerry M., Terese (T.C.) Richmond, and Gary W. Yohe. *Climate Change Impacts in the United States: The Third National Climate Assessment.* U.S. Global Change Research Program, 2014.

Mohapatra, Sanket, George Joseph, and Dilip Ratha. "Remittances and Natural Disasters: Ex-Post Response and Contribution to Ex-Ante Preparedness." *Environment, Development and Sustainability*, 14 2012: 365-387.

Noy, Ilan. "The Macroeconomic Consequences of Disasters ." *Journal of Development Economics* 88 (2009): 221-231.

Sicular, Terry, Yue Ximing Renmin, Björn Gustafsson, and Li Shi. "The Urban-Rural Income Gap and Inequality in China." *Review of Income and Wealth* 55, no. 1 (2007).

StataCorp. "xtunitroot — Panel-data unit-root tests." In *Stata Multivariate Statistics Reference Manual: Release 13*. College Station, TX: Stata Press, 2013.

te Velde, Dirk Willem. *Foreign Direct Investment and Income Inequality in Latin America.* London: Overseas Development Institute, 2003.

Todaro, Michael P., and Stephen C. Smith. *Economic Development.* 10th Edition. Reading, MA: Addison Wesley, 2009.

Torres-Reyna, Oscar. *Panel Data Analysis: Fixed & Random Effects using Stata (v. 4.2).* Princeton University, 2010.

Tsounta, Evridiki, and Anayochukwu I. Osueke. *What is Behind Latin America's Declining Income Inequality?* IMF Working Paper WP/14/124, Washington, DC: International Monetary Fund, 2014.

UNDP Bureau for Development Policy. *Humanity Divided: Confronting Inequality in Developing Countries.* United Nations Development Programme, 2013.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: The MIT Press, 2002.

Young, Alwyn. "Inequality, the Urban-Rural Income Gap, and Migration." *The Quarterly Journal of Economics*, 2013: 1727–1785.

World Development Indicators 2015. Washington, DC: World Bank. Accessed between August 2015 and December 2015.

Budgetary Lags in Nigeria

Vincent Nwani, The Lagos Chamber of Commerce & Industry

Abstract

Using methodological Triangulation of desk research and open-ended interviews, this study examined the characteristics of lags and inefficiencies presently experienced in the Nigerian budget process. The paper finds that budget preparation, submission, consideration, approval and signing lags are predominant to an extent that the budget is seldom not available for implementation in the 1st quarter and a better part of the 2nd quarter over the past one and half decades. The solution requires formulation and adoption of a new integrated budgetary model for all stakeholders.

Introduction

For approximately one decade, there appears to be a lingering concern among major players in the Nigerian private sector regarding the uncertainty of budget approval timelines and sub-optimal levels of central budget implementation. It is rare for official budget implementation of the annual budget of the federal government to commenced in January being the start the fiscal year (Ekeocha, 2012).

In addition, the yearly implementation of budget in Nigeria is almost always mired in controversy and poor performance is merely stating the obvious (Akintoye, 2008). Even the actual preparation is not spared either, and unfortunately, this is usually the stage where the foundation for the controversy and low budget implementation is set.

The relative size and significance of the federal government budget has been a key factor for gauging business confidence by domestic and foreign investors in the Nigerian economy (Osafo-Kwako & Soji, 2009). Likewise, the fiscal pronouncements that accompany the federal government annual budget such as debt level, exchange rate, inflation rate, GDP growth expectations, and sectoral spending allocations are far-reaching due to their impact on the business performance of firms and industries (Olomola, 2012).

As provided in the (Nigerian Constitution, 1999) as amended, the central government alone accounts for 54% of the national budget relative to the aggregate budget figure of the 36 States and 744 Local Government Councils in the country. Furthermore, approximately 70% of total spending in the Nigerian economy is tied to the public sector, making it an indispensable business driver for corporate leaders (Abu & Usman, 2010). Tellingly, the size of the national budget grew by 546%, from US$4.7 to US$30.9 billion between 2000 and 2012, compared to private capital flow which grew by 400%, from US$1.64 to US$8.2 billion over the same period (Central Bank of Nigeria, 2012).

Unpredictable budget approval timelines engender uncertainty among economic agents, thereby raising the risk profile of new contracts and investment plans. It makes fiscal planning difficult, and therefore causes inefficient inter-temporal allocation of economic resources across the economy (Thornton , 2008). It is generally accepted that budget approval delay in Nigeria or elsewhere is detrimental to the economy. Yet, it is very difficult to put a figure on the economic cost of budgetary inefficiency due to the scant of empirical research on the subject.

Thus, there seem to be limited empirical research that has examined the impact of budget approval lags on private sector performance across sectors and jurisdictions. Thus, the objective of this paper is to examine the characteristics of inherent lags and inefficiencies presently experienced in the Nigerian budget process.

Theoretical and Empirical Understanding

A deeper look at the fiscal stance of government in any budget year is crucial to make correct assessments on the timeliness of information and approval of fiscal policy stance (Cimadomo, 2008). The standard is that in each fiscal year t−1, executives and legislatives prepare, consider and approve the budget for year t. That is, budgeting laws and practices does not only naturally involve a lag but are also designed on the basis of ex-ante projections on the state of the economy (Cimadomo, 2008). Therefore, the achievement of planned fiscal measures usually documented in the annual budget depends greatly on preparation, consideration approval and implementations timelines.

Khan & Hildreth (2002) noted that if the contents of government budget is relevant to portfolios managers in the private sector for taking deal positions on regular basis then lag in the approval and implementation of budgets may influence their portfolio decision (Khan & Hildreth, 2002). For instance, Cimadomo (2008) gauge the "intentional" stance of fiscal policy in OECD countries based on an information set which closely mimic the one available to governments at the time of fiscal planning (Cimadomo, 2008). His findings suggest that the use of updated observations would point to a pro-cyclical fiscal

policy stance in industrialized countries over the 1994-2006 periods, whereas real-time data indicate that fiscal policy was intended to be counter-cyclical, especially as long as economic expansions are concerned. That is, extended budget approval and implementation lags have a negative consequence on the economy as a whole, the private sector inclusive.

Problems often occur when fiscal policies are placed at the discretion of politicians (Rainey, Backoff, & Levine, 1976). Scholars have identified technical problems associated with discretionary fiscal policies and classify them into three types of lags (Friedman, 1948). The first is the lag between the need for government action and recognition of the need. This lag arises because politicians need economists to analyse the economy; economists need data to conduct the analysis whereas data are post-ante, always behind the reality. On the other hand, policymaking relies heavily on forecasting, which is persistently a "failing" effort (Blinder, 2006). The fact that there is always a lag between data and the reality on ground, interventions and actual results are often associated with lapses. The magnitude of lag depends on the discretionary policy proposal and this lag can be very long or very short.

The second type is the lag between recognition of the need for action and when the action is actually taken. Though, appropriate construction and administration of tax and transfer systems especially with advanced information technology can potentially reduce this lag to a negligible extent. For discretionary actions, the length of the lag depends on the action to be taken. Factors such as shifts in expenditure policy lead to extended lag lengths. Thus, if the policy involves capital construction, it necessarily takes time to get projects under way without causing much efficiency loss.

The third type of lag is between the action and the policy such that in this type of lag, both automatic and discretionary cause of actions are present. Besides, there are political problems with discretionary fiscal policies which are closely related to the second type of lag discussed above. The political aspect of the problems occur due to regular conflicts between the executives and the legislative arms of the government. This is more prevalent when the executive and the legislative branches of the central government are controlled by different political parties in a democratic setting. Heavy influence of interest groups and competition among the federating units for federal resources also may give rise to and fuel the conflicts. This is why there is a lead time between recognition of the need for action and taking policy actions that address the problems. Seidman (2003) cited examples of shorter lags (1975 and 2001), when emergencies prompted quick action by both the President and the US Congress.

The Nigerian private sector relies heavily on the federal government budget as the basis for critical investment and financing decisions. This is hanged on the fact that government spending alone account for about 70% of aggregate spending in the Nigerian economy (Ademola, 2014). Thus, any shock on government fiscal administration exerts significant adverse effect on the private sector. Poor budget implementation seems to have reached such a record level in 2012 that the Federal House of Representatives threatened to impeach the sitting President (Dr. Goodluck Ebele Jonathan) for consistently breaching the Budget Appropriation Act. A major fall out between the then president and the Federal House of Representatives stem from the revelation that as at the end of third quarter in 2012, only 19% of the fund appropriated capital budget had been released (Federal House of Representative, 2012). No doubt, budget administration crisis is partly responsible for the steady slowdown in business momentum, competitiveness and the country's steady drop in the World Bank Doing Business rankings (World Bank, 2012). Combined with shortcomings in infrastructural development and poor structural reforms, inefficient administration of Nigeria's central budget is increasingly inhibiting productivity (Fiscal Responsibility Commission, 2011).

Research Approach

After exploring the literature, it appears that there is limited empirical research that examined the impact of budget approval lags across the sectors in any jurisdiction. Thus, adopting or modifying an existing methodology for this study is somewhat impractical. Therefore, this study adopted methodological triangulation which involves the use of multiple qualitative and quantitative research frameworks as follows:

i. Step-One: Desk Research: Leveraging on the literature, data from the Nigerian Budget Office (NBO) and Fiscal Responsibility Commission (FRC), this research demonstrates that several lags exist and that the country is "neck deep" in budgetary crisis over the years.
ii. Step-Two: Interviews: To validate the findings from table research and allow for robustness, this research sampled the Organised Private Sector (OPS) in Nigeria on a face-to-face, open-ended interview basis. The institutions include the National Chamber of Commerce, selected state chamber of commerce and selected trade and professional associations covering 70% activities across all the sectors respectively. All the sampled institutions gave initial consent to participate in a paper interview and the respondents are top executives occupying positions that are relevant to the subject.

The benefit of methodological triangulation approach is that results from different sources of data such as desk research, questionnaires, interviews and discussions are compared to see if the findings from the various approaches are consistent or identical. If the conclusions from each of the methods are the same or closely related, then validity is established (Jick, 1979).

Face-to-face open ended interview covered ten Organised Private Sector (OPS) institutions with the largest coverage in Nigeria. The institutions include the four top regional chamber of commerce, four leading trade and professional associations, one bi-lateral chambers of commerce and the national umbrella for industry associations. The respondents were given a code name and listed numerically as OPS-1 to 10 and then characterized within their respective profile as follows:

- Age and type of the institution;
- Title and role of the respondent in the institution, and;
- Length of time respondent has spent on the current function.

The research ensured that all the participants are in senior positions of authority that confers them oversight responsibility and information the member group they represent. This also includes the number of years the respondents have spent on the current position. Table 1 presents the summary of participants in the open-ended interview survey.

Table1: Interview Participant Profile

	Org. Age	Org Type	No. of Members	Position	Function	No. of years in Position
OPS1	128	City Chamber	1,750	DG	Management	7
OPS2	22	City Chamber	800	CE	Research	8
OPS3	35	City Chamber	600	Director	Advocacy	7
OPS4	24	City Chamber	480	Snr. Mgr.	Trade promotion	10
OPS5	28	Trade Association	100	Director	Trade Missions	5
OPS6	16	Trade Association	2,000	Director	Planning	3
OPS7	31	Trade Association	150	Director	Govt. Relation	4
OPS8	30	Trade Association	150	President	Management	2
OPS9	41	National Chamber	-	Director	Research	6
OPS10	10	Bi-lateral Chamber	-	Director	Trade promotion	4

LEGEND

1. Organised Private Sector = OPS
2. Director General = DG
3. Senior Manager =Snr Mgr
4. Chief Economist = CE
5. Age of the organization = Org. Age
6. Type of the organization = Org Type

Source: Nwani (2015)

The interview feedback is analyzed using created using Mindjet MindManager (Mindjet MindManager, 2011) software. Findings from the desk research and interviews are triangulated to find comparison and determine consistency. The mixing of data types from different methodological approaches is often thought to help in validating the claims that might arise from a none existing or understudied areas such as the present research (Jick, 1979).

Results and Explanation

This paper attempts to answer the following research question: "What are the characteristics of inherent lags and inefficiencies presently experienced in the Nigerian budget process?".

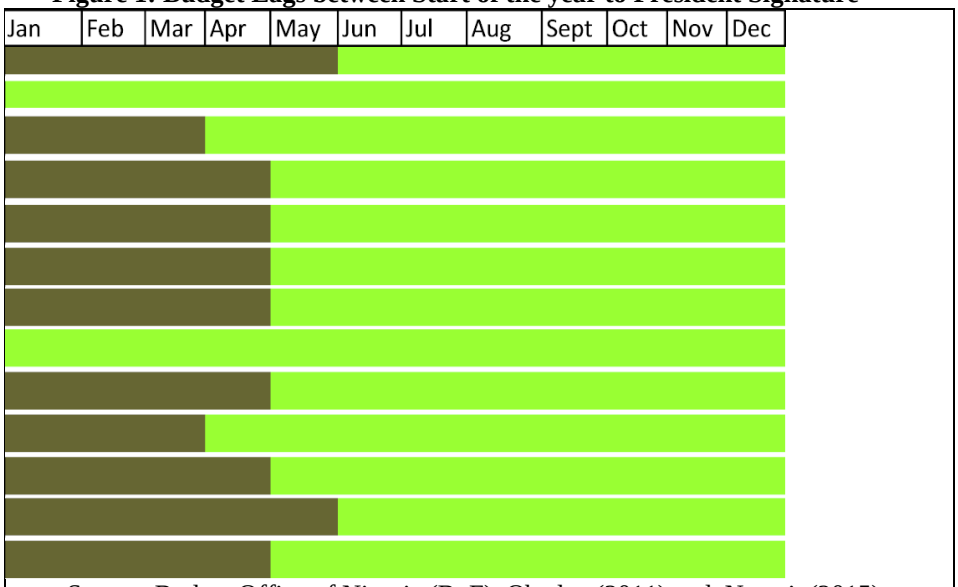
Findings from data collated from secondary (literature and databank of authorities) sources shows that Nigeria is "neck deep" in various forms of lags in the budgetary process. The research identified the existence of five budgetary lags which include:

1. Approval lag (No budget Period): Time lag between start of the fiscal year to President assent (signing) of the Annual Appropriation Bill = 4months 7days;
2. Submission lags: Time lag between start of fiscal year and submission of draft budget by the executive to the legislature = 1 month 7days (International benchmark is minimum of 3 months with legal backing);
3. Consideration lag: Time lag between submission of draft budget to legislators by the executive and legislative approval of the budget = 3 months 18 days;
4. Signing lag: Time lag between legislative approval of the budget by the legislature and signing of the approved budget by the President = 19 days, and;
5. Implementation lag: Average quarterly capital budget implementation levels = Q1-1%, Q2-8%, Q3-22% & Q4-41% (25% is the ideal quarterly implementation threshold).

Existence of Budget Preparation and Approval Lags

Figure 1 shows that over the period 2000 and 2012, the Nigerian economy operated without central budget for the first three and half months on the average. Years 2000 and 2011 witnessed the longest time lag of five months and two weeks. This leaves the civil servants and other stakeholders with shorter time to implement the approved budget. The Federal Responsibility Act (2007), provides that fiscal year in Nigeria commences on 1st January and end on December 31st each year. The president is required to sign the budget as approved by the legislature before it can become operational. Yet, hardly has the operation of Federal Government (FG) budget commenced on 1st January.

Figure 1: Budget Lags between Start of the year to President Signature



Source: Budget Office of Nigeria (BoF), Obadan (2011) and Nwani, (2015)

Lag between Budget Submission to Legislators and President's Signature

Table 2 shows that the average time lag between submission of the draft budget to the legislators and President's signature over the period, 2000 to 2013 is four and half months. The years 2000, 2005, 2008 and 2011 respectively were more telling with approximately six months between the budget presentation to the legislators and the president's assent. In the same vein, 2002, 2003, 2004, 2010, and 2013 took about four months before approval. Only in 2001, 2006, 2007, 2009 and 2013 respectively, has the budgetary process taken relatively lesser time of one to two months. A key omission in the Nigerian fiscal appropriation law, the Fiscal Responsibility Act (2007) is the absence of specific timelines for submission, consideration, approval and signing of the budget. Amendment of the existing law to incorporate critical timelines for budgeting is one of the identified essentials for addressing budgetary inefficiency.

Lag between Start of Fiscal Year and Submission of Draft Budget to the Legislature

Table 3A shows arbitrary timelines of submitting draft budgets by the executive to the legislature. The average timeline for submitting draft budget over the period 2000 to 2013 to the legislature for consideration is one month, 21 days from the start of the fiscal year. Interview respondents considered the period between draft budget submission and start of fiscal year to be too short for a robust consideration of the draft budget by legislature. It has been reported that legislatures have often found that they need ample time and technical help to play credible roles in budgeting (Posner & Chung-Keun, 2007). The legislature can be overwhelmed by the sheer magnitude of budget documents, their technical complexity, and the years of expertise possessed by executive budget offices. Accordingly, legislatures in some countries have chosen to invest in separate legislative offices that specialise in budgetary reviews, finding that independent units can put the legislature on a more equal footing with the executive (Anderson, 2005). Thus, another element of a new budgetary model is the establishment of legislative budget office in Nigeria to complement the work of executive budget office. This have become the global best budgetary practice.

Table 2: Lag between Budget Submission to Legislators and President's Signature

Years	Date of Budget Submission by Executive to legislature	Date President Signed the Budget into Law	Time Lag between Budget Submission to Legislators and President Signature
2000	24th Nov., 1999	5th May, 2000	5 months 11days
2001	9th Nov., 2000	21st Dec., 2000	1 month, 12 days
2002	7th Nov., 2001	28th March, 2002	4 months, 21days
2003	20th Nov., 2002	10th April, 2003	4 months, 21days
2004	18th Dec., 2003	21st April, 2004	4 months, 3 days
2005	12th Oct., 2004	12th April, 2005	6 months
2006	6th Dec., 2005	22nd April, 2006	2 months, 16 days
2007	6th Oct., 2006	22nd Dec., 2006	2 months, 12 days
2008	8th Nov., 2007	14th April,2008	5 months, 7 days.
2009	2nd Dec., 2008	10th March, 2009	3 months, 8 days
2010	23rd Nov., 2009	22nd April, 2010	4 months, 29 days
2011	15th Dec, 2010	26th May, 2011	5 months, 11 days
2012	13th Dec, 2011	13th April, 2012	4 months
2013	12th Oct., 2012	26th Jan, 2012	4 months, 14 days

Source: Nigerian Economic Summit Group {(NESG), 2013); Nwani (2015)

Table 3A: Budget Submission Lags in Nigeria (2000-2013)

Start of Fiscal Years	Date of Budget Submission by Executive to legislature	Lag between Submission date to Start of Fiscal Year
1st Jan. 2000	24th Nov., 1999	1 month, 6 days
1st Jan 2001	9th Nov., 2000	1 month, 21 days
1st Jan 2002	7th Nov., 2001	1 month, 23 days
1st Jan 2003	20th Nov., 2002	1 month, 10 days
1st Jan 2004	18th Dec., 2003	13 days
1st Jan 2005	12th Oct., 2004	2 months, 18 days
1st Jan 2006	6th Dec., 2005	25 days
1st Jan 2007	6th Oct., 2006	2 months, 24 days
1st Jan 2008	8th Nov., 2007	1 month, 22 days
1st Jan 2009	2nd Dec., 2008	29 days
1st Jan 2010	23rd Nov., 2009	1 month, 7 days
1st Jan 2011	15th Dec, 2010	16 days
1st Jan 2012	13th Dec, 2011	18 Days
1st Jan 2013	12th Oct., 2012	2 months, 18 days

Source: Nigerian Economic Summit Group {(NESG), 2013); Nwani (2015)

Table 3B gives highlight of official timetable for budget submission to the legislature across selected countries. Minimum timeline between draft budget submission to legislators and start of fiscal year for most countries is three to four months. US official submission timeline is spectacular with eight months (submission date of 1st Monday in February) as specified in the Budget and Accounting Act of 1921. Unlike Nigeria, official draft budget submission timeline is clearly specified in the budget laws and constitutions of the countries. This omission is a key legal and policy action point suggested by this research for budgetary reform in Nigeria.

Lag between Legislative Approval and President's Signature

Table 4 shows the time lag between legislative approval and the signing of the approved budget by the president. There is ample lag (18 to 60 days) between budget approval date by the legislature and the day it is signed into law by the president over the years. This is with exception to years 2001, 2002, 2004 and 2007 when the budget was signed by the president the same day it was approved by the legislature. Enquiry by the research shows that advocacy through mass media campaign by the OPS and electioneering campaign are partly the factors responsible for "real time" signing of the budget by the president for these years (2001, 2002, 2004).

Table 3B: Official Timetable for Budget Preparation – Nigeria Vs Other Countries

Countries	Start of Fiscal Year	Date of official Budget Submission to legislature	Lag between Submission date to Start of Fiscal Year	Source
France	1 January	First Tuesday of October	About 3 months	Organic budget law (Loi Organique relative aux Lois de Finances, LOLF)
Japan	1 April	During January	More than 3 months	Law (Public Finance Act)
Korea	1 January	2 October	About 3 months	Constitution
Mexico	1 January	8 September	About 4 months	Law
Netherlands	1 January	Third Tuesday of September	More than 3 months	Constitution
Sweden	1 January	Not later than 20 September	More than 3 months	Law (Parliament Act)
United Kingdom	1 April	None (at the discretion of HM Treasury)	–	None
United States	1 October	First Monday in February	About 8 months	Law (Budget and Accounting Act)
*Nigeria	1 January	None (at Executive discretion)	None	

Source: (Posner & Chung-Keun, 2007); Nwani (2015)

Table 4: Budget Signing Lags in Nigeria (2000-2013)

Start of Fiscal Years	Date Budget is Approved by the Legislature	Date President Signed the Budget into Law	Lag between Legislative Approval and President Signature
1st Jan. 2000	14th April, 2000	5th May, 2000	19 days
1st Jan 2001	21st Dec., 2000	21st Dec., 2000	Same day
1st Jan 2002	28th March, 2002	28th March, 2002	Same day
1st Jan 2003	11th March, 2003	10th April, 2003	30 days
1st Jan 2004	21st April, 2004	21st April, 2004	Same day
1st Jan 2005	18th March, 2005	12th April, 2005	25 days
1st Jan 2006	21st Feb., 2006	22nd April, 2006	60 days
1st Jan 2007	22nd Dec., 2006	22nd Dec., 2006	Same day
1st Jan 2008	27th March, 2008	14th April,2008	18 days
1st Jan 2009	3rd Feb., 2009	10th March, 2009	35 days
1st Jan 2010	25th March, 2010	22nd April, 2010	28 days
1st Jan 2011	25th May, 2011	26th May, 2011	1 days
1st Jan 2012	15th March, 2012	13th April, 2012	29 days
1st Jan 2013	22nd Dec 2012	26th Jan, 2012	35 days

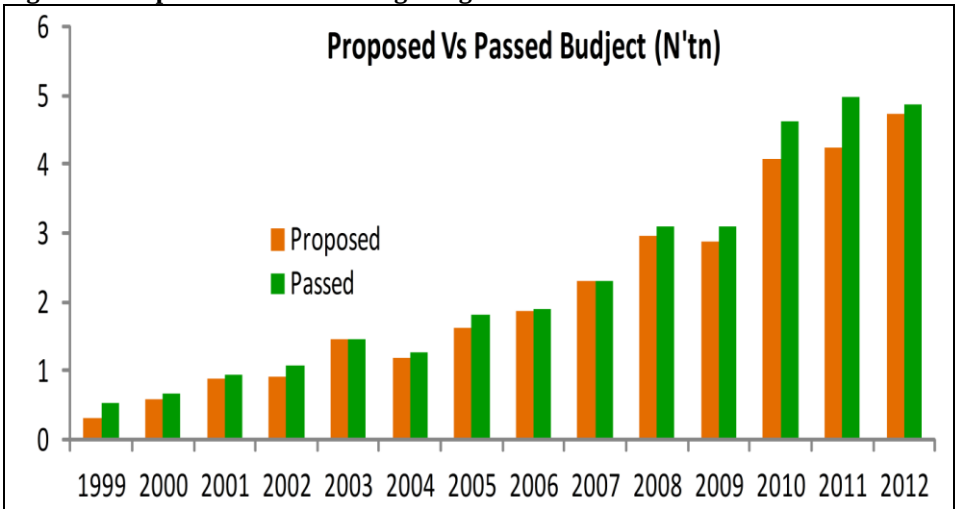
Source: Nigerian Economic Summit Group (NESG), 2013; *Nwani (2015)*

Feedbacks from OPS interviews reveal that the executive arm of the government in Nigeria has persistently accused the legislators of significantly altering the content of the draft budget and making it impossible to implement. Over the period between 1999 and 2012, the legislators have consistently increased and approved a higher budget figures opposed to the draft estimate submitted by the executive (see Figure 4.2). This is largely responsible for sustained imbroglio between the arms of government and persistent delay of president's assent of the budget. This research observed that the enabling laws in Nigeria are surprisingly silent on the extent to which the legislature can alter or tamper with the content of the draft budget.

Budget Adjustment and the Role of Legislature

Notwithstanding the potential scope of legislative authority with regard to the extent of legislative modifications to executive proposals appears to be quite modest when examining just the formal record of actions taken in response to executive proposals (Posner & Chung-Keun, 2007). For instance, Posner & Chung-Keun showed that in Korea, legislative amendments constituted 3.8% of the executive budget proposal, while Japan reports no legislative amendments changing budgetary allocations in recent years. Mexico, on the other hand, reports that recent legislative amendments to the president's budget constituted 26% of the total.

Figure 2: Proposed Vs Passed Budget Figures



Source: MarkMonitor (2013)

As Richard Neustadt said, presidential systems are best characterised not as separation of powers but as separated institutions sharing powers (Neustadt , 1960). Each actor has powerful leverage: the legislature can refuse to approve the president's priorities while the president can veto the appropriations. Accordingly, both sides have strong incentives to engage in bargaining and negotiations to avoid a confrontation that can disrupt programmes and operations and cause public disaffection. Legislative officials often exercise informal influence over executive budgets before they are formally announced. Executive officials have incentives to anticipate and take into account potential legislative views and reactions as they develop their proposals, either through informal negotiation or through anticipatory behaviour (Meyers, 1999).

In cases when the president does veto appropriations, considerable disruption can occur in the operations of MDAs, as questions can be raised about whether government agencies must shut down. In 2005, the lack of adequate institutional mechanisms and legal provisions resulted to a constitutional crisis as the Mexican president vetoed the budget that had been approved by the Congress (controlled by the opposition party) (Posner & Chung-Keun, 2007). When the Mexican president used his veto power in 2005, the House of Representatives did not recognise it and a constitutional controversy took place in the country. This left the government without a budget for ten months until the Supreme Court ruled that the veto by the president was constitutional.

However, most countries have established provisional budget authority to cover instances when there are delays in legislative approval beyond the legal or conventional deadlines (Lienert, 2005). For example, in France and Korea, the provision guarantees that the executive can spend on the basis of the previous year's approved budget. Yet, considerable ambiguity continues to exist about the scope of such provisional budget authority. In the United States, there is no standing provisional authority to spend public funds in the absence of an appropriation signed by the president (Posner & Chung-Keun, 2007).

Accordingly, Congress and the president must agree on a continuing resolution to sustain operations pending enactment of full appropriations, or the government will have to shut down. This in fact, occurred during the Clinton presidency when the Republican Congress failed to enact appropriations that the Democratic president could sign (Posner & Chung-Keun, 2007). The government shut down for nearly three weeks until both sides were able to reach agreement. A similar scenario also played out in 2014 between President Obama (Democratic Party) and Republican Congress. In Nigeria, there is no record in the literature where the president or the legislature has vetoed appropriation bill but the two arms of government are constantly on the "border-line" to override each other.

In Nigeria, part of the flexibility recently introduced through mutual understanding is that the executive can spend on the basis of the previous year's approved budget provisions up to March of current fiscal year. This provision covers only recurrent expenditure especially salaries and wages leaving out the capital aspect of the budget which is the main driver of the private sector. The present research noted that the provisional spending provision is yet to be backed by any legislation. Scope of legislative work on the budget and the limit of its alteration of draft budget content are also missing in the existing appropriation law.

Lessons for Policy

This research is guided by the fact that budget is one of the most critical parts of national and management control tools for promoting efficient use of resources and providing support for other critical functions of any nation. Its role in an economy cannot be overemphasized as a formal instrument of the government (Faleti, Faleti, & Ojeleke, 2014). Therefore, any empirical solution directed at strengthening an ailing budgetary system is useful.

This paper finds that budget preparation, submission, consideration, approval and signing lags are predominant in the Nigeria's central budget administration to an extent that the budget is seldom not available for implementation in the 1st quarter and a better part of the 2nd quarter over the past one and half decades.

Table 5: Perspectives for Prompt Approval and Optimal Implementation of FG Budget

Perspectives for Prompt Approval of Budget	Perspectives for Optimal Implementation of Budget
1. Amend the Fiscal Responsibility Act in line with global best budget practices	1. Straighten capacity of key agencies on management of public projects
2. Set legal timelines along with milestones for all budgetary processes	2. Streaming processes of releasing funds for government projects
3. Improved collaboration between the executive and legislators	3. Pass a bill that specifically deal with monitoring budget performance
4. Better quality of project concept and budget preparation by the agencies	4. Add successful projects oversight as one of the performance indicators of the Legislators
5. Diversify the economy away from oil as the mainstay	5. Amend the FRC Act to reform its functions and operation
6. Early release key indicators in the budget draft by Nigeria's Budget Office	6. Institutes measure that eliminate/reduces corruption in the system

Source: Nwani (2015)

Overall budget approval delay hinders the ability of businesses to effectively plan and operate within the context of government's fiscal policy actions from the start of the year. This is more because of Nigeria's peculiar economic structure of import dependency has amplified the imperative of early fiscal pronouncements that accompany the annual budget. Budget delay is adversely infectious to the private sector in an economy such as Nigeria where government is the lead spender of up to 70% of aggregate annual spending (World Bank, 2013). Table 5 highlights the perspectives of top economic stakeholders for prompt approval and optimal implementation of FG budget in Nigeria.

References

Budget Office of Nigeria. 2012. Various Years (2000- 2012). *Budget Office of the Federation Presidential Budget Speech*..

NESG. 2013. *The Nigeria Economic Summit Group Brief*.

Abu, N., & Usman, A. 2010. Government Expenditure And Economic Growth In Nigeria, 1970-2008: A Disaggregated Analysis. *Business and Economics Journal*.

Ademola, F. 2014. Economic Review and Projections. BGL Financial Monitor (January Issue, 2014).

Akintoye, I. R. 2008. Budget and Budgetary Control for Improved Performance. *European Journals of Economics, Finance and Administrative Science*(8), 9-11.

Anderson, B. 2005. "The Value of a Nonpartisan, Independent, Objective Analytic Unit to the Legislative Role in Budget Preparation". *Oversight,* 131.

Blinder, A. 2006. The case against discretionary fiscal policy. In R. W. Kopche, G. Tootell, & R. Triest (Eds.), *The macroeconomics of Fiscal policy.* Cambridge, MA: The MIT Press.

Central Bank of Nigeria. 2012. *Business Expectation Survey, Q1-2012.* CBN.

Cimadomo, J. 2008. Fiscal Policy In Real Time. *European Central Bank Working Paper Series No 919*.

Ekeocha, P. C. 2012. *An Analysis of the Federal Budgeting Process in Nigeria: Implications for Institutional Reforms for Achieving Timeliness.* National Institute for Legislative Studies (NILS). National Assembly, Abuja.

Faleti, K. O., Faleti, H. O., & Ojeleke, R. O. 2014. Budgetary and Management Control System for Improved Efficiency in Public Sector: The Implications of "Babariga–Style". *Budgeting Approach Scholarly Journal of Business Administration, 4*(2), 44-52.

Federal House of Representative. 2012. *Budget and Appropriation Report.* Nigeria Federal House of Representative.

Fiscal Responsibility Commission. 2011. *Federal Budget Performance Review.* Fiscal Responsibility Commission.

Federal Responsibility Act.. 2007. The Nigerian Fiscal Responsibility Act. Nigeria.

Friedman, M. 1948. A monetary and fiscal framework for economic stability. *The American Economics Review, 38*(3), 245–264.

Jick, T. D. 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly, 24*(4), 602-611.

Khan, A., & Hildreth, W. B. 2002. *Budget theory in the public sector.* California: Quorum Books, Greenwood Publishing Group, Inc.

Lienert, I. 2005. *"Who Controls the Budget: The Legislature or the Executive?", .* Washington DC: IMF Working Paper No. 05/115, June, International Monetary Fund .
MarkMonitor, L. 2013. Federal Budget and the Private Sector. *FinIntel Magazine., 1*(8), March 2013, pp. 34-35.
Meyers, R. T. 1999. *Handbook of government budgeting.* Jossey-Bass.
Mindjet MindManager. 2011. *MindManager: Version 9 for Mac. San Francisco.*
Neustadt, R. E. 1960. *Presidential power.* New American Library.
Nigerian Constitution. 1999. Constitution, Nigerian Constitution of the Federal Republic of Nigeria." Abuja, Federal Ministry of Information and National Orientation.
Obadan, Michael. I. 2011. Africa's recent growth performance and development challenges. *CODESRIA bulletin, 3*(4), pp. 14-19.
Olomola, A. S. 2012. State Budgetary Allocations: An Appraisal Of Budget Implementation And Effects In Nigeria. *Nigerian Institute of Social and Economic Research seminar series (NRSS), 11*. Ibadan.
Osafo-Kwako, P., & Soji, A. 2009. *Nigeria: Country Assessment Research Background Paper on the Political Economy of the Budget Process.* Centre for the Study of Economies of Africa (CSEA).
Posner, P. L., & Chung-Keun, P. 2007. Role of the Legislature in the Budget Process: recent trends and innovations. *OECD Journal on Budgeting, 77*(3), 1-26.
Rainey, H. G., Backoff, R. W., & Levine, C. H. 1976. Comparing Public and Private Organizations. *Public Administration Review, 36*(2), 233-244.
Seidman, L. S. 2003. *Automatic fiscal policies to combat recessions.* New York: M.E. Sharp.
Thornton, J. 2008. Inflation and Inflation Uncertainty in Argentina, 1810-2005. *Economic Letters*(98), 247-252.
World Bank. 2012. *Ease of Doing Business Report.*
World Bank. 2013. World Development Report.

Appendix

Questionnaire (Open-ended Interview for BMOs and Trade Associations)
PART A- Respondent's Profile:
Name of the BMO/Trade Association: ……
What is your Role/Position in your organisation: ……..
How Long have your organisation been in Existence: ……
How many years have been in your organisation: …….
PART B- Questionnaire Continued
Please provide your full opinion for each of the following questions:

1. Generally, what impacts do FG budget approval delay and poor implementation have on the economy?

Impact of FG budget approval delay on the private sector	*Impact of poor implementation of FG budget on private sector*

4. List the aspects of FG budget that your members mostly care about
5. What factors do you think are responsible for Delays of FG budget and poor implementation in Nigeria?
6. What are your suggestions for prompt approval and adequate implementation of FG budget in Nigeria?

Suggestions for prompt budget approval	*Suggestions for adequate implementation of the budget*

Source: Nwani (2005)

Managerial Decisions and Player Impact on the Difference between Actual and Expected Wins in Major League Baseball

Rodney J. Paul, Greg Ackerman, Matt Filippi, and Jeremy Losak, Syracuse University

Abstract

In the game of baseball, the difference between runs scored and runs allowed generally translates directly into wins and losses. A mathematical relationship call Pythagorean Win Percentage maps the difference between runs scored and runs allowed into an expected win percentage for Major League Baseball teams. With fierce competition for playoff spots, outperforming or underperforming a team's expected winning percentage could make the difference in earning a playoff spot. This research examines potential determinants at the margin, including relief pitching, bench performance, and managerial decision-making, as it pertains to outperforming or underperforming a team's expected win percentage for a season.

Introduction

In competition, in sports, business, and otherwise, small advantages that one party may be able to achieve could lead to large returns. The sport of baseball has undergone a statistical revolution through most of its' teams' front offices that has led to deep empirical research on players and game outcomes. This revolution was highlighted by the Michael Lewis book, Moneyball, and the movie of the same name where market inefficiencies were exploited to allow for a small market team to compete with big market teams in a sport where a salary cap does not exist.

One angle of General Manager of the Oakland A's, Billy Beane, that was illustrated in Moneyball was the formation of a roster that generated a certain number of expected runs scored and runs allowed. By creating a roster that could generate more expected runs than the pitchers and defense would be expected to allow would generate an anticipated winning record and a higher probability of a playoff appearance. This, in turn, would give the team a chance at a championship.

The conversion of runs scored and runs allowed into an expected win percentage has been known in sports for a long time. Early pioneers such as Bill James transformed these on-field statistics into winning percentages based upon historical data in a method called Pythagorean Wins. Although amended over time, this calculation remains logically the same today, with the only key difference being the exact magnitude of the exponent used in the calculation. We use the calculations of Pythagorean win percentage and actual win percentage from www.baseball-reference.com.

Given that a particular run differential (runs scored – runs allowed) converts into an expected win percentage and this run differential yielded an actual win percentage in a season, it is possible to study the difference between actual and expected wins to ascertain if any particular variables lead to a team outperforming or underperforming its run differential. If teams construct a roster, given certain budget constraints, to attempt to maximize the difference between runs scored and runs allowed, there are still possibilities that even if correct, the team could miss out on the playoffs if they underperform their run differential or a team that achieves a smaller run differential could still make the playoffs if their true record is better than their run differential would suggest.

The goal of this research is to attempt to identify factors which help to explain how and why a team outperforms or underperforms their run differential. By using a series of regression models with the difference between actual and expected win percentage as the dependent variable, we investigate player and managerial performance at the margin which may help to decipher why some teams win or lose more games than they "should". To model this relationship we choose to investigate players who may have large impacts at the margin in close games and how the decision-making of managers may influence team performance compared to run differential.

Teams that outperform or underperform their expected wins may have this result due to their record in close games. These close games may be decided at the end of a baseball game by relief pitchers and hitters coming off the bench. Assuming that the bulk of runs scored and runs allowed come from the players who are on the field the most often, these relievers and bench players may influence games at the margin and lead to greater or fewer actual wins compared the team's Pythagorean wins expectation.

In addition to bench players and relievers, we investigate if managerial decision-making plays a role in team performance that differs from expectations generated by run differential. If managers employ certain strategies or display a certain type of behavior that influences wins and losses beyond the sum of runs scored and runs allowed over the course of the season, it may be possible to identify these factors and these strategies could be adopted to allow for the greatest likelihood of maximizing

wins given a specific level of talent on a roster. The Bill James Handbook includes statistics for each manager in Major League Baseball for each season. We use these statistics as independent variables to test if any of these strategies and decisions have statistical significance, and in what direction, for explaining the difference between actual and expected wins in a season.

When considering the role of a manager in Major League Baseball as it relates to the success of the team, a few key factors come into play. One is the role of making the best decisions at the optimal time. This could come about due to the use optimal use of a pitching staff, bench player management, and proper timing when it comes to risk-taking (taking an extra base, stealing a base, etc.) or playing it safe (issuing intentional walks, playing the outfield deep or along the lines to guard against extra-base hits, etc.).

The role of risk in managerial decision-making has been studied in the management, economics, and finance literature on multiple fronts. March and Shapira (1987) investigated the varied perspectives on risk and risk taking from the perspective of managers. In their study, they found that decisions by managers are primarily impacted by their attention to performance targets and that manager's make a clear distinction between the act of taking risks and that of gambling. How a manager is impacted by being risk-averse has been studied in various finance studies including the choices made by managers when it comes to financing projects (Lewellen, 2006) and the distortions that occur in the marketplace when managers who are risk-averse make the decision to undertake risky projects (Parrino, et al., 2005). The psychological perspective on risk-taking and how it impacts an organization was studied by Kahneman and Lovallo (1993) who delved into the rationale behind overly cautious and overly optimistic decision-makers. The role of loss aversion, being more sensitive to losses as opposed to gains (wins) and its impact on risk taking was investigated by Thaler, et al. (1997) who found a relationship between frequency of feedback and level of risk-taking (and returns). All of these studies may relate to baseball managers as they are in a highly competitive position and typically take a large portion of the blame when a team performs poorly.

This paper is arranged as follows. The next section investigates the role of player performance at the margin and how it impacts the difference between actual and expected win percentage. The variables used to measure player performance focuses on relief pitching and bench performance. The following section adds the managerial performance data to the model and investigates what managerial decisions, if any, leads to teams outperforming or underperforming their run differential for the season. The final section discusses the implications of the results and concludes the paper.

The Role of Relief Pitching and Bench Performance on Actual vs. Expected Win Percentage

Before attempting to illustrate the role of managerial decision making in explaining the difference between actual and expected wins, it is useful to attempt to identify potential player contributions to this relationship. Given that actual runs scored and runs allowed, offensive and defensive productivity in baseball, would mainly be driven by the players who receive the majority of the playing time, this second-order effect that we are investigating will focus on the impact of role players who may have considerable impacts at the margins impacting late-innings runs scored and runs allowed. To do this, we focus on bench players who may pinch-hit, pinch-run, and/or be defensive substitutions in late innings along with the bullpen (who pitch after the starter has been removed from the game). To attempt to allow this distinction to be as exogenous as possible, we defer to the website www.baseballreference.com and use their categories of players for this purpose. The website divides players into tiers for both hitters and pitchers. For hitters (non-pitching position players), the website lists the players who started the majority of games and the bench players. For pitchers, they categorize these players as starters, relievers, and a depth group. The relief pitchers category is the five-most-used relief pitchers including the pitcher(s) who served as closers. The last category of pitchers we classify as "depth" pitchers on the roster.

The logic for focusing on these players to attempt to explain why teams may win or lose more games than their run differential would suggest is that bench players and relievers may cause swings in games that are close and tight in the late innings. Timely pinch hits may change the result in a game as could a good or poor performance from the relief pitchers. The players who receive the majority of the playing time will normally account for majority of the runs scored and runs allowed during the season, but these bench/relief players may help to decide outcomes of one-run games, which ultimately can skew a positive run differential into a losing record or a negative run differential into a winning record.

Although different possible offensive and defensive metrics were considered and tried when constructing the model, we ultimately decided upon OPS+ for hitting and show alternative results for FIP, ERA, and K-BB for pitchers. OPS+ is on-base plus slugging average for hitters, adjusted for park effects and scaled to the league average. FIP is fielding-independent pitching and attempts to account for pitching performance that the pitcher can control, independent of the fielders around him. ERA is a traditional stat that stand for earned run average (number of earned runs allowed per 9 innings of work) and K-BB is the strike-out (K) to walk (BB) ratio for the pitcher. Each of the results of the simple regression models are shown below. Coefficients are presented for each independent variable with t-stats in parentheses. Statistical significance is noted by *-notation. The dependent variable in each regression model is the difference between actual winning percentage and the expected (Pythagorean) win percentage of the team. The expected win percentage is taken from the team's run differential. Table 1 presents the summary statistics of the variables used and table 2 presents the regression results.

Table 1: Summary Statistics of Team, Bench, and Reliever Performance Variables – 2000-2013

	Difference (Act – Exp Win %)	***OPS+***	***FIP-Relief***	***FIP-Depth***	***ERA-Relief***	***ERA-Depth***	***K/BB-Relief***	***K/BB-Depth***
Mean	-0.0004	82.0315	3.9482	4.7816	3.7298	4.9805	4.1851	2.7602
Median	-0.0003	82.4957	3.9090	4.8003	3.6544	5.0098	4.1608	2.5719
Standard Deviation	0.0250	12.1169	0.5559	0.6569	0.7409	0.9524	1.2454	1.2113

Table 2: Regression Results: Bench and Pitching Performance on the Difference between Actual and Pythagorean Win Percentage – 2000-2013

I		***II***		***III***	
Intercept	0.0206	Intercept	0.0071	Intercept	-0.0102
	(1.4307)		(0.5750)		(-1.0729)
OPS+ (Bench)	0.0001	OPS+ (Bench)	0.0001	OPS+ (Bench)	0.0001
	(0.1104)		(0.1911)		(0.1795)
FIP – Relief	-0.0044**	ERA – Relief	-0.0026	KBB – Relief	0.0014
	(-1.9130)		(-1.5358)		(1.3735)
FIP – Depth	-0.0009	ERA - Depth	0.0001	KBB - Depth	0.0007
	(-0.4956)		(0.0756)		(0.7328)

Coefficients are presented along with t-statistics in parentheses. Statistical significance is noted by *-notation at the 10% (*), 5% (**), and 1% (***) levels.

The only regression result that revealed any statistically significant results was the first specification involving measuring pitcher performance, Fielding-Independent Pitching (FIP). The other specifications, using Earned Run Average (ERA) and Strikeout to Walk Percentage (KBB) were not found to have significant results. Given that FIP measures the performance of a pitcher independent of the fielders around him, this pitcher quality measure may be preferred to the other possibilities and could be why it helps to explain the difference between actual and expected wins of a team.

In specification I, the relief pitcher performance measured by FIP was found to have a negative and significant effect on the difference between actual and expected (Pythagorean) wins. As FIP increases, indicating poorer relief pitcher performance, the team is more likely to lose more than their run differential would indicate, leading to fewer actual wins compared to expectations. The depth pitching FIP variable was also shown to have a negative effect, but was statistically insignificant. In a similar fashion, the role of the bench players was shown to have a small positive effect on actual vs. expected wins, but was not shown to be statistically significant.

Overall, it appears that relievers have the ability to impact losses beyond those indicative of a team's run differential. At the margin, a reliever that gives up a key run or two could have a much greater impact on the win/loss record of a team as that run (or runs) are allowed late in the game, which may impact which team actually wins the game.

To explore how much the best relief corps compare to the worst relief corps in our sample, a simple table comparing the top 10% of teams that outperformed their run differential and the bottom 10% of teams that, in this case, underperformed their run differential is shown below. The values of each of the variables studied above is given for each subsample along with the % difference in the samples (top 10% compared to bottom 10%).

Table 3: A Comparison of Team Performance Variables when MLB Teams Outperform and Underperform Their Run Differentials – 2000-2013

Sample	***OPS+ (Bench)***	***FIP – Relief***	***FIP – Depth***
Top 10% Seasons – Outperform Run Differential	81.5630	3.6622	4.6913
Bottom 10% Seasons – Underperform Run Differential	81.6380	3.9818	4.7697
% Difference Between Samples	-0.09%	-8.03%	-1.64%

The biggest difference, percentage-wise, between the team performance variables between the top 10% and bottom 10% seasons of run differentials is with the FIP for Relief pitchers. An over 8% difference in Fielding-Independent Pitching exists between the best run differential and worst run differential teams. This illustrates the importance, at the margin, of having effective relief pitching to protect leads and to give teams changes to overcome run deficits in late innings of baseball games.

Decision of Baseball Managers and Their Role in Actual vs. Expected Winning Percentage

For this section of this study, we add managerial decision-making data for Major League Baseball. Data is presented for managers each season in the Bill James Handbook. Bill James is the innovator of advanced statistics and SABR-metrics for baseball and an annual volume of informative statistics are still printed each year after the conclusion of the season before the next season begins.

The managerial data is difficult to find on-line in a summarized form, so the Bill James Handbook serves as an excellent source for this data. Data that is presented in the handbook and was used in this study cover various aspects of the game of baseball where the manager has a direct influence on the game. These statistics are pinch hitters used, pinch runners used, defensive substitutions, relief innings pitched, stolen bases attempted, sacrifices ordered, and pitch outs ordered.

Pinch hitters and pinch runners used denotes the managerial decisions involving the use of bench players to give a better opportunity to score runs, typically in late innings. These decisions will also capture the relative quality of the bench players on the team, as managers may not have as much flexibility if their bench is poor. Defensive substitutions are similar managerial decisions, but on the defensive side of the game. Defensive substitutions remove a starter from the game, who may have greater offensive prowess, and replaces the player with a higher-quality defensive player. This, again, may give the team a better chance to win close games in the late innings.

Relief innings pitched is a variable that illustrates the use of the bullpen by the manager, but as with some of the other stats mentioned above, this will also reflect the quality of the bullpen the manager has at his disposal. Stolen bases attempted signifies the degree of risk-taking a manager is willing to tolerate on the base paths, which could potentially lead to more runs, but could also remove a runner from base and eliminate a potential scoring opportunity. Sacrifice attempts also shows the risk preference of a manager by essentially giving up an out to provide a better opportunity for a runner on base to score (advancing the runner from first to second through a bunt). The last variable, pitch outs ordered, denotes the number of times the manager has called for a pitch out which wastes a pitch, but if timed correctly, may lead to the removal of a runner currently on base through a caught stealing possibility. Like some of the other variable, the decision to call for a pitch out may reflect the risk-preference of the manager as those who call for more pitch outs may be considerably more risk averse in their role as manager of a team.

The regression results are presented below. The dependent variable is the same as before, the difference between actual win percentage and expected (Pythagorean) win percentage based upon team run differential. Table 4 presents summary statistics of the managerial variables and Table 5 presents the regression results. Coefficients are presented with t-stats in parentheses. *-notation denotes statistical significance.

Table 4: Summary Statistics of Managerial Decision Variables – 2000-2013

	Pinch Hitters Used	***Pinch Runners Used***	***Defensive Substitutions***	***Relief Innings Pitched***	***Stolen Bases Attempted***	***Sacrifices Attempted***	***Pitch Outs Ordered***
Mean	188.1949	31.1620	33.6152	453.8608	134.7924	69.3873	21.5013
Median	228.0000	30.0000	32.0000	460.0000	131.0000	70.0000	18.0000
Standard Deviation	89.0850	14.3574	15.3624	51.2778	56.5625	26.1723	15.7355

When managerial decision are included in the model, the impact of the team performance variables are not found to have the same directional effect on actual vs. expected wins, but are not statistically significant. The only managerial variables shown to significantly impact the difference between actual and Pythagorean wins are the number of defensive substitutions used and the number of pitch outs ordered. The number of pinch hitters and runners, relief innings pitched, stolen bases attempted, and sacrifices ordered were not found to have a significant impact on the dependent variable.

Table 5: Regression Results: Bench and Pitching Performance on the Difference between Actual and Pythagorean Win Percentage Including Managerial Decisions – 2000-2013

Variable	***Coefficient (t-stat)***	***Variable***	***Coefficient (t-stat)***
Intercept	0.0200	Defensive Substitutions	0.0003***
	(-0.9657)		(3.5602)
OPS+ (Bench)	-0.0001	Relief Innings Pitched	-0.00004
	(-0.2786)		(-1.4258)
FIP – Relief	-0.0023	Stolen Bases Attempted	0.000007
	(-0.9656)		(0.3458)
FIP – Depth	0.0012	Sacrifices Attempted	-0.00002
	(0.5889)		(-0.2871)
Pinch Hitters Used	-0.000004	Pitch Outs Ordered	-0.0002**
	(-0.1621)		(-2.0751)
Pinch Runners Used	0.00004		
	(0.4072)		

Coefficients are presented along with t-statistics in parentheses. Statistical significance is noted by *-notation at the 10% (*), 5% (**), and 1% (***) levels.

The number of defensive substitutions used was found to have a positive and significant effect on the difference between actual and expected wins. Teams that used more defensive substitutions were found to be more likely to outperform their run differential and win more games than expected. Having roster flexibility, where late-inning defensive players can replace good offensive players who may not have great defensive abilities, appears to allow a team to win more close games and outperform their run differential. This finding supports the research showing that roster flexibility is valuable for a Major League Baseball team (Chan and Fearing, 2013).

The other variable that was shown to have a statistically significant impact on the difference between actual and expected wins was the number of pitch outs ordered. Managers who ordered more pitch outs were found to lose more games than their run differential would suggest. This could be due to the team having a catcher who may not be effective at throwing out base stealers leading to the manager wasting a pitch from the pitcher to compensate for this inferior skill. It could also suggest a pitching staff who cannot effectively hold runners, or a combination of this factor and poor catcher ability, leading to more losses than expected. Alternatively, or in addition to the other possibilities noted above, this variable may serve as a proxy for risk-averse behavior on the part of managers. Managers who choose to call for more pitch outs may be more risk-averse overall, which may lead to losing more close games due to "playing it safe" in certain situations throughout the game.

Table 6: A Comparison of Key Managerial Decision-Making Variables when MLB Teams Outperform and Underperform Their Run Differentials – 2000-2013

Sample	***Defensive Substitutions Used***	***Pitch Outs Ordered***
Top 10% Seasons – Outperform Run Differential	36.0732	17.2927
Bottom 10% Seasons – Underperform Run Differential	26.1892	21.5000
% Difference Between Samples	37.74%	19.57%

As can be seen in Table IV, there is a large difference in managerial decisions when it comes to using defensive substitutions and ordering pitch outs. In the top 10% of seasons where teams outperformed their run differential, managers used over 37% more defensive substitutions than managers of teams in the bottom 10% of the sample. Likewise for pitch outs ordered by managers, there was nearly a 20% difference between samples as managers of the teams who were best at outperforming their run differential ordered fewer pitch outs than those who significantly underperformed their run differential.

Conclusions

In the highly-competitive sport of baseball, the difference between winning and losing is often small. The greater the level of competitive balance within the league, as talent is more evenly distributed across teams, the more important small advantages at the margin become in determining winners and losers in the sport. This research examined the determinants of the difference between actual and expected win percentage using the Pythagorean win percentage based upon runs scored and runs allowed.

Teams which outperformed their run differential won more games than the difference in runs suggested, while teams that underperformed their run differential lost more games than their run differential suggested.

We examined the possible determinants of the difference between actual and expected win percentage in two steps. We first focused on bench players and relievers under the assumption that the main bulk of runs scored and runs allowed comes from the players who play the most innings, the starting players on the field and the starting pitchers. It was assumed that teams who outperform or underperform their win differential may receive contributions at the margin at pivotal times in the game, which could be due to successful relief pitching and bench hitting performance. Through a simple regression model using the actual minus expected win percentage as the dependent variable and various measures of relief pitching and bench hitting performance, it was shown that the only variable that was statistically significant was relief pitching FIP, fielding independent pitching. FIP was shown to have a negative effect on actual minus expected wins implying that higher (lower) levels of FIP led to more (fewer) losses than implied by the team's run differential.

Next, we examined the role of managerial decision-making in attempting to explain the difference between actual and expected wins. Using data from the Bill James Handbook on manager decisions on a variety of statistical fronts, we included these variables in the regression model to attempt to determine if managers have any direct influences on game outcomes at the margin that impact how teams either outperform or underperform their run differential. The number of defensive substitutions was found to have a positive and significant effect on the difference between actual and expected win percentage as managers who used late-inning defensive replacements were found to have more success than the team's run differential would suggest. This could be due to astute changes by the manager or could simply be due to having a deep roster which allows for more flexibility during the game, whose success could be more a function of the general manager and front office.

The other managerial variable that was found to have a statistically significant impact on the difference between actual and expected win percentage was the number of pitchouts ordered during the season. The pitchout is a purposeful decision to give up a pitch for a ball, to attempt to time when a baserunner may steal and increase the likelihood of throwing that runner out. Managers that ordered more pitchouts were found to lose more game than their run differential suggested. This negative impact of pitchouts ordered may reveal poor defensive team aspects of the catcher and pitcher for holding runners, but could also suggest overall risk averse behavior by the manager of the team. If this risk averse behavior manifests itself in other situations, in addition to throwing more pitchouts than other teams, this risk aversion may lead to teams losing more close games and having these managers and teams significantly underperform their run differential during the season.

References

Chan, T. and Fearing, D. (2013) "The Value of Flexibility in Baseball Roster Construction." *MIT Sloan Sports Analytics Conference* (online: http://www.sloansportsconference.com/?p=10180)

Kahneman, D. and Lovallo, D. (1993) "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science*, 39(1), 17-31.

Lewellen, K. (2006) "Financing Decisions when Managers are Risk Averse." *Journal of Financial Economics*, 82(3), 551-589.

Lewis, M. (2003) *Moneyball: The Art of Winning an Unfair Game*. New York: W.W. Norton.

March J. and Shapira, Z. (1987) "Managerial Perspectives on Risk and Risk Taking." *Management Science*, 33(11), 1404-1418.

Parrino, R., Poteshman, A., and Weisbach, M. (2005) "Measuring Investment Distortions when Risk-Averse Managers Decide Whether to Undertake Risky Projects." *Financial Management*, 34(1), 21-60.

Thaler, R., Tversky, A., Kahneman, D., and Schwartz, A. (1997) "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test." *The Quarterly Journal of Economics*, 112(2), 647-661.

Comparative Return Distributions:Strangles versus Straddles

Glenn Pettengill and Sridhar Sundaram, Grand Valley State University

Abstract

Straddles and strangles are option strategies holding an equal number of puts and calls. The straddle holds options with a strike price, closest to the current stock price. The strangle holds out-of-the-money options. On a per option basis the strangle has less invested risk, at the cost of lower possible dollar return which is the comparison made in extant comparisons using profit diagrams, emphasize this comparison. But an investor is concerned with return on investment, not per unit dollar return. On a return basis, the lower cost of the strangle provides leverage, more risk and greater potential return. An empirical study shows the strangle with greater risk and return potential leading us to propose the use of return diagrams in place of profit diagrams.

Introduction

Straddles and strangles represent two similar complex option strategies. Both strategies involve long positions in an equal number of puts and calls. Both strategies are designed to benefit from a large price movement in the underlying stock in either direction. The investor is assuming that a forthcoming event will cause a large change in the stock price but is unsure in which direction the price movement occurs. The long put position will benefit the investor if prices fall substantially. The long call position will benefit the investor if prices rise substantially. Both positions are profitable only if the market underestimates the volatility of the underlying stock.

Despite these similarities, the two complex strategies offer substantially different return potential. A straddle holds an equal number of puts and calls with the same strike price, generally at the strike price closest to the price of the underlying stock. In contrast, the strangle holds an equal number of calls and puts purchased at different strike prices. Both of the options are purchased out of the money generally with similar distances between the strike prices and the current stock price. Thus, the call is purchased with a strike price above the current stock price and the put is purchased with a strike price below the current stock price.

Five important contrasts exist for the investor between the use of a strangle and a straddle. Because the options purchased for the strangle are out of the money, these options are cheaper than the options purchased for the straddle. Thus, each strangle contract is cheaper than the corresponding straddle contract on a per unit basis. Second, because the strangle is cheaper than the straddle on a per unit basis the maximum loss on the straddle is always higher than for the strangle. This first two contrasts seem to favor the use of a strangle. A third contrast definitely favors the use of the straddle. Because the straddle uses puts and calls with the same strike price, there is a single point at which the investor would lose 100% of the investment if the position is terminated at option expiration. The straddle suffers a return of -100% at expiration only if at that time the stock price is equal to the strike price. If the stock price at expiration is above the strike price the call is in the money and the investor gains some income from the sale of the call. If the stock price at expiration is below the strike price the put is in the money and the investor gains some income from the sale of the put. On the other hand, for the strangle there is a range of expiration stock prices for which the investor experiences a return of -100%. The call is purchased with a strike price above the current stock price and will be worthless at expiration if the stock price at that point is less than the strike price. The put is purchased at a strike price below the current stock price and will be worthless at expiration if the stock price at that point is more than the strike price. Thus at any point between the two strike prices the return to the strangle strategy is -100%. A fourth contrast deals with the expiration break-even point. Although the price of the strangle is cheaper and thus requires a smaller gain to break even than does the straddle, the straddle starts cutting into the maximum loss immediately while the strangle suffers the maximum loss over the entire range of stock prices between the two strike prices. Thus, the straddle is likely to break even more quickly. This advantage for the straddle may seem dominant when examining a profit diagram because after each strategy breaks even both strategies gain dollar for dollar, on a per option basis, as the stock price either falls or rises. But a fifth a critical comparison exists. Because the options purchased for the strangle are cheaper, the dollar for dollar gain to the strategy represents a higher gain on a percentage basis for the strangle as compared to the straddle. Thus, on a return basis the contrast between the strangle is much more likely to have a return of -100% but will have greater returns if the strategy turns positive relative to a straddle.

This fifth contrast presents the critical issue for this paper. A profit diagram, using a per unit cost basis, obscures the return on the total investment for the investor. (And there is no reason to assume that an investor will use a different initial investment

when choosing between the two strategies). Indeed comparisons between the strangle and the straddle present a classic example of the risk-return tradeoff. The strangle provides leverage with a much higher risk level but with the potential for superior returns. In this paper we conduct an empirical study comparing returns of a strangle with a straddle. To date, to our knowledge, there has been no such study. Our purpose is not to identify a superior strategy but rather to provide investors with a more complete picture of the difference in the risk-return tradeoff between the strategies.

The rest of the paper is organized as follows. In Section II we provide a comparison of the straddle and strangle strategies using the traditional profit diagram created for IBM options. Our purpose in this analysis is to provide reference for additional analysis. In Section III we provide a discussion of the short extant literature that compares these two strategies which is found entirely in the practitioner literature. In Section IV we describe our sample and in Section V we discuss our empirical results. Our empirical results lead us to advocate that a return diagram rather than the traditional profit diagram should be used to compare these two strategies. In Section VI we illustrate the use of this strategy. A conclusion follows.

A Comparison of a Strangle and Straddle using Profit Diagrams

To illustrate the comparisons between strangles and straddles made in the previous section we create traditional profit diagrams for each of these two strategies. A profit diagram shows the cost of establishing an option strategy at some point before expiration and the value of the strategy at the expiration date given various prices for the underlying stock at expiration. We illustrate the profit on a per option basis rather than for a contract. In this comparison we purchase a strangle and a straddle using options for IBM stock. The options, purchased on July 3, 2015, have an expiration date of October 15, 2015. When the options were purchased IBM was trading at $165.09. Thus, the straddle was established using put and calls with a $165 strike price. The strangle was established using a call with a $170 strike price and a put with a $160 strike price. Table 1 shows the prices for the October IBM call and put options with strike prices of 160, 165 and 170 using last prices reported on Yahoo midday July 3, 2015.

Table 1: Prices of October IBM Options Midday July 3, 2015

Option Type	**Strike**	**Price**
Call	160	$8.51
Call	165	$5.81
Call	170	$3.56
Put	160	$4.76
Put	165	$6.95
Put	170	$9.60

Figure A shows the profit diagram for the straddle. The straddle buys the 165 call for $5.81 per option and the 165 put for $6.95 for a total cost of $12.76. The investor losses $12.76 per straddle if at expiration IBM closes at $165, but only if IBM closes at $165. The return to the strategy is -100% for a single point because at any other expiration price for IBM other than $165 either the call or the put will be in the money. If the price of IBM at expiration is greater than $165, the call has intrinsic value and the investor can sell the call for its intrinsic value. But, the investor starts with a loss of $12.76 so in order for the investor to break even the price of IBM must increase to $177.76 at expiration. Likewise, if the price of IBM at expiration is less than $165 the put has intrinsic value and the investor can sell the put for its intrinsic value. But again, the investor starts with a loss of $12.76 so in order for the investor to break even the price of IBM must decrease to $152.24 at expiration. At any expiration within the rather wide range of $152.24 to $177.76 the investor suffers a negative return. At any expiration price for IBM above $177.76 or below $152.74 the investor has a positive return. As the expiration price for IBM goes above $177.76 or below $152.24 the straddle investor gains dollar for dollar with the price movement on IBM on a base investment of $12.76. A return of positive 100% is gained if the price increases to $190.92 or falls to $139.48.

The strangle also purchases a call and put on the same underlying stock but these options are purchased out of the money. The long position in the call and in the put held by the strangle results, as did the straddle, in a complex strategy that has a negative return over a substantial range of expiration prices for the underlying security, but with positive returns if the price increases sufficiently or decreases sufficiently. But the use of two out-of-the-money options results in some significant differences between the two strategies, as illustrated in Figure B, in the profit realized.

Figure A: Profit Diagram – Buying the Straddle for IBM: Buy 165 Call and Put

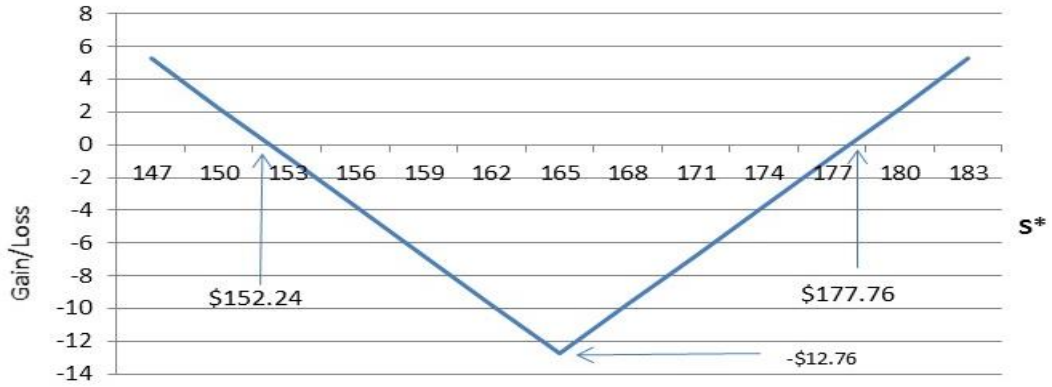


Figure B: Profit Diagram – Buying the Strangle for IBM: Buy 170 Call and 160 Put

The cost of a strangle is $8.32 ($3.56 for the 170 call and $4.76 for the 160 put) as opposed to a cost of $12.76 for the straddle. It is tempting to view this cost differential in favor of the strangle as an advantage to the strangle. Except for the most constrained of individual investors it is not. An investor can buy two straddles or three strangles for approximately the same amount. The initial investment cost differential is not material. The concern to the investor is the return and risk that is generated by the strategy.[1]

The key impact on return is the greater leverage provided by the strangle suggesting higher risk but higher potential return for the strangle. As shown in Figure B, the strangle definitely provides more risk. The strangle is much more likely to provide a return of -100%. The straddle provides a return of -100% only if at expiration the price of IBM is exactly $165. The investor holding a strangle has a profit of -$8.32 and loses 100% of the investment if the closing price of IBM at expiration is anywhere between $165 and $175. The fact that the maximum loss of $8.32 on a per unit basis is less than the loss of $12.76 on a per unit basis for the strangle is immaterial. The investor cares about his/her loss on the total investment.[2] The strangle only begins to have value at expiration, reducing the maximum loss, if the put has value because IBM closes below $160 or the call has value because IBM closes above $170. This higher risk should associate with higher return potential. It does but the profit diagram would tend to suggest otherwise.

As shown in Figure B, the put (call) has sufficient value to offset the cost of the strangle only if the price of IBM at expiration is $151.68 ($178.32). In contrast the straddle strategy, as noted above, breaks even with an IBM price at expiration of $152.24 ($177.76). The strangle has returns of -100% for a longer time period than the straddle and has a negative return for a longer time period than the straddle. Where then is the return advantage for the strangle? After either strategy breaks even, both the strangle and the straddle gain dollar for dollar as the price of IBM at expiration goes up or down further. The strangle would seem to never catch up! But, here is the key! The strangle is gaining the same dollar as is the straddle but on a lower invested value. The strangle uses more leverage consistent with the higher risk described above. But leverage provides higher potential return. On a percentage basis the dollar for dollar gain experienced by both strategies provides a higher and more rapid percentage growth rate for the strangle than for the straddle. We noted above that the straddle earns a 100% return with closing

prices for IBM at expiration of $139.48 or $190.92. The corresponding values for the strangle are $143.36 and $186.64. The return advantage for the strangle continues to grow as the closing expiration price for IBM becomes more extreme.

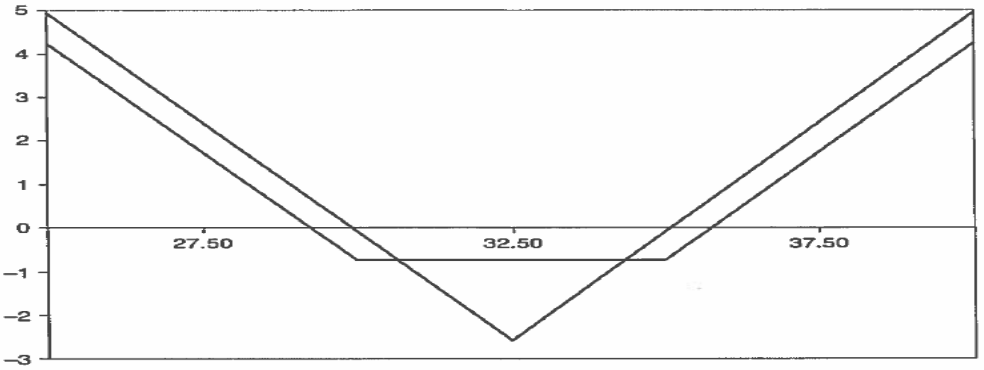
Comparisons Straddles and Strangles Previous Discussions

To our knowledge there is no comparison of straddles and strangles in the academic literature. As cited more fully in the reference section, we find five recent practitioner investment guides which discuss and compare straddles and strangles: Bouzoubaa and Osseiran (2010), DraKoln (2009), Fontanills (2005), Rhoads (2010), and Sinclair (2010). All five provide very similar comparisons.

Each of the guides emphasizes the similar goal of the strategies to benefit from an event providing a major change in the price of the stock regardless of the direction of the price change. Rhoads (2010, p. 89) goes so far as to state that: "They (referring to strangles and straddles) are usually considered interchangeable as the trading motivations behind both strategies are very similar." Likewise, all of the guides emphasize the results as illustrated in a profit diagram, identifying that both strategies have unlimited profit potential with maximum loss limited to the initial investment. Thus, in contrast to what one would conclude by looking at return patterns, the investor is presented with a description that identifies a low risk investment from losing only the initial investment with the potential for a large gain. Consistent with this view all guides represent the short position in the strategies as being riskier than the long position in the strategies, despite the much higher probability of a positive return to the short position.

The difference between the strategies is also consistent with the view from comparing the profit diagrams which show a smaller maximum loss for the strangle but with a higher break-even point for an increase in the underlying asset and a lower break-even point for a decrease in the underlying asset. The following quote from Sinclair (2010, p. 109), comparing the strangle to the straddle, illustrates this consistent contrast: "The profit as function of both time and volatility is the same for the long strangle position. The only difference is that we need a bigger move in the underlying for the position to profit. To compensate for this the premium paid for the position will be smaller than for a straddle." All guides make generous use of profit diagrams to illustrate this contrast. As shown for example by Rhoads (2010, p. 93), reproduced below as Figure C.[3]

Figure C: Comparison Straddle and Strangle Using Profit Diagrams



Because these guides emphasize the per unit profit from a strategy, the investor does not see the high risk of both strategies in terms of expected return. Nor is the investor made clearly aware of the real contrast between the two strategies in terms of the risk-return tradeoff. The investor should be made aware of the much higher risk with the strangle with the consequent potential for higher reward. Our empirical study is designed to illustrate this contrast.

Before moving to our study we must identify that there is some implicit recognition of the risk-return tradeoff in the guides. Most explicitly, Rhodes (2010, p. 92) recognizes that: "However, on the use of capital basis the strangle is actually more profitable." Bouzoubaa and Osseiran (2010, p. 101) implicitly recognize the higher risk for a strangle, indicating that the user of the strangle expects more volatility in the underlying stock.

We should also indicate that our study limits risk-return tradeoffs to a position held to expiration and that these guides generally identify the strategies as single event strategies and instruct the investor to terminate the position after the event. DraKolin (2009, p. 73) identifies an advantage for the strangle in this regard stating that the strangle provides greater flexibility in determining when to terminate the position. Ironically, this is an implicit recognition of the greater risk of the strangle as the flexibility is provided by the greater likelihood that the risk would show a large loss after the event takes place.

Sample and Methodology

Our sample selection procedure is designed to foster comparability and data availability. We begin by selecting the 30 Dow Jones Industrial Average (DIJA) securities as our sample universe. To foster comparability in establishing spreads for the strangles we discard the five stocks with the highest market prices and the five stocks with the lowest market prices. For the remaining twenty securities we discard the five companies with the lowest open interest for calls at the near term expiration data. The remaining fifteen securities constitute our sample firms and the list of these securities is provided in Appendix I.

For these fifteen securities we create straddles and strangles and calculate returns for these strategies under the assumption that the position is held until expiration. Strategy positions are established on a weekly basis (each Wednesday) for a three-month period commencing with the second week in September and ending with the second week in December during 2015. Straddles are formed with calls and puts with strike prices nearest the current strike price. The strangles are formed with a $10 spread with a call purchased at a strike price $5 above the strike price used for the straddle and with a put purchased with a strike price $5 below the strike price used for the straddle.[4] Each week we form two strangles and two straddles for each of the fifteen sample securities. The first set of complex option positions is established with a time to expiration as close as possible to one-month away with the proviso that the time to expiration must be at least three weeks. The length of time to expiration will vary across securities given a particular security's expiration date pattern. The second set of complex strategies is formed with a January expiration date which is available for most securities.[5] Using this procedure we calculated a total of 416 strategy pairs. The holding period for the trials ranged from three weeks (only two observations) to 18 weeks for the sample created September 9 with a January 15 expiration date.

Returns are determined on a per contract basis, ignoring transaction costs. The cost of the contract for the strategy is based on the last price for the option as reported on Yahoo. We recognize that the last price reported on Yahoo may be stale and may be biased downward as our purchase price would likely occur at the ask and the last price could either have been at the bid or the ask. This price bias should apply equally to the determination of the cost of the strangle and the straddle so no bias should exist in our comparisons of the strangle and the straddle strategies. Price staleness should occur more frequently with the strangle as trading should be more active close to the strike price, but we know of no reason that this greater staleness would have a tendency to increase or decrease the return of the strangle as compared to the return of the straddle. We hold each strategy until the expiration date of the option and determine end value of the contracts by the intrinsic value of the options given the closing price of the underlying strike.

Empirical Results

Profit diagrams compare results between investing in strangles and straddles on per unit dollar basis. These comparisons and comparisons made in extant literature argue that strangles are less risky and receive less reward than straddles. Strangles are seen as less risky because the per unit dollar loss on strangles is less than the per unit dollar loss on straddles. Consistent with the comparison, the return reward for strangles is seen as lower as both strategies gain dollar for dollar with a stock price increase or decrease after a break-even point is reached. Because the break-even occurs for the straddle before the break-even for the strangles, on a per unit dollar basis the gain for the strangle is always less than the gain for the straddle. Thus from the perspective of the profit diagram, the strangle is seen as having less risk with the consequent lower potential return. But this comparison is simply wrong. Just as one would not consider investment in Berkshire Hathaway to be riskier than investing in penny stocks because the per unit dollar loss potential is much higher for Berkshire Hathaway, one should not compare strangles and straddles on a per unit dollar gain or loss. Just as one should make a risk-return comparison between Berkshire Hathaway and penny stock on a return basis, one should compare the risk-return pattern between strangles and straddles on a return basis. When one does so, the comparison stands on its head. Strangles are the much riskier investment, but strangles also have greater return potential than straddles. The 416 return calculations for paired strangles and straddle investments show this relationship clearly.

To illustrate the relative risk and return of the two strategies, Table 2 reports the return distributions for both strategies along with the distribution of the return difference between the two strategies calculated as the return to the strangle minus the return to the straddle for each paired trial. The high risk of the strangle is clearly shown in the return distribution. In 225 of the 416 trials (54.09% of the trials), the strategy provides a return of -100%. In over 50% of the trials an investor would have lost the entire investment. In nearly 65% of the cases an investor would have received a negative return. The downside risk is much less for the straddle, but still substantial. The straddle never loses the total investment and a return of less than -50% occurs with less than half of the frequency of the total loss for the strangle. Indeed the probability that investors in strangle will lose 100% is greater than the probability that investors in the straddle will have a negative return.

Consistent with the risk-return tradeoff principle, the strangle has the greater return potential. In 104 of the 416 trials, exactly a quarter of the total trials, the return to the strangle is greater than 100%. And in seven cases an investor would earn a return

of over 1,000%. The strangle provides some truly amazing leverage. The return potential is decidedly less for the straddle, but still substantial. Sample results also report high returns for the straddle as in 18.75% of the trials the straddle has returns of over 100%. Contrary to what one would surmise from profit diagram comparisons, comparisons of return distributions show the strangle to be much riskier than the straddle with an accompanying higher return potential.

Table 2: Return Distributions: Strangle vs Straddle

Return Values	Observations Strangle	Observations Straddle	Observations Return Difference
-100%	225	0	36
"-99% to -50%"	22	110	102
"-50% to 0%"	22	105	148
percent negative	64.66%	51.68%	68.75%
"0% to 50%"	23	77	39
"50% to 100%"	20	46	23
"more than 100%"	104	78	68
"more than 1000%"	7	0	2
percent positive	35.34%	48.32%	31.25%
percent greater than 100%	25.00%	18.75%	16.35%
percent greater than 1000%	1.68%	0.00%	0.48%

Note: for the return difference column, the -100% row represents less than -100%. For these 36 observations the return for the strangle was generally -100% (in four cases returns were between -72% and -98%) and the straddle earned a positive return.

The column labeled return difference shows the frequency distribution of the return to the strangle minus the return to the straddle for the 416 paired observations. Consistent with the higher risk for the strangle, in close to 70% of the trials this variable is negative. In the vast majority of cases an investor would have earned a greater return with the straddle than with the strangle. In over 9% of the trials (36 out of 416) the return to the straddle is more than 100% greater than the return to the strangle.[6] These cases exist because a break-even point is achieved more quickly for the less risky straddle than the strangle. In most of these trials the straddle is earning a positive return while the stock price has not moved sufficiently for either the out-of-the-money call option or the out-of-the-money put option to have non-zero intrinsic value, providing the strangle with a return of -100%. The strangle is riskier than the straddle. The distribution of the return differences also shows that the riskier strangle has greater return potential than the straddle. In 68 trials the return to the strangle exceeded the return to the straddle by more than 100%. In two cases the return to the strangle exceeded the return to the straddle by over 1,000%. Once the return turns positive the greater leverage involved with the strangle leads to a much higher return.

Table 3 which reports summary statistics of the returns for the strangle and the straddle and the return differential confirms the higher risk and the higher return potential for the strangle relative to the straddle. The risk for both the strangle and the straddle is quite high. But in terms of total variation the risk is much higher for the straddle. The standard deviation of returns for all trials is more than twice as high for the strangle as for the straddle. Risk is also shown in the calculation of the median returns. For both the straddle and the strangle the most likely return is negative. But the median return shows greater risk for the strangle. More than half of the returns for the strangle are -100%. In contrast half of the returns for the straddles are below -5.87%. Still negative, but much healthier than the likely total loss for the strangle strategy. And the median for the return differential is -24.97%. In more than half of the cases the strangle outperforms the strangle by 25% or more. The strangle is indisputably riskier than the straddle.

Summary statistics also indicate that the high risk strangle has higher return potential. The maximum return for the strangle in the sample is a very healthy 1,578% which is more than twice the highest return for the straddle of 739%. In one trial the return to the strangle is more than 1,250% higher than the return for the straddle. These extremely high returns cause the average returns to be positive for both the strangle and straddle strategies. Consistent with the risk-return tradeoff the average return is higher for the strangle than for the straddle, 46.90% for the strangle versus 30.01% for the straddle. Contrary to what one may surmise from profit diagrams and extant literature the strangle is riskier with higher expected return than the straddle.

Table 3: Return Summary Statistics: Strangle vs Straddle

	Strangle	Straddle	Return Difference
Mean	46.90	30.01	16.99
Standard Deviation	268.29	133.86	158.64
Median	-100.00	-5.87	-24.97
Minimum Value	-100.00	-99.55	-181.06
Maximum Value	1577.97	739.27	1254.89

The key to the difference in the risk-return tradeoff between the strangle and the straddle is the use of the out-of-the-money options by the straddle. These out-of-the-money options require a greater movement in the price of the underlying asset before the options have positive intrinsic value at expiration. Thus, the strangle has, as displayed in the sample, a large chance of earning a -100% return and a greater chance of a negative return than does the straddle. The positive effect from using out-of-the-money options results from their cheaper price providing leverage and greater return potential once the returns for the strategy become positive. Figures D and E illustrate the negative and positive effects for the strangle from using the out-of-the-money options. Figures D and E together show the return for the strangle and for the straddle for each of the 416 paired trials. The observations are ordered by the return to the strangle from lowest to highest return. Figure D shows the paried returns for the 269 trials in which the return for the strangle is negative. The greater risk for the strangle is clearly evident. As the returns remain -100% for the strangle, the returns for the straddle grow and in a few cases turn positive. As the price of the underlying stock shows sufficient movement either the call or the put held by the strangle gains intrinsic value at the expiration of the trial. With a few exceptions the return to the straddle is less negative or positivie and the straddle provides superior results.

Figures D: Paired Returns: Strangle vs Straddle – Negative Returns for Strangle

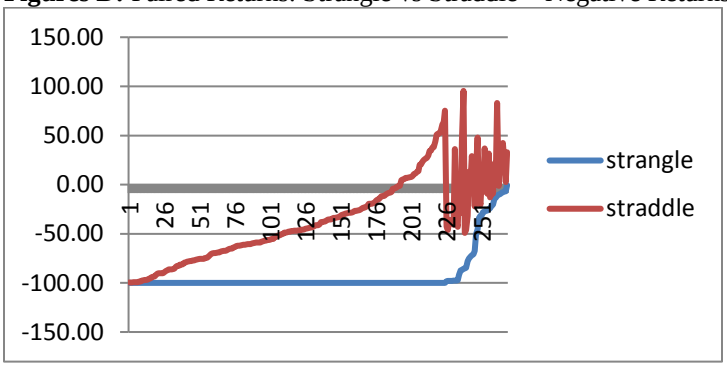


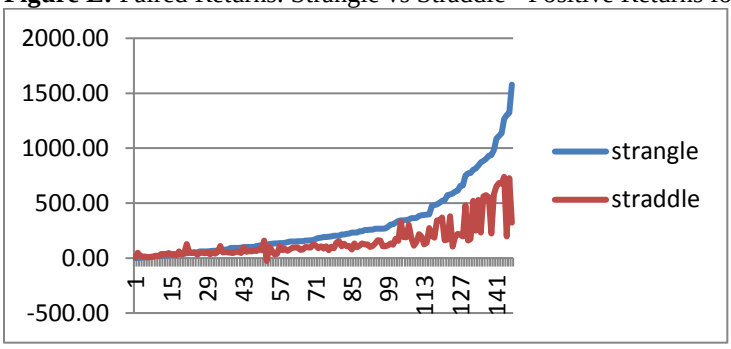
Figure E shows the positive effect of leverage when returns to the strangle turn positive. When returns to the strangle turns positive, the investor with the strangle quickly catches up with the investor holding the straddle. The leverage held by the strangle results in a much higher return gain for the strangle relative to the straddle even though each strategy is gaining the same dollar amount as the price of the underlying asset continues to gain or lose. At extreme price movements for the underlyng asset the return difference between the strangle and the straddle becomes large. As shown in Table 2, in some cases the return to the strangle exceeds the return to the straddle by more than 1,000%. In direct contrast to what is suggested by profit diagrams, the strangle has much more risk than the straddle but has a much higher return potential. The inaccurate picture provided by the profit diagrams begs for a better explanatory tool to illustrate the difference between the two strategies.

Use of the Return Diagram

The risk-return calculations found in our empirical studies provides a stark contrast to extant discussions concerning the comparison of the straddles and strangles. We illustrate that the strangle has a very high probability of providing a return of -100% and a greater chance of a negative return, but with a higher return potential. Because these important findings contrast to what an investor may view in a profit diagram, we suggest the need for a new diagram, a return diagram. In this diagram we

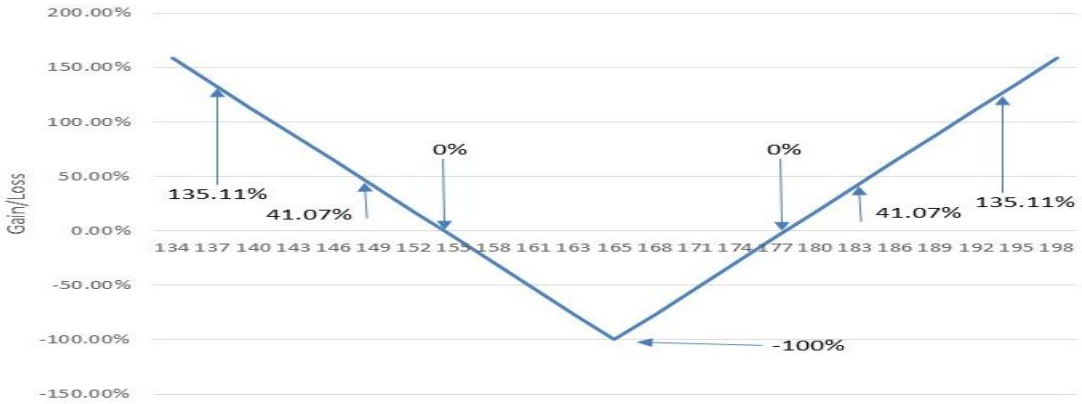
still show the result of the strategy assuming the strategy was held to expiration and with the vertical axis showing the price of the underlying asset at expiration. However, we substitute a percentage return for the dollar value typical for the profit diagram.

Figure E: Paired Returns: Strangle vs Straddle - Positive Returns for Strangle



To illustrate, we use the data on which we created profit diagrams in Section II to create comparable return diagrams. Figure F represents the return diagram for the straddle and Figure G represents the return diagram for the strangle.

Figure F: Return Diagram – Buying the Straddle for IBM: Buy 165 Call and Put.



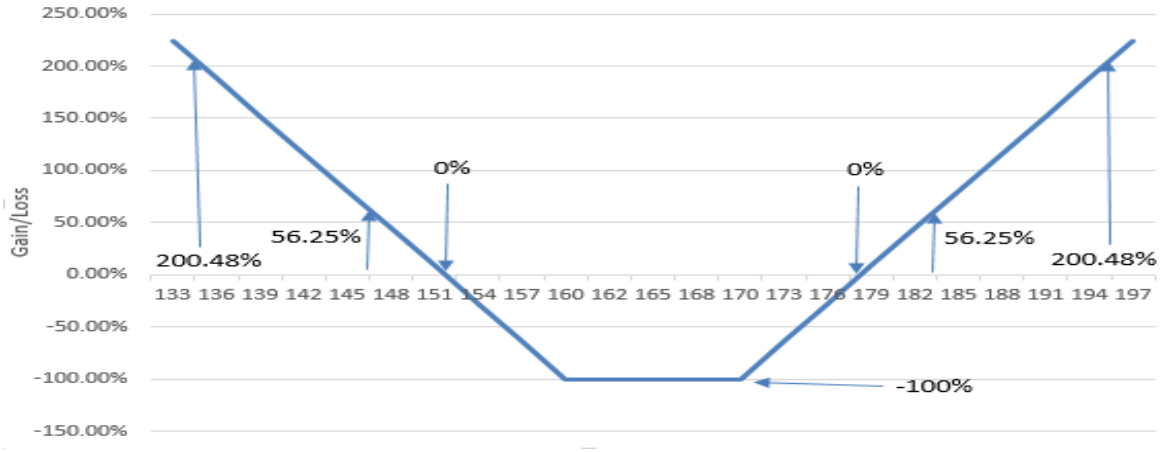
With the percentage return we emphasize the greater risk for the strangle. There is a single point where the straddle yields a return of -100% but there is a wide range of the maximum loss for the strangle. When comparisons are made using profit diagrams, as shown in Figure C, it appears that the straddle has a larger loss for a substantial period of time. This is only true on a per unit basis not on the basis of total return which is the variable of return for the investor!

As shown in Figure G the strangle shows no gain in return between expriation prices of $160 through $170. As shown in Figure F the straddle steadily reduces losses during this period as either the call or the put is in the money. The return diagram emphasizes the greater risk for the strangle.

If the risk-return tradeoff holds the strangle ought to be rewarded with greater potential return. Comparisons of Figures G and F clearly show this relationship. The profit diagrams obscure the greater return potential of the strangle, as these diagrams show the same dollar for dollar gain for the strangle and the straddle. And because the break-even occurs later for the strangle than the straddle, it appears that the straddle never catches up. In fact, once the strangle breaks even it quickly surpasses the straddle in terms of percentage return. Because of the greater leverage the strangle is gaining at a much faster rate once either the call or the put is in the money. For the strangle, after an option is in the money, for every dollar decrease or increase in the price of the stock the strangle gains 12.02% in return. In contrast the straddle gains at a smaller rate of 7.84%. Thus, if at expiration the stock price is either $147 or $183, the return to the strangle is 56.25% as compared to a return of 41.07% for the straddle. As shown on Figures G and F, this gap will continue to widen with either a further increase or decrease in the stock price. Thus, if at expiration the stock price is either $135 or $195, the return to the strangle is 200.48% as compared to a return of 135.11% for the straddle. As shown in the empirical results reported above, high potential returns are rare for the strangle.

The more likely result is a return of -100%. But this is exactly the tradeoff that an investor faces with the choice between the strangle and the straddle. This tradeoff is clearly shown in the return diagram but it is hidden in the profit diagram.

Figure G: Return Diagram – Buying the Strangle for IBM: Buy 170 Call and 160 Put



Conclusion

In this paper we compare the return distributions of two complex strategies: strangles and straddles. The strategies are similar in that they both involve long positions in calls and puts that will produce positive returns if the price of the underlying stock moves substantially either above or below current market price. The strategy are different in that buying a straddle involves buying a call and a put with the same strike price, generally at the strike price closest to the current price of the underlying stock. In contrast investing in the strangle involves buying a call and put which are both out of the money**.** The call has a strike price above the current price of the underlying stock and the put has a price below the price of the underlying stock.

We create paired samples establishing strangles and straddles with the same underlying stock and with the same expiration dates. We calculate invested wealth assuming that the strategy involves one contract, using option price data from Yahoo. We compute returns for these investments by assuming the position is held until expiration allowing us to use intrinsic value to find terminal wealth. Our results show that the strangle has much higher risk with higher return potential relative to the straddle.

Our findings are in stark contrast to current discussion of the relative risk and return tradeoff between strangle and straddles. Current discussions make comparison on a per unit basis and argue that strangles which have lower per unit costs are less risky given the lower maximum per unit loss. Current comparisons also cite lower return potential for strangles as once the strangle and straddle break even they both gain on a dollar for dollar basis with the change in the price of the underlying stock.

We argue that current comparisons are inaccurate because comparisons are being made on a per unit basis. Investors are concerned with return distributions not with per unit changes in value. The lower cost of the strangle provides greater leverage and the greater risk and return potential that we illustrate with our sample. We argue that the bias which leads to inaccurate current interpretations results from the use of standard profit diagrams which show profit on a per unit basis. As we show this bias disappears with the use of return distributions rather than profit diagrams. Although we know of no investment text or investor guide that uses return diagrams, we suggest that use of return diagrams would provide students and investors a clearer picture of expected results from option strategies.

Notes

1**.** As Pettengill, Gondhalker and Wingender (2015) argue a similar confusion has plagued textbook discussions of the relative risk of buying a put relative to shorting a stock. The argument is frequently made that buying a put is less risky than shorting a stock because all one can lose with a long position in a put is a price of the put, but examining the return distribution for the put relative to shorting the stock shows that investing in the put is indeed much riskier. The put very frequently provides a return of -100% with a small increase or a decrease for an out-of-the-money put in the price of the underlying stock. For a short stock position to provide a return of -100% (assuming a 50% margin) the stock price would have to increase by 50%.

2. We recognize that a lower per unit cost may allow the user of the strangle to diversify across several underlying securities. But strangles and straddles are event strategies that are unlikely to engender portfolio diversification issues.

3. Note that Rhoads labels the diagram as a payoff diagram instead of the more common usage where payoff diagram is reserved for a graph indicating intrinsic value of the strategy at expiration and where profit diagram, as is the case with this diagram, illustrates the profit subtracting the initial cost of the investment from the intrinsic value at expiration.

4. On occasion it was not possible to form a $10 spread for the strangle as a last price was not reported for either the call or the put at the desired strike price. In such cases we formed the strangle using a spread as close as possible to the $10 target.

5. If available we use the third Friday of January option expiration date. If this date is unavailable we use the closest possible January expiration date.

6. In four of the 36 trials the return to the strangle was greater than -100% but the return to the straddle was sufficiently positive so as to make the return difference over 100%.

References

Bouzaoubaa, Mohamed, and Adel Osseiran. 2010. *Exotic options and hybrids: A Guide to structuring, pricing and trading*. Hoboken, NJ: Wiley.

DraKoln, Noble. 2009. *Trade like a pro: 15 High-profit trading strategies*. Hoboken, NJ: Wiley.

Fontanills, George A. 2005. *Option course: High profit and low stress trading methods (2nd edition).* Hoboken, NJ: Wiley.

Pettengill, Glenn, Vijay Gondhalekar, and John Wingender Jr., "Teaching Comparative Risk of Put Buying vs. Short Selling," *Journal of Financial Education,* Summer 2015: 80-103.

Rhoads, Russell. 2010. *Option spread trading: A comprehensive guide to strategies and tactics*. Hoboken, NJ: Wiley.

Sinclair, Euan. 2010. *Option trading: Pricing and volatility strategies and techniques*. Hoboken, NJ: Willey.

Appendix

Sample Securities

Symbol	**Name**
AAPL	Apple Inc.
CAT	Caterpillar Inc.
DD	E.I. du Pont de Nemours and Company
DIS	The Walt Disney Company
JNJ	Johnson & Johnson
JPM	JPMorgan Chase & Co.
MSFT	Microsoft Corporation
NKE	NIKE, Inc.
PG	The Procter & Gamble Company
TRV	The Travelers Companies, Inc.
UTX	United Technologies Corporation
V	Visa Inc.
VZ	Verizon Communications Inc.
WMT	Wal-Mart Stores Inc.
XOM	Exxon Mobil Corporation

Towards A Global Framework for Impact Investing

Marc Sardy and Richard Lewin, Rollins College

Abstract

Recent trends suggest several trillion dollars are being directed into impact investing in enterprises of all forms. This is based on the active effort to improve investment prospects, not simply through corporate social responsibility but through a corporation's holistic approach to doing business. Our paper explores whether the investment community has discovered that there is a further dimension to which the traditional risk-reward tradeoff could be extended. We propose a framework that includes impact which may prove to shift the traditional risk-return tradeoff into a more optimal position. We examine the bottom line impacts of such a pioneering stakeholder model.

Introduction

Since 2008 the financial sector has taken a significant public pummeling as the stock market crash and great recession have raised serious doubts regarding the ability of financial markets to act as forces for social improvement. Many investors have also discovered that firms making impact investments have yielded higher returns (Snider, 2015). While investors have the greatest influence over the social, environmental and economic challenges of societies, they operate within a market infrastructure and investment ecosystem where the stock investment returns have been divorced from the social, environmental and economic impacts (Grace, Wood and Thornley, 2012). To redress this, impact investing is an emerging investment approach intentionally seeking to reconnect financial returns with positive social impact, which has the potential to reconcile such key shortcomings identified in traditional financial markets, (Sikken, 2011).

As institutional investors allocate ever more capital to impact investments that deliver both financial return and social and/or environmental enhancements, the need for a global standard has become paramount, as recently identified by the World Economic Forum Investors Industries report (2013). Impact investing, a broad scope of investment approaches that take 'impact' as a primary focus also targets segments of the economy typically under-served by traditional businesses, this merits the status of a new asset class for investors, (Brandenburg *et al.*, 2010). This approach has dual benefits, both improving the investment horizons by which fund managers and investors may act to nudge firms towards more sustainable business decisions and practices, whilst thereby generating more substantial long-term returns (Bugg-Levine and Emerson, 2011).

In section 2 we discuss some of the metrics currently used to measure impact. In section 3 we consider the traditional risk and return framework, and the potential modifications that highlight the effect of adding in an impact dimension. In section 4 we outline a framework for analyzing the interaction of risk, return & impact. In section 5 we explore other considerations including examining the extra value attainable over above the standard returns metric. In section 6 we conclude our paper.

Metrics Measuring Impact

Measuring impact is the distinguishing feature in assessing whether impact will have an effect beyond the risk & return framework. Thus the metric used to measure impact is of critical importance (Chung and Jed Emerson, 2013). To help standardize measuring and reporting, the Global Impact Investing Network (GIIN, 2013), a nonprofit organization dedicated to increasing the scale and effectiveness of impact investing, created Impact Reporting and Investing Standards (IRIS), a catalog of more than 400 generally accepted performance metrics (Achleitner *et al.*, 2011). Concrete social and environmental performance data, alongside financial performance data is needed in order to fully understand the performance of investees and to screen investment choices (E.T. Jackson and Associates, 2012).

IRIS serves as the taxonomy, or set of terms with standardized definitions, that govern the way companies, investors, and others define their social and environmental performance. Housed at GIIN it incorporates sector-specific best practices, is updated regularly based on user and expert feedback, and produces benchmark reports that capture major trends across the impact investing industry. PULSE Impact Investing Management Software is a portfolio management tool, administered by Application Experts (App-X), and is widely available to clients and comes pre-loaded with the IRIS metrics.

The Global Impact Investing Ratings System (GIIRS) is an impact ratings tool and analytics platform that assesses companies and funds on the basis of their social and environmental performance. It is based on IRIS definitions, and generates data that feed industry benchmark reports. These tools are also critical if the impact investing industry is to mature and have integrity around its dual value proposition, Clark, Emerson & Thornley (2012). Impact investing, which prioritizes positive social and environmental impact over investment returns (Freireich and Fulton, 2009), will see new capital inflows ranging from $0.5 to $2 trillion in the next 10 years, according to the JP Morgan 2014 report on 'Spotlighting the Market for Impact Investing' (Saltuk *et al.*, 2011, 2013). A firm's management may find that the bridge to understanding impact investing requires

them to think about it in an unconventional way. A degree of commitment is required by those investors intentionally looking to allocate capital towards impact investments. The intended goal of these efforts will be to continue to move impact investing from the margin and into the mainstream (Grace *et al.*, 2011).

Another methodology has been established by the Sustainability Accounting Standards Board (SASB, www.sasb.org). The objective of SASB is to establish metrics for measuring the sustainability practices of firms. Data is collected on firms using SIC and industry sub-classifications. Firms can be compared within their industry peer group and their sustainability practices then become available to investors who consider these data important to their investment practices.

Yet a third method is one developed by R. Paul Herman as a measure of High Impact investing practices (HIP). His methodology includes 5 measures of impact; health, wealth, earth, equality and trust (Herman, 2010). These metrics are pulled from a combination of direct measures from financial reports and from other softer measures garnered from reports and articles of the firm. Herman has shown that HIP firms have regularly outperformed the S&P index and have also provided stakeholder value beyond the basics of the risk-return framework.

International Impact Investing

While most of the measurement systems have been developed largely for the analysis of equities, impact investing has taken other forms in different countries. A UK Social Impact Investment Taskforce, set up under the UK's presidency of the G8 in 2013, set out to see how investments made into businesses, non-governmental organizations (NGOs) and charities can generate a measurable social return, as well as a profit (Cohen, 2013). Impact investing may lead the way, even in cash-strapped times, to solve some of society's most entrenched issues while also allowing investors to also collect a dividend. The Social Impact Investment Taskforce Report in September 2014 announced that $1tn (£615bn) of social investment funds could be unlocked around the world; it seems that there is global awareness and a welcome impetus to the idea that the power of enterprise that can be harnessed to benefit - rather than hinder - society as a whole, without sacrificing investor returns.

The concept of social investing has gathered pace in times of austerity as governments struggling to cope with failing healthcare systems, poverty, crime and poor education increasingly look to the private sector, with around $50bn in social investments already under management globally. However these fall more in the NGO sector as opposed to the mainstream financial sector. These include an incredibly diverse mix of investing opportunities, including microfinance, affordable housing development and renewable energy finance, to name just a few (Bannick and Goldman, 2012). Evidence to date shows that impact bonds have high yields and they are often less risky (Gustafsson-Wright *et al.*, 2015).

The UK has been a frontrunner in this area, pioneering the first-ever social impact bond (SIB) in 2010, in a pilot scheme to reduce reoffending amongst prisoners in Peterborough. A SIB is an innovative financial tool that enables government agencies to pay for programs that deliver results, utilizing a payment-by-results model, whereby private investors fund preventative social projects, usually aimed at improving the lives of at-risk individuals, and paying those investors back - with interest - from public funds if these targets are met. Underpinning this is the idea that successful projects will cut government spending over the long term. Thus profitability can be directly tied to a reduction in government spending in ways that decrease the government portion rather than expand it.

In the UK, for example, a youth offender is estimated to cost the state around £21,000 a year, while a successful project to reduce recidivism means that this could be reduced to as little as £7,000. The £5m of private money invested in the Peterborough SIB cut reoffending rates by 8.4% compared with national averages. Investors in the second phase, in this instance a number of charitable foundations, will receive their money back and a small return should the hurdle rate of a 7.5% reduction be beaten (Cohen, 2014). The applications, however, can be extended much further than criminal justice. The Think Forward scheme in East London provides educational support to almost 1,000 vulnerable teenagers, meaning 60% of the group now achieve five A*- C grade General Certificates of Secondary Education (GCSEs). By improving their employment prospects, young people are now much less likely to end up on unemployment benefits.

Beyond the UK, SIB programs have been announced in Australia, Belgium, Canada, Germany, Holland, and across the US. Social Impact Bonds (Bouri *et al.*, 2013) are thus an arrangement between one or more government agencies and an external organization where the government specifies an outcome (or outcomes) and promises to pay the external organization a pre-agreed sum (or sums) if it is able to accomplish the outcome(s), (Rangan, Appleby and Moon, 2012). The number of projects in the US will expand considerably if Congress passes the Social Impact Bond Act, which will grant a further $300m to SIB projects, (Olson and Philips, 2013).

Theoretical Framework

Traditionally firms face a trade off in terms of risk & return for projects they undertake. But are these the only options to enhance stakeholder returns? Impact investing suggests there may be an alternative third dimension, not previously addressed

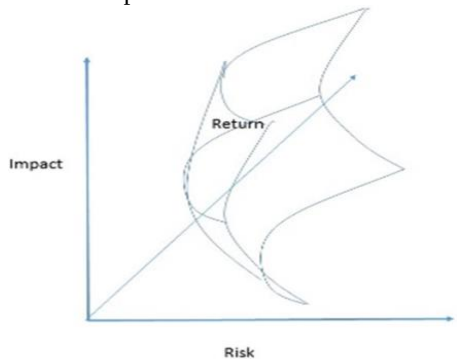
within the literature (Palandjian, 2010). What would happen if a vertical line was extended out of the traditional 2 dimensional risk-return frontier curve? It could be interpreted that this would look rather like a sail in the vertical plain, on the same basis as risk-return, but instead of being vertically straight (implying no 'impact' enhancement to the previous risk and return graph) a bell-shaped supply curve convex towards the origin as the curve moves to the left, may allow the risk-return tradeoff to be optimized through impact investment.

Traditional theory and analysis says that this should not happen. Why? Let's dig deeper into what happens when we make an impact investment to say cut costs at a traditional company such as Dow Chemical. Historically, the firm wasted chemicals in its manufacturing process, and in consequence polluted the environment. When Dow Chemical decides to make an impact investment to reclaim, recycle and reduce such pollution, they actually lowered the ongoing cost of materials used in production, discovered new revenue streams for the chemicals that were previously discarded and thereby increased their internal profitability and competitiveness, whilst lowered the risk that might have come from a potential future environmental scandal. In effect this investment in cost cutting leads to a firm being both more environmentally responsible and in the process developing a more profitable, sustainable, less wasteful and less risky entity, thereby raising investors returns and lowering risk concomitantly. One could say that the impact investment shifted the traditional risk-return curve to the left.

Impact investors should be looking at efforts that induce the risk-return curve to shift towards the axis (or for the sail shape to billow inwards towards the axes); understanding implicitly or explicitly that the firm making such an effort can change its underlying cost, profitability and risk curves. Conceivably a vertical representation of impact investing, could provide capital to a firm that is making investments that may not be possible otherwise based on the conventional framework. The impact investor's profit maximization actually becomes a cost minimization or externality minimization intervention.

Imagine Apple Inc. decides to opt for impact investments which increase battery life and consume less power, lower associated costs while decreasing the carbon footprint of Apple users worldwide. Or another investment which uses recycled materials and thereby lowers material costs. Apple is more environmentally conscious through reducing operational costs and would thus be shifting the cost curve of the organization down, thereby making it even more financially successful and an even greater value proposition to its investors.

Figure 1: Relationship between risk, return and impact



In effect by doing so additional capital would be invested, and outcome that would not otherwise transpire. The asset-liability equation clearly increases equity and increases assets, not just in physical plant, but also in the liquid assets which provide bigger and better efforts for ongoing cost minimization, rather than cost maximization. For example, Walmart introducing solar panels across the roofs of their stores, thereby lowering ongoing operational costs. Such initial capital expenditure is higher initially of course, but this is tempered by the long-term return on the investments made via lowered associated operating costs, as well as providing long term impacts to reduce CO_2 emissions. As traditional retailer margins have suffered due to the internet economy, these types of ongoing operating costs reductions will become more pressing to restore the balance and sustainability of profitability.

Impact investors typically take a longer-term view; thus these investors provide lower capital costs to businesses, leading to more profitability and more sustainable, higher returns, relative to the risk profile of the firm. In effect, the impact curve should shift to the left with the graph curving in two dimensions, but with greater impact from cumulative investment, the graph itself would billow in. If impact does not increase incremental investment, an interesting argument which is economics and finance associated, is asking whether an impact investment by a firm leads to enhanced profitability? If impact investors lead to a more profitable valuation of the firm, such that more impact can be undertaken, then this should become a virtuous cycle. What happens if firms can improve their competitive position through impact investment capital from an investor with a more 'patient' capital objective, is that investment that might not otherwise be forthcoming in a capital rationing framework becomes feasible.

The curve may drift backwards in smaller firms or social entrepreneurs, as they look to individually make a change in the way that industry has executed previously as they do not just see value creation but also the potential to do social good. Initially it may seem as pursuing a risky path for the firm, to some extent, as it may even seem to lower potential immediate returns although the impact itself may ultimately lead to a much higher permanent shift to the left of the risk-return frontier as it moves up the impact curve, which would not be otherwise captured under conventional analysis.

The traditional model is thus too static in respect of risk and return, as no impact measures of internal or external stakeholders are to be considered. Spending on something does not raise profit and instantaneously lower cost in an increasingly competitive environment, however impact investors have already seen a new way to win the battle through impact investment. This tighter economic argument is potentially generalizable globally.

Other Considerations

A measure for impact measurement, the impact of every firm can be found when you look at the iso-curve in all possible combinations, but what would that look like? Vertically a bizarre surface with billowing sail-shaped peaks and valleys. GIIRS is one such methodology already provided, but unfortunately they are unwilling to share their assessment metrics. Paul Herman uses health, wealth, earth, equality and trust as five part sections of a score for each composite of 20 points. High HIP total scores demonstrate firms driving the curve further inwards by yielding greater returns.

The traditional profit formulation is:

$$\Pi = P - \sum_{j=1}^{n} C_j \tag{1}$$

Where costs "Cj's" are explained by:

$$Cj = \sum_{j=0}^{n} VCj + FCj + OCj \tag{2}$$

Where variable costs VCj are the sum of the Labor costs LCj, Materials costs MCj and can be reduced by impact investment in variable costs:

$$VCj = \sum_{j=0}^{n} Li + Mi - Ivc \tag{3}$$

Fixed costs FCj are the sum of the fixed labor costs LCj, fixed overhead costs OMj and can be reduced by impact investment in fixed costs:

$$FCj = \sum_{j=0}^{n} Li + OMi - Ifc \tag{4}$$

Other costs OCj are the sum of the fixed labor costs KCj, fixed other capital costs Kj and can be reduced by impact investment in other costs:

$$OCj = \sum_{j=0}^{n} Li + Ki - Ifc \tag{5}$$

The cost reducing sum of the impact investment is:

$$Impact\ Investment = \sum_{j=1}^{n} Ivc + Ifc + Ioc \tag{6}$$

Costs associated with products or services that we denominate as C, where all components that decrease any one or group of c's, creating an overall reduction in aggregate C. The sum of all of the impact investments thus reduces the overall cost function and thus the volatility of profits. Lower carbon foot print overall may also lower costs thereby providing a better value proposition with longer life battery being of more benefit to the consumer. Cost per charge, time per charge is affected with

costs associated overall lower in our analysis. Higher costs per hour, lower life battery cheaper materials lighter, lower cost process of doing several interventions simultaneously lower costs product or with the same cost yields greater benefits as measured as a change in the demand curve versus others competitors. Cheaper in terms of costs if more environmentally responsible. Optimize how much utility profit margin on it on other hand in a competitive environment lower price more competitive with what is out there without affecting profit market. Thus to the degree that profits improve returns of the firm with respect to the market:

$$R_i = \beta \mathrm{e}_i + \varepsilon \tag{7}$$

Where R_i represents the returns of firm "i" and β represents the relationship between the firm's returns and the earnings of that firm "i", and ε represents the unexplained variation between earnings and returns. To the degree that β also represents the variance of earnings and covariance of earnings and returns the relationship between R & e will decline as the variance and covariance converge. As we know from the CAPM literature the lower β will also lower risk of the firms returns. Thus as impact investment increases earnings will increase and risk will decline shifting the risk return tradeoff closer to the origin.

What if impact investors note that there is extra value attainable over above the standard returns metric which implies that the firm may be currently under investing? A measure for the impact tradeoff requires a framework and measurement mechanism to verify for the sway of impact investors to what degree is an impact investment performing. Impact investing drives up cost services in the short term but sub externalities will benefit consumers. At issue is that to take the initial increase in investment, most non-impact investors may perceive value destruction based management. Not recognizing that the traditional cost curve is growing.

More savvy investors may direct more capital in line with direct evidence to date of a higher returns model as alluded to by Hope Consulting (2010), who illustrate the increases in the value of utility as perceived by the consumers in respect of firms 'doing good as well...' If however connection to impact leads to better results, then cost may be a function of the sub components of investment. Investment may be used to minimize the cost function in a type of mini-max solution. Impact measures in terms of ultimately higher returns and lower risk are expressed on the plane of the three dimensional curve surface of risk, return & impact and may be well above and bellied out along the curve for firms undertaking impact investments. Walmart for example has an outsize effect due to the footprint of impact, smaller than large Fortune 500 companies' also have the ability to have a large impact due to impact scale of the business.

By putting investment above and beyond what is normally allocated may actually raise long-term profitability, thereby attracting further investment funds through a high impact approach that is not only sound investing but will also yield lower inherent risk. This may sustainably increase long term profitability and thus is not a violation of a firm's fiduciary responsibility to shareholders. Traditionally arguments are made that any manager of a for profit firm who is paying too much attention to perceived "peripheral" concerns, should be fired because he is not doing what is in the company's best interest (maximizing shareholder returns). Impacting investing suggests this position needs reviewing in the light of the tradeoff between short and long-term profitability measures.

The old school way of thinking thus only values a firm based on shareholder value, but in Europe for example boards are already required to reflect stakeholder values, not exclusively shareholder values. Whilst purists might argue that they are being frivolous with these funds, it turns out that making precisely those kinds of investments is effectively like taking out a type of insurance for the risk of the sustainability of the business model. Like old fashioned currency hedging; if you don't hedge and the market turns against you, you may look foolish. Investment is required to protect the firm against currency risk typically via a hedging strategy such as the use of forwards or options. This investment is never questioned in the traditional finance framework even though short term costs are borne by the organization, but these are justified on the rational basis of uncertain long-term benefits resulting in the stability of the firm. Impact investing lends itself to this analogy in that investments may lead to improved operating parameters for organizations and therefore form a legitimate part of the arsenal of management techniques to enhance enterprise value alongside external benefits sought in the wide community.

As one of the first banks in the United States which has an environmental and social mission, First Green Bank was founded in 2009 by a team of experienced bankers after being given the last bank charter in the state of Florida. First Green Bank focuses on making loans to people for impact investments and making impact investments themselves in their own operational model. From outside a typical branch the bank appears to be located in a good building, but competitor banks also run these type of buildings at whatever their cost of operation. First Green bank however has adopted a different mantra by commissioning the first Platinum LEED status bank building which aims to reduce long-term cost savings. First Green Bank made investment beyond the conventional building cost assessment by uprating insulation and installing solar paneling on the building to provide for the highest LEED certified building status. This investment has subsequently resulted in a more profitable operational model as the bank now has structurally lower operating costs than its competition and as such its inherent long-term profitability is higher.

Indeed First Green bank attained profitability in a remarkably short period (a mere 19 months from inception), with current assets over $292 million as of December 2014, and it has lent in excess of $251.5 million to local businesses and people in the community. Moreover First Green Bank has received the coveted 5-star Bauer Financial rating from the nation's leading independent financial institution rating service. First Green Bank adheres to a values-based business model which endeavors to do the right thing for the environment, community & shareholders. They offer a 'never pay for power again' investment facility through their residential solar program. This Solar Loan program offers a great, long-term fixed rate to encourage customers and employees to install solar panel systems for reducing energy use.

First Green Bank offers discounted interest rates for commercial and residential projects that also meet green building criteria of Leadership in Energy and Environmental Design (LEED) certification by the U.S. Green Building Council. As a non-profit the sponsored 'First Green Foundation' provides assistance to community members for installation of solar panel systems in addition to providing assistance to community supported agriculture, such as local start-up farming projects as well as projects that better manage scarce water resources.

Conclusions

Impact investment is not a silver bullet but there may be huge potential for benefits for both investors and society. Social investment to do some good in the world is an industry that could potentially unlock trillions of dollars in pursuit of positive social and environmental impact as well as financial return. Metrics play a critical role in distinguishing good companies from good marketing, and thus enable management, investors and other stakeholders to judge performance objectively and inform decisions on the basis of social and environmental impact in addition to profitability. This is particularly critical for impact investments, as they are by definition, designed to generate impact beyond financial return. Impact ratings (or performance standards) for asset managers and owners, who reported lacking the tools needed to assess their pipeline and active portfolios on the basis of non-financial performance require standardized definitions of impact performance measures.

We have outlined an investment approach that intentionally seeks to create both financial return and positive social or environmental impact that can be actively & accurately measured. A framework for measuring impact (in-line with the IRIS standards, and in GIIRS ratings) but with an agreed-upon standard of what social impact data would be collected by impact funds. An impact investing sector without agreement on what constitutes impact and what minimal data should be collected by impact funds is inherently handicapped making it inaccessible for mainstream investors which might otherwise provide an avalanche of additional investment capital were the story properly told. We conclude that the most important outcome is universal, quantifiable data that will provide one-level-deeper insights to balance the scales and thus be weighed by investors against purely financial return metrics.

References

Achleitner, Ann-Kristin, Andreas Heinecke, Abigail Noble, Mirjam Schöning and Wolfgang Spiess-Knafl. 2011. "Social Investment Manual: An Introduction for Social Entrepreneurs", *Technical University of Munich and Schwab Foundation for Social Entrepreneurship*, September, Available at SSRN: http://ssrn.com/abstract=1884338.

Bannick, Matt and Paula Goldman. 2012. "Priming the Pump: The Case for a Sector Based Approach to Impact Investing", *Omidyar Network, September*.

Bouri, Amit, Abhilash Mudaliar, Min Pease and Yasemin Saltuk. 2013. "Perspectives on Progress: The Impact Investor Survey", J.P. Morgan and the Global Impact Investing Network.

Brandenburg, Margot, Nick O'Donohoe, Christina Leijonhufvud, Yasemin Saltuk and Antony Bugg-Levine. 2010. "Impact Investments: An Emerging Asset Class", J.P. Morgan, Rockefeller Foundation, & GIIN.V.

Bugg-Levine, Antony and Jed Emerson. 2011. *Impact Investing: Transforming How We Make Money While Making a Difference*, Jossey-Bass (San Francisco, CA), John Wiley & Sons.

Calvert Foundation. 2012. "Gateways to Impact: Industry Survey of Financial Advisors on Sustainable and Impact Investing", (Href: http://www.calvertfoundation.org/storage/documents/Gateways-to-Impact.pdf, last viewed on 4/28/2016).

Chung, Amy and Jed Emerson. 2013. "Grants to Groundbreaking: Unlocking Impact Investments", Living Cities, Impact Assets Issue Brief #10.

Clark, Catherine H., Matthew H. Allen, Bonny Moellenbrock, and Chinwe Onyeagoro. 2013. "Accelerating Impact Enterprises: How to Lock, Stock and Anchor Impact Enterprises for Maximum Impact", CASE i3 at Fuqua School of Business, SJF Institute, and O-H Community Partners Investors' Circle conference, San Francisco.

Clark, Catherine H., Jed Emerson, and Ben Thornley. 2012. "A Market Emerges: The Six Dynamics of Impact Investing", Pacific Community Ventures, CASE at Duke University's Fuqua School of Business, and Impact Assets.

Cohen, Ronald. 2014. 'Impact Investment: The Invisible Heart of Markets, Harnessing the power of entrepreneurship, innovation and capital for public good, *Social Impact Investment Taskforce* September.

E.T. Jackson and Associates Ltd. 2012. "Accelerating Impact: Achievements, Challenges and What's Next in Building the Impact Investing Industry", (Href: http://www.rockefellerfoundation.org/blog/accelerating-impact-achievements last accessed 05/14/16).

First Affirmative Financial Network, LLC. 2012. "2013 To Be the Year of 'Impact Investing'" (Href: http://www.firstaffirmative.com/resources-news/news/2013-to-be-year-of-%E2%80%9Cimpact-investing/ last accessed 2/2/2016)

Freireich, Jessica and Katherine Fulton. 2009. "Investing for Social & Environmental Impact", Monitor Institute.

Global Impact Investing Network (GIIN). 2013. IRIS Data Brief: Focus on Employment (Href: http://www.thegiin.org/cgi-bin/iowa/resources/research/527.html last viewed 2/2/2016).

Grace, Katie, Ben Thornley, David Wood and Sarah Sullivant. 2011. "Impact Investing: A Framework for Policy Design & Analysis", Pacific Community Ventures, Institute for Responsible Investment (Harvard) & Rockefeller World Economic Forum.

Grace, Katie, David Wood and Ben Thornley. 2012. "Impact at Scale: Policy Innovation for Institutional Investment with Social and Environmental Benefit", *Pacific Community Ventures*, Institute for Responsible Investment at Harvard University, and Rockefeller Foundation.

Gustafsson-Wright, Emily, Sophie Gardiner and Vidya Pucha. 2015. "The Potential and Limitations of Impact Bonds", Brookings Institute.

Herman, R. Paul. 2010. *The HIP Human Impact & Profit Investor, Make bigger profits by building a better world*, John Wiley & Sons (Hoboken, NJ).

Hope Consulting. 2010. "Money for Good: Impact Investing Overview" (Href: http://www.hopeconsulting.us/moneyforgood. last accessed, 05/14/16).

ImpactBase: Global Online Directory of Impact Investment Vehicles 2013. (Href: http://www.impactbase.org, last accessed 5/14/16).

Karamchandani Ashish, Harvey Koh and Robert Katz. 2012. "From Blueprint to Scale: The Case of Philanthropy in Impact Investing", Monitor Group, (Href: http://acumen.org/idea/from-blueprint-to-scale/ last accessed 5/14/2016).

Lilly Family School of Philanthropy, 2013. "Leveraging the Power of Foundations: An Analysis of Program-Related Investments", Indiana University (Href: http://www.philanthropy.iupui.edu/research-by-category/PRIreport last viewed 3/2/2016).

Olson, John and Andrea Phillips. 2013. "Rikers Island: The First Social Impact Bond in the United States", *Community Development Investment Review*, Federal Reserve Bank of San Francisco.

Palandjian, Tracy P. 2010. "Investing for Impact: Case Studies Across Asset Classes", The Parthenon Group; Bridges Ventures and the Global Impact Investing Network, March.

Rangan, V. Kasturi, Sarah Appleby, and Laura Moon. 2012. "The Promise of Impact Investing." *Harvard Business School*, Background Note 512-045, February (Revised July 2012).

Saltuk, Yasemin, Amit Bouri and Giselle Leung. 2011. "Insight into the Impact Investment Market: An In-Depth Analysis of Investor Perspectives & Over 2,200 Transactions", *Global Social Finance*, JP Morgan and the Global Impact Investing Network.

Saltuk, Yasemin, Ali E. Idrissi, Amit Bouri, Abhilash Mudaliar and Hannah Schiff. 2013. "Spotlight on the Market: The Impact Investor Survey", *Global Social Finance*, J.P. Morgan Securities plc and Global Impact Investing Network.

Sikken, Bernd Jan. 2011. "Accelerating the Transition Towards Sustainable Investing - Strategic Options for Investors, Corporations, and Other Key Stakeholders", World Economic Forum Paper, SSRN: http://ssrn.com/abstract=1891834.

Snider, Anna. 2015. "Impact Investing: The Performance Realities" *Wealth Management Institute*, Merrill Lynch Global Wealth and Retirement Services, November.

World Economic Forum Investors Industries. 2013. "From the Margins to the Mainstream: Assessment of the Impact Investment Sector and Opportunities to Engage Mainstream Investors", Industry Agenda, Deloitte Touche Tohmatsu.

Another Clue in the Market Efficiency Puzzle

Karen Sherrill, Sam Houston State University

Abstract

The Efficient Market Hypothesis states that all publicly available information is immediately incorporated into stock prices and that there is no opportunity to generate a profit by trading on this information. This assertion has been difficult to test and there exist studies with findings that both prove and disprove this hypothesis. I add to the literature by examining investor reaction to the same information set disseminated twice on different dates and in different formats. I find that positive news is immediately incorporated into the stock price, however negative news is not immediately incorporated and the second announcement of the same data causes a second stock price reaction.

Introduction

The Efficient Markets Hypothesis (EMH) states that all publicly available information is assimilated into the markets immediately and therefore it is not possible to earn abnormal returns by trading based on this information. Jensen (1978) further defines the EMH as, "A market is efficient with respect to information set θ_t if it is impossible to make economic profits by trading on the basis of information set θ." If we take θ_t to be all publicly available information then there is a large volume of work showing support for this theory. In fact, Jensen (1978) concludes that no hypothesis has more solid empirical evidence behind it and that previous studies using data from a variety of markets consistently prove this theory to be correct. This is also supported by Fama (1970) who reviews the existing literature and splits the EMH into three segments, a random walk, a semi-strong form, and a strong form. He defines the semi-strong form of market efficiency as being markets where all available public information is assimilated into prices immediately. He reviews the existing literature and finds overwhelming support for this hypothesis.

However, Brown (1978) finds that the EMH does not always hold. He notes, using a CAR based methodology that the markets fail to adjust immediately to new information on earnings per share. He further finds that contrary to Jensen (1978) one can indeed make a profit on these inefficiencies as the excess returns are greater than the transaction costs. This is consistent with Jones and Litzenberger (1978) who find that the market does not adjust immediately to public information. They also conclude find that public information is not correctly priced by investors.

There are also contradictions regarding the EMH in the behaviorist's literature. Barberis, Shleifer and Vishny (1997) present a model where news is incorporated slowly into prices. Their model shows that whether the news is good news or bad news the investor underreacts which leads to autocorrelation over time. In other words, good news leads to long term positive returns and bad news leads to long term negative returns.

Hong, Lim, and Stein (2000) hypothesize that bad news is incorporated more slowly than good news. They suggest that this is particularly relevant for firms with less analyst coverage, which are generally the firms of smaller size. The managers of these firms will be happy to divulge any positive news and therefore push it out to the public. They may not be as eager to share negative news, therefore they do not publish it. This leads to positive news being assimilated into prices more rapidly as there is more information available about the positive news, while the negative news is assimilated more slowly as there is less information about this.

Ie add to this literature with a unique approach. I examine an information set that is made available to the public in two different formats on two different dates. The data is the same. If the EMH holds true then I expect that once the information is disseminated the first time, the information will be assimilated into stock prices and the second announcement to the public of the same data will have no impact on stock prices. These findings are consistent with Hong, Lin, and Stein (2000). I find that the EMH holds true for positive news but negative news is not immediately assimilated into the stock price and there is an opportunity for investors to take advantage of this inefficiency in the market.

The unique situation that I examine is around earnings announcements. Firms are required to issue quarterly and annual earnings report and the data is made available to the public on the earnings announcement date on the SEC EDGAR (electronic gathering, analysis, and retrieval system) website.

Finance.yahoo.com is a popular site for investor. Finance.yahoo.com contains up to the minute stock market quotes as well as historical data, analyst opinions, relevant firm specific data and news. This includes copies of the firms' financial statements which finance.yahoo.com pulls from the SEC website via an automatic feed and populates the Yahoo website with the information in an easy to read format. Finance.yahoo.com does not add any new information, they simply post the financial statements at a date later than the announcement date. This Finance.yahoo.com posting (hereafter referred to as the Yahoo date) can be as soon as two trading days after the earnings announcement date or as much as thirty trading days after the earnings announcement date. The average is fourteen trading days. When the information is available on the Finance.yahoo.com site, a news announcement is included on the firm's main Finance.yahoo.com page alerting investors to its availability.

Thus when investors search a specific stock on finance.yahoo.com they will see the headline stating that the financials for that firm are available. The investor can then examine the financial statements from the finance.yahoo.com web page either by selecting the statement they wish to see from the menu bar on the left of the page under the heading Financials or by clicking on the news headline itself. The income statement, balance sheet, and statement of cash flows are all available as separate web pages.

This provides an opportunity to study the impact of the same information disseminated to the public on two different dates, in two different formats, by two different venues. If the EMH is true, then there should be no abnormal returns in the stocks at the time of the posting on the Finance.yahoo.com site since this is information that investors have already been made aware of via earnings announcements and it is available on the SEC's EDGAR website. If the market is semi-strong form efficient then this information should already be incorporated into the stock price and no further opportunity to generate an abnormal profit from this information should be available to investors. There should be no abnormal returns after the Yahoo date. If the EMH is not true, then the information posted by Yahoo should generate abnormal returns.

I find that the firms that have a negative abnormal return after the initial earnings announcement also have a negative and statistically significant abnormal return after the Yahoo data is posted. However, firms that have a positive abnormal return after the initial earnings announcement do not have a statistically significant abnormal return after the Yahoo data is posted.

These results suggest that the markets incorporate positive information in an efficient manner but negative information is not efficiently incorporated. This gives investors an opportunity to take advantage of market inefficiencies by short selling stocks that have negative abnormal returns after their earnings announcements.

Data

I use five hundred random firms and searched Finance.yahoo.com to find the date when the Edgar online data became available at Finance.yahoo.com. I then searched the SEC website to find when the firms actually filed their annual or quarterly report to get that data's availability date. I obtain return data for each of the firms from Finance.yahoo.com from the period before the initial earnings announce to after the Yahoo data is posted.

The firms have to be U.S firms, and have to have return information and price information in CRSP for twenty days prior to the earnings announcement date. This gives me a sample size of 447 firms. The mean size of the firms in my sample is significantly larger than the mean size of firms in the market in general. The Hong, Lim, and Stein (2000) hypothesis states that firms with less analyst coverage will disseminate negative information more slowly. Smaller firms generally have less analyst coverage. Therefore, having a sample that has a greater mean size than the population ensures that I am not capturing a phenomenon seen only in small stocks.

Methodology

I calculate a three day cumulative abnormal return (CAR) for the earnings report date. The three days includes day 0 which is the day the earnings information is made available to the public, trading day 1, and trading day 2. I determine the normal return for each stock by calculating an average return for the stock for twenty trading days prior to the earnings announcement date. The abnormal return is then the return for day 0 minus the average return, the return for trading day 1 minus the average return, and the return for trading day 2 minus the average return.

I do the same calculation for the Yahoo data. We calculate a three day CAR using day 0 which is the day that finace.yahoo.com posts the data from the EDGAR website on the Yahoo site, trading day 1, and trading day 2.

The CAR is simply the sum of the abnormal returns over the three day period, either the three days after the earnings announcement or the three days after the Yahoo data is available. The three day CAR is the mean of the summed abnormal returns for all of the sample firms, shown in equation 1.Here, t_1 represents the first day of the event and t_3 represents the final day of the event. N is the total number of firms. CAR_i is the CAR for firm *i*.

$$\overline{CAR}\,(t_1, t_3) = \frac{1}{N}\sum_{i=1}^{N} CAR_i\,(t_1, t_3) \qquad (1)$$

To check for robustness, I do the same calculations but use a different estimation window for the normal results. I take the average return for ten days prior to the earnings announcement date, however I include a five day window prior to the earnings announcement date which I do not use in the estimation. Thus our normal return is calculated as the average of trading days -15 to -6. Trading days -5 to -1 are the five day window prior to the earnings announcement that I do not use.

I calculate the abnormal return as the return on day 0 minus the average of the ten days in the estimation window plus trading day 1 minus the average of the ten days in the estimation window plus trading day 2 minus the average of the ten days

in the estimation window. I calculate a three day CAR after the earnings announcement date and after the Yahoo date in the same manner as described earlier.

Since earnings announcement drift is a documented event, I further test the robustness of the results by requiring that there be at least seven trading days between the earnings announcement data and the yahoo date so that I do not capture earnings announcement drift in the post yahoo date three day CAR. I further test this by requiring that there be at least fourteen trading days between the earnings announcement date and the yahoo date.

When requiring that there be seven or fourteen days between the two events, I use the first definition of normal, the average of the returns for each stock for twenty days prior to the earnings announcement date.

I split the data into two sets; firms that have positive CARs after the earnings announcement and firms that have negative CARs after the earnings announcement. I make the assumption that the positive abnormal returns following an earnings announcement is indicative of god news and that a negative abnormal return following an earnings announcement is indicative of bad news. This is not based on the actual earnings announcement content per se, but is based solely on investors' reactions to the news.

I calculate a mean and median three day CAR for both the earnings announcement data and the Yahoo data for each data set, positive and negative. I determine the statistical significance of the mean by using a t-test. I determine the statistical significance of the median using a signed rank test (M test in SAS).

Results

The primary results are shown in TABLE 1. There are 207 firms that have positive abnormal returns after the earnings announcement. The mean abnormal return for these firm is 11% and the median is 5%, both statistically significant at the 1% level. However, when I examine these firms after finance.yahoo.com posts the same data I find that the abnormal return is not positive nor is it statistically significant whether I use the mean or the median results.

Table 1: 3 Day CAR data

Panel A: 3 Day CAR data based on earnings date

	N	***Mean***	***T-stat***	***p- value***		***Median***	***Sign M***		***Standard deviation***
Positives	207	0.1102	3.85	0.0002	***	0.0469	<.0001	***	0.4114
Negatives	240	-0.0709	-12.81	<.0001	***	-0.0452	<.0001	***	0.0857

Panel B: 3 Day CAR data based on Yahoo date

	N	***Mean***	***T-stat***	***p- value***		***Median***	***Sign M***		***Standard deviation***
Positives	207	-0.0030	-0.55	0.5828		-0.0087	0.2661		0.0781
Negatives	238	-0.0154	-2.71	0.0072	***	-0.0198	<.0001	***	0.0875

* significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level

But when I examine the firms with negative abnormal returns after the earnings announcement I find a different result. After the earnings announcement the firms show a mean abnormal return of -7% and a median abnormal return of -5% both of these statistically significant at the 1% level. Examining the abnormal returns after finance.yahoo.com posts the exact same data as that which was announced in the earnings report I find that both the mean and median abnormal returns are -2% and these are both statistically significant at the 1% level.

It is possible that I am picking up post earnings announcement drift in the 3 day CAR after the Yahoo date. I do a couple of robustness checks to determine if this is the case. The first test is to only use data for firms that have at least seven trading days between the earnings announcement date and the Yahoo date. The results for this analysis are shown in Table 2. The results are consistent with the results for the entire sample.

The positive firms have a mean 14% abnormal return for the 3 day CAR after the earnings announcement. The median abnormal 3 day CAR is 6%. Both of these results are statistically significant at the 1% level. The mean abnormal return after the Yahoo data is made available is -1% and is not statistically significant. The median 3 day CAR is -2% and is significant at the 1% level.

However, once again we see different results for the firms with negative abnormal earnings announcement returns. The mean 3 day CAR is -7% for the earnings announcement and the median is -5.5% both statistically significant at the 1% level. The mean 3 day CAR is -2% for the yahoo data and the median is -3%, both statistically significant at the 1% level. Again this suggests that the market is not efficient at assimilating negative news.

Table 2: 7 Trading days between events

	N	Mean	T-stat	p- value		Median	Sign M		Standard deviation
Panel A: 3 day CAR data based on earnings date									
Positives	111	0.1420	2.70	0.0081	***	0.0581	<.0001	***	0.5552
Negatives	141	-0.0681	12.17	<.0001	***	-0.0551	<.0001	***	0.0664
Panel B: 3 day CAR date based on Yahoo date									
	N	Mean	T-stat	p- value		Median	Sign M		Standard deviation
Positives	111	-0.0082	-0.97	0.3319		-0.0228	0.0076	***	0.0882
Negatives	141	-0.0223	-3.16	0.0019	***	-0.0275	<.0001	***	0.0839

* significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level

We do this same type of analysis requiring 14 trading days between the earnings announcement date and the Yahoo date as a further robustness check. This reduces our sample size but once again we find the same results. The positive news is incorporated efficiently by the market but the negative news is not. The results are shown in Table 3.

Table 3: 14 trading days between events

	N	Mean	T-stat	p- value		Median	Sign M		Standard deviation
Panel A: 3 day CAR data based on earnings date									
Positives	94	0.0917	8.29	<.0001	***	0.0586	<.0001	***	0.1072
Negatives	122	-0.0669	-12.10	<.0001	***	-0.0550	<.0001	***	0.0610
Panel B: 3 day CAR date based on Yahoo date									
	N	Mean	T-stat	p- value		Median	Sign M		Standard deviation
Positives	94	-0.0026	-0.27	0.7848		-0.0217	0.0298	**	0.0909
Negatives	122	-0.0149	-2.07	0.0405	**	-0.0241	<.0001	***	0.0794

* significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level

The positive earnings announcement data associated with a positive 3 day CAR has a mean 3 day CAR of 9% and the median is 6%, both statistically significant at the 1% level. The 3 days CAR after the Yahoo date is negative and only the median is statistically significant. However, the negative news associated with a 3 day CAR after the earnings announcement is negative has a mean of 7% and a median of 6%, both statistically significant at the 1% level, and the 3 day CAR after the Yahoo date is also negative and significant with a mean of -1% and a median of -2% both statistically significant.

I do one further robustness check by changing our measure of abnormal returns. Instead of using the average of the firm's stock price twenty trading days before the earnings announcement date I use an estimation window that begins fifteen trading days prior to the earnings announcement date and ends five trading days prior to the earnings announcement date. This provides ten trading days of data to obtain an average stock price and excludes the trading days immediately prior to the earnings announcement date as these days may include positive (or negative) run up based on expected news. These results are shown in Table 4.

The results are consistent with the previous results. For the positive news I have significant positive CARS after the earnings announcement, 11% as the mean and 4% as the median and both are statistically significant, but negative CARS after the Yahoo date. The mean is -1% and is not statistically significant and the median is -.5% and is only significant at the 10% level. The negative news however, continues to have significant CARS after both the earnings announcement and the Yahoo date. The earnings announcement CARS are -7% mean and -4% median both significant at the 1% level. The Yahoo date CARS are -1% for both the mean and the median and again both are statistically significant at the 1% level.

Conclusion

These findings suggest that the markets are not always efficient. The same data presented on a different date in a different manner should not produce abnormal returns for investors. Yet I find that if the initial news causes a negative stock price

reaction, then announcing the same news at a subsequent date produces a second negative stock price reaction. These results are inconsistent with the semi-strong form of the Efficient Markets Hypothesis.

Table 4: Different estimation window

Panel A: 3 day CAR data based on earnings date

	N	Mean	T-stat	p- value		Median	Sign M		Standard deviation
Positives	207	0.1056	3.79	0.0002	***	0.0437	<.0001	***	0.4006
Negatives	213	-0.0710	-12.06	<.0001	***	-0.0434	<.0001	***	0.0859

Panel B: 3 day CAR date based on Yahoo date

	N	Mean	T-stat	p- value		Median	Sign M		Standard deviation
Positives	206	-0.0056	-1.48	0.1397		-0.0043	0.0813	*	0.0541
Negatives	212	-0.0091	-2.15	0.0326	***	-0.0102	<.0001	***	0.0618

* significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level

These result are consistent with Hong, Lim, and Stein (2000) who posit that negative news is more slowly assimilated into the markets than positive news. These findings indicate that positive news is immediately incorporated but negative news is not and a second reaction occurs. As this sample data set consists of stocks that have a higher mean size than the mean size of all stocks in the market, it suggests that I am not just capturing a small firm effect.

This suggests that investors do have an opportunity to generate a positive return based on public information. If the initial stock price reaction is negative, the investor should short the stock and when the second announcement of the information is made the investor can take advantage of the second stock price reaction.

This study used five hundred random stocks. Further research could be done using a bigger data set. This would allow for more refined analysis such as sorting by size, by industry, and by market status. It would also be interesting to examine if this good news versus bad news phenomenon happens in other markets besides the U.S. and if it is exacerbated by market conditions. Is bad news more worrying in a bear market?

References

Barberis, Nicolas, Andrei Shleifer, and Robert Vishny. 1998. "A Model of Investor Sentiment." *Journal of Financial Economics*, Vol. 49, pp. 307 -343.

Brown, Stewart L. 1978. "Earnings Changes, Stock Prices, and Market Efficiency." *The Journal of Finance*, Vol. 33, No. 1, pp. 17-28.

Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance,* Vol. 25, pp. 383-417.

Hong, Harrison, Terence Lim, and Jeremy Stein. 2000. "Bad News travels slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies." *The Journal of Finance*, Vol. LV, No. 1 pp. 265 – 295.

Jensen, Michael C. 1978. "Some Anomalous Evidence Regarding Market Efficiency." *Journal of Financial Economics*, Vol. 6, Nos. 2/3 pp. 95- 101.

Jones, Charles p. and Robert H. Litzenberger. 1978. "Quarterly Earnings Reports and Intermediate Stock Price Trends." *The Journal of Finance*, Vol. 25, No. 1, pp. 143-148

Optimizing Strategies for Monopoly: The Mega Edition Using Genetic Algorithms and Simulations

Shree Raj Shrestha, Daniel S. Myers, and Richard A. Lewin, Rollins College[1]

Abstract

Monopoly Mega Edition adds two new mechanics: a Speed Die and Bus Tickets, which give players additional choices in the dice roll and token movement. These new features and their associated choices limit the usefulness of Markovian models as a tool for analyzing the Mega Edition. This paper discusses the method and challenges in modeling this new version of the Monopoly game using evolutionary algorithms and computer simulations, and analyzes the strategic implications of the data obtained from these simulations. In particular, we discuss the difference between aggressive versus strategic gameplay, and the expected cost of using a bus ticket.

Introduction

Monopoly is a popular board game that involves elements of strategy, skill and luck. Numerous guides have been published that instruct the reader with an optimal strategy of how to play the original Monopoly game. Authors like Koury (2012) and Orbanes (2007) have published guidebooks on winning strategies for Monopoly. This game has also been used to teach the strategies behind Real Estate. Janik (2009) published her book on profitable investing that is based on the original Monopoly game. Similarly, Orbanes (2013) uses the Monopoly game in his guide to make smarter financial decisions. On the other hand, Monopoly has also been used to teach Markovian chains in mathematics. Johnson (2003) published a detailed paper that describes how to use Monopoly, and other similar board games, to teach stochastic models.

There have been many analyses of the original game of Monopoly that include simplifying assumptions to translate the game into a mathematical model that can be used to verify the legitimacy of existing strategies. The early work of Ash and Bishop (1972) provides a rigorous analysis of the mathematics behind the game, using limit frequencies of convergence for a simplified model of the game using eigenvalues. Stewart (1996a) published a paper on the fairness of Monopoly in the Scientific American journal where he concluded that the game was fair since the steady state probability of the game approaches 1/40. However, in a subsequent publication, Stewart (1996b) concluded that some squares are more likely to be visited than others if rules like Go to Jail, doubles, Community Chest and Chance cards are included. Abbott and Richey (1997) published a similar analysis of the Monopoly game where they suggested "the accuracy and usefulness of a particular model depends largely on how well the realities of the system survive the translation into a mathematical language". In their analysis they pose questions like how to accommodate within the model such non-Markovian aspects like the Chance cards that direct players to the nearest railroad where they consequently pay double rent. Murrell (1999) conducts a similar analysis for 100 dice rolls in a simplified version of the game, and explains how the landing frequency provides but an initial analysis of the game, and it is important then to consider the cost and revenue generated from each property as well.

However, analysis of the Mega Edition of Monopoly has not been carried out in the literature. This new and alternative version of the game includes randomness and other elements of strategy in the dice roll itself. Introduction of new rules like the Bus Ticket give players more control of movement within a side of the board, making strategy a crucial aspect of the game. If a player rolls a Bus Ticket on the Speed Die, the player gets to skip squares on the current side of the board or keep a Bus Ticket for later use. The new version is bigger, faster, and provides more capital to invest as well as freedom of movement to the player, making for much richer strategic gameplay. The Mega Edition introduces an additional die and 12 new squares with new rules and properties. Analyses of strategies for playing the original version of the game have been carried out using Markovian models and computer simulations; however, the Mega Edition of the Monopoly game has not previously been analyzed for its strategic implications. Nevertheless, guides on how to play the Mega Edition have been published including the U.S. National Monopoly Champion, Matt McNally's "Winning Tips to Own It All." What has not been undertaken in the literature is a rigorous update to the probabilities associated with the revised Mega Edition, which is the focus of this research. We began by first obtaining the steady-state probability of visiting each square for Monopoly: The Mega Edition using the static Markovian model, similar to the approximation used by previous analyses including the seminal work of Stewart (1996b).

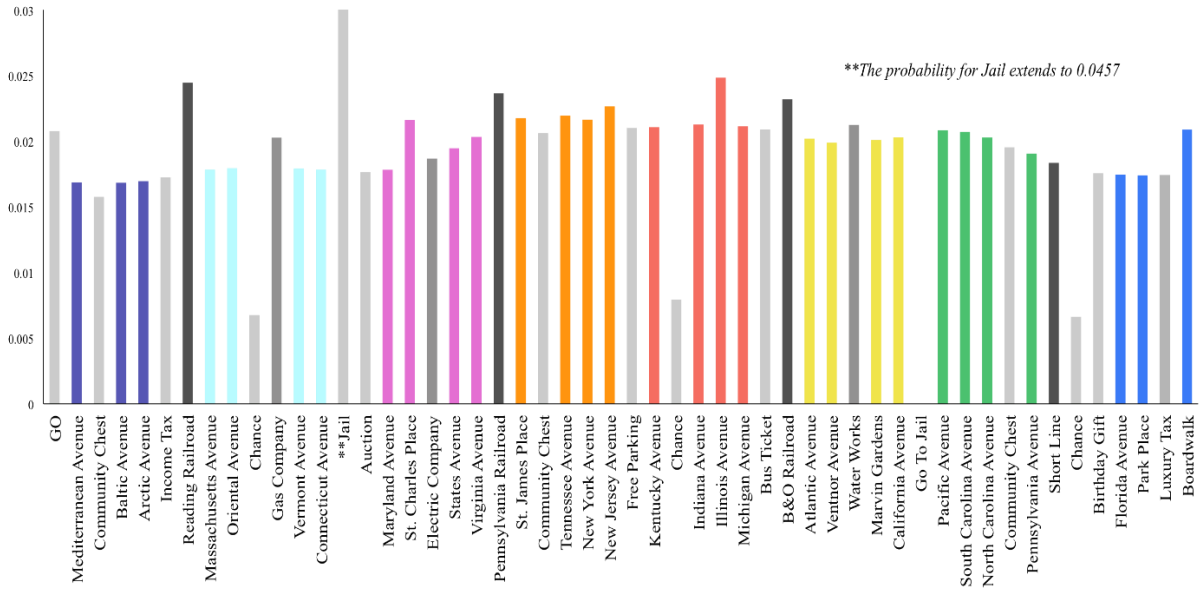
Preliminary Research

The Markovian approximation estimates the expected long-run probability of visiting each square taking into consideration the movement of players due to random dice rolls, Jail, and cards that redirect the player to another location (Shrestha, Lewin and Seitzer 2015).

Using a computer simulation of 250 dice rolls (the average used by previous analyses), we calculated the probability distribution for the 52 squares on the board under the Markovian model. Following Frayn (2005), we model Community Chest and Chance Card by assuming they are drawn at random from a complete stack, as opposed to modeling the composition of the Chance and Community Chest decks at each moment in the simulation.

Properties that are more likely to be landed on are regarded as being more valuable since these properties act as a source of income in the form of rent to the owner of the property. A summary of the results obtained is given below in Figure 1.

Figure 1. Long term probability distribution of landing on each square for each 52 squares of the Monopoly: The Mega Edition game. The color on each bar corresponds to the color group of the property. In addition, black refers to Railroads, dark-gray to Utilities and light-gray to non-property squares. (See Appendix 5 for an updated result with rules for Mr. Monopoly, Bus Ticket, doubles and triples included).



According to the relative probability distribution of the color groups, it can be concluded that the orange, red, yellow and green groups have the highest chance of being landed on, which implies a higher return on investment for these properties in the long run. This preliminary analysis agrees with the strategies suggested by Matt McNally, the U.S. National Monopoly Champion, for the Monopoly Mega Edition (McNally, 2007). Please refer to Appendix IV for the complete Java code and full analysis of results.

Although other complex rules like the triples, doubles, Mr. Monopoly, Bus, and Bus Ticket were not accounted for, the analysis of the game purely by chance enlightens our postulation of simple strategies that one can implement to make one's opponent go bankrupt faster in the long run. However, the long term probability of visiting each property does not in itself lead us to any insight in respect of the short term strategies a player can implement to take advantage of the additional rules in the Mega Edition like the Bus Ticket, which probabilistically occurs once every 3 turns, as one face of the Speed Die is a Bus Ticket.

An Evolutionary Approach

Evolutionary algorithms have been effectively used to optimize complex systems with a large number of variables. These algorithms draw inspiration from the biological process of evolution, whereby an initial population evolves into a new generation, with bias to members of that population that perform well in the environment. Each new generation has, on average,

higher fitness than its parents, and therefore represents a better set of candidate solutions to the problem under consideration. Frayn (2005) used this computation to analyze the valuation of properties for the original game of monopoly. In the scope of our experiment, the "environment" resembles the game of Monopoly and is implemented using a computer simulation.

This method can take into account complex rules easily and effectively, including those that are a challenge to incorporate into more formal Markovian models, such as Go to Jail, doubles, Community Chest and Chance Cards. Each member of the population is an array of numbers that represents the strategic behavior of an individual Mega Edition player. In our simulation, each array consists of ten values between 0 and 1, one for each of the eight color groups, the utilities, and the railroads. Each number represents the probability of purchasing any property belonging to its associated category. For example, a population member with a value of .75 in the entry column corresponding to orange-colored properties will choose to purchase properties in that group 75% of the time and decline to purchase the remaining 25% of the time. Each time a simulated player lands on a purchasable property, the simulation generates a random value between 0 and 1, and then uses the player's strategy vector to make a purchasing decision.

Upgrade decisions are handled in a similar fashion. Each player's strategy vector contains ten additional values between 0 and 1 representing the value that player assigns to upgrading properties in each of the ten categories. Therefore, the player's complete strategy vector is described by twenty values in (0,1). On a player's turn, that player may spend cash to upgrade the properties it owns, with the priority for competing upgrades resolved based on the value the player assigns to properties in each category. Players never voluntarily drop their cash reserves below a set minimum threshold to avoid bankruptcy for example.

Each member of the population competes in simulated games of Monopoly against three other random individuals. The player receives points based on its finishing position in each game. Following Frayn (2005), we award four points for a first place finish, two for second place, one for third place, and zero points for the fourth place. The total fitness of an individual is the sum of the points earned in 100 of simulated games against random opponents. This process is repeated for all the members of the initial population.

Once the fitness scores for all members of the initial population have been tabulated, we apply natural selection, crossover and mutation to form the next generation. One percent of each generation with the highest fitness scores in a generation survive to the next by right as top performers. The remaining 99 percent of the next generation is formed by crossover between the members of the current generation such that the process is biased for individuals that have a higher fitness score. Then, with a small probability, we change independently and mutate each element of the strategy vector for all the members of the new generation. This introduces some randomness into the process to prevent solutions from becoming trapped within local optima.

The process is then repeated indefinitely, or until a termination criterion is achieved. The following algorithm summarizes the methodology used in the genetic process:

1. Generate a random initial population of 1,000 players.
2. Evaluate fitness for each individual in the population.

a. Fitness is the total score of the individual for 100 games played against randomly generated opponents.

3. Apply selection and crossover on best-fit individuals to generate the next generation.

a. The fittest 1 percent continue to the next generation by default.

b. 99 percent of the new generation is made by crossover between randomly chosen individuals, with higher-fitness individuals having a greater selection probability.

4. Apply mutation on the newly generated population to escape local optima.

a. Each parameter of each member of the population is mutated with the probability of 1 percent

5. If termination criterion is not met, repeat albeit from step 2.

The constants for generating a new population – the selection percentage and the mutation rate are arbitrary values; however, certain values converge faster than others. Upon adjusting the variables numerous times, the 1 percent default allows convergence to be achieved in a reasonable amount of time.

Initial Findings

Our initial simulation model included support for several key rules, including extra turns for rolling doubles, Go to Jail, being In Jail, Community Chest and Chance Cards, but did not include support for the more complex Bus Tickets, which allow movement to any square ahead of the player on the same side of the board. Temporarily ignoring Bus Tickets allowed us to test if our results agreed with the Markovian analysis carried out in previous studies. The next section removes this restriction and considers the strategic trade-offs of when to use available Bus Tickets.

Triples pose an additional complexity. Upon rolling triples, the player may move to any square on the board. To account for this rule, we assumed the player will move to the unowned square that it perceives as being the most valuable; if all the properties are owned, the player deliberately moves to the Go to Jail square where the player then goes to Jail and waits to roll

doubles for 3 consecutive turns. This strategy, well known among competitive players, allows the player to remain in Jail to decrease the chances of paying high rents in the late phase of the game when all properties are owned.

Upon running the algorithm for 70 generations, it was observed that the probabilities for both the perceived value of purchase converged towards 1, implying that *all* properties are valuable to buy and upgrade, independent of their steady-state landing probabilities. This result contradicts the results from the Markovian analysis that has been presented in the preliminary research section, which estimated that certain property groups were more valuable than others based on empirical differences in the steady-state landing probabilities.

An interesting observation from this result can be made that has an important strategic implication on the new upgrade rules. The Mega Edition has four properties for each color group, but a player only needs to own three to begin upgrading properties within a group. Therefore, players must purchase aggressively to prevent their opponents from collecting three properties within a group, and if a player has purchased two properties in a group, it makes strategic sense to acquire at least one more to enable upgrades. Further, in a game against three opponents, a player cannot realistically expect to return to a property again before it is visited and possibly acquired by another player. Therefore, players generally face a one-time yes/no decision to purchase each property they land on, with the practical consequence that a no decision entails abandoning that property to an opponent for the duration of the game. Thus, players are incentivized to play aggressively and always purchase properties they land on, given sufficient financial resources.

Including Bus Tickets

The Bus Ticket allows the player to jump ahead to any forward square on the current side of the board. To get a Bus Ticket, the player has to roll a Bus on the Speed Die (or land on the Bus Ticket square) and choose to retain the ticket for later use. The introduction of this new rule not only speeds up the gameplay but also introduces a new strategic element to the game. There are risks involved in adopting this approach given that 3 of the Bus Tickets cause any others drawn to expire.

The analysis in this section considers the strategic implications of using a Bus Ticket in the late phase of the game, when all properties have been purchased. Further, we consider the *worst-case* situation, where opponents own all reachable properties. In this case, a player's safest move is to always advance to a corner square if possible, because none of the corner squares require the payment of rent. Further, advancing to GO awards $200 and advancing to the Go To Jail space places the player in Jail, which is the safest place to spend turns in the advanced stage of the game without the danger of paying rent on opponent-owned properties (Frayn, 2005).

Suppose that a player is currently on square *S* of the Mega Edition board and has one Bus Ticket. There are three strategies the player might pursue over the next two turns:

1. Use the Bus Ticket to jump to the end of the current side and then move using a normal die roll on the next turn
2. Move using a normal die roll on this turn, then use a Bus Ticket on the next turn.
3. Move using two normal die rolls.

To evaluate the relative trade-offs of these three approaches we consider the worst-case expected rent a player might have to pay under each strategy. Figure 2 presents simulated results of these values for each starting square *S*. For a complete code that implements the algorithm in Python, please refer to Appendix III.

Figure 2 shows that using a Bus Ticket decreases the expected rent paid for all of the squares. This is expected, as the player always uses the Bus Ticket to transition to a square that does not require any rental payment. Further analysis of the difference between the rents shows us that using a Bus Ticket on the third side of the board, i.e. from Free Parking to California Avenue, to move to the Go To Jail square significantly decreases the expected rent paid in the next die roll to almost 0. Similarly, there is a comparative advantage in using the Bus Ticket to skip the fourth side of the board, from Pacific Avenue to Short Line, where the expected rent paid on a die roll is the highest of all the properties on the board.

Incorporating Auctions

An auction occurs in the case where a player lands on a property (or lands on the Auction square when there are still unowned properties available), but chooses not to purchase it; by the official rules, the property should be immediately auctioned off. Since the auction starts at a bid of $1 (made by the declining player), to the maximum bid offered, players must subsequently decide how much to pay. This leads to two strategies based on the face value of the property. First, the player bids less than the face value of the property, and the second when the player elects to overpay by some amount.

The simulated routine of the gameplay used in our previous analysis was thus modified so that a property is auctioned off when a player lands on it but declines to purchase it. For this purpose, a maximum bid factor parameter was added to each player. A maximum bid factor of 1 indicates that the player will an amount up to and equal to the face value of the property.

Similarly, a bid factor of 0.5 would indicate half and a bid factor of 2 double of the face value respectively. A factor of 0.0 indicates that a player enters a bid of 0 in all auctions.

The simulation was run for 500,000 rounds of the game. For each round, the maximum bid factor of one player was controlled while the other three opponents choose bid factors in the range [0.0, 2.0]. According to the results from the evolutionary analysis, players who play aggressively, i.e. who choose to purchase every property they land on given sufficient funds, performed better than players who do not. In the analysis for the Auction rule, it is assumed that all players play aggressively. The expected fitness score over 100 games obtained by each bid factor is presented in the table below.

Table 1: Maximum Bid Factor versus Expected Fitness Score over 100 Cumulative Games with Random Players

Maximum Bid Factor	0.0	0.25	0.50	0.75	1.0	1.25	1.50	1.75	2.0
Expected Fitness Score	165.77	172.30	173.94	175.17	176.28	176.55	176.70	176.38	175.95

A fitness score of 400 would indicate perfect performance, i.e. no losses in any match with a maximum score of 4 per match. The results show that not participating in auctions has a clear detrimental effect on fitness. Since the values greater than 1.0 are relatively close, it can be seen that bidding more than 1.5 times the face value does not improve the fitness score significantly. Furthermore, the results show that fitness is improved at a bid factor of 1.5, indicating a willingness to pay up to 50% above a property's face value during an auction which corroborates with observations at tournaments.

This indicates participating in auctions and being willing to bid above the face value of a property is a very good strategy. This makes intuitive sense because an auction is effectively an opportunity to acquire a property outside of a normal turn. This increases the ability of the player to acquire property since the chance that the player will get an opportunity to purchase that same property later in the game is extremely low.

However, bidding too much early on in the game can have adverse effect on subsequent purchases, rent payment and auctions putting the player in risk of bankruptcy. The results are consistent with this observation as there is no significant advantage in bidding more than 1.5 times the face value of the property which is demonstrated by the limiting effect observed in the table above. Moreover, this begs the question of what auction should players use if all opponents bid as high as possible? The table below represents the expected fitness score over 100 cumulative games.

Table 2: Maximum Bid Factor versus Expected Fitness Score over 100 Cumulative Games.

Maximum Bid Factor	0.0	0.25	0.50	0.75	1.0	1.25	1.50	1.75	2.0
Expected Fitness Score	168.68	176.78	178.98	180.15	180.58	181.22	181.14	181.53	181.43

The data in the table above confirm the results obtained in previous table - fitness is improved by participating in auctions and being willing to bid 25-50% above the face value of a property.

Dynamics of Property Trading

In the real game, players are allowed to trade properties with other players outside their normal roll. Players can trade properties in all color groups, including the utilities and the railroads, which can allow the player to upgrade properties in cases where the player does not own enough properties in the color group to allow for upgrades already.

Under this analysis, we use the following strategy for all the players. The player commencing the trade offer is referred to as the trader and the player accepting the trade offer is referred to as the client hereinafter. The trader makes a deal with the client if and only if:

1. Both the trader and the client have the same number of properties in a color group, the trader is willing to both offer and receive a property from the color group.
2. The trader owns more properties in a color group than the client, the trader is willing to receive a property from the color group.
3. The trader owns less properties in a color group than the client, the trader is willing to offer a property from the color group.

The trader does not offer upgraded properties or properties that he or she received from previous trades but is willing to offer color groups that he or she has offered to other clients. Furthermore, the trader makes offers to random players until an offer is accepted, or until no players are left to make an offer to. Furthermore, among the properties agreed upon between the trader and the client, the trader prefers to receive properties that he or she perceives has a higher purchase value.

In our previous analysis it was established that a player should purchase all properties as long as the player has the money to buy and maintain a minimum cash reserve to avoid bankruptcy. However, for this analysis, the perceived values of purchase

and upgrade are allowed for in the range of [0.9, 1.0]. A perceived purchase or upgrade value of 0.95 for a property X means that the player will choose to purchase or upgrade the property X with a 95 percent probability respectively.

All properties need to be owned in a color group and upgraded to hotels in order to build a skyscraper. Similarly, Railroads can only be upgraded once by a single Train Depot. Furthermore, as in the real game, players can only upgrade properties evenly within the color group. For example, in a given color group, a player cannot build four houses on one property and have none in the other. In other words, players aim to upgrade all properties in a color group to the same level.

In this new analysis, some rules ignored in the evolutionary analysis of our basic strategies were added, specifically, Mr. Monopoly and Triples. In the case of rolling a Mr. Monopoly, a player moves to the next unowned property, or to the next owned property on which they must pay rent if all properties have already been bought. The addition of the Mr. Monopoly rule makes it really advantageous for players who have higher upgrades later in the game when all properties have been bought. Players cannot only collect high rent out of the normal dice roll and bankrupt players effectively, but also may expect opponents to throw Mr. Monopoly one third of the time, as the third die has two faces for Mr. Monopoly. In the case of triples, the player moves to any unowned property or the Go To Jail square if all properties have been owned. Technically the triples allows for the player to move to any square on the board, but in later gameplay the advantages of remaining in Jail are overwhelming. Furthermore, a player always opts to stay in Jail if he gets in Jail because of rolling triples.

The simulation for the game was modified to base the score in order of ranking, as opposed to awarding points in the power of 2 as we did in the evolutionary analysis for our purchasing strategy. The game still involved four players as in our previous analyses. The player who remains in the game is awarded the rank of 1 whereas the player who gets bankrupt first is awarded the rank 4. In cases when players are not bankrupt until 400 turns per player, ranking is based on total assets with the player owning the most assets being awarded rank 1. The expected ranking for a completely fair game is, hence, 2.5 which is the average of the total score over the number of players.

The simulation was allowed to occur without any players being controlled. Instead, games where trades occurred were observed and data on the trade behaviors were extracted from each game. Specifically, the number of times players were engaged in trade, both offering and receiving, was recorded. The number of trades made by the players versus the average rank of the player over 50,000 games is presented in the table below.

Table 3: Number of Trades versus Average Rank over 50,000 games

Number of Trades	0	1	2	3	4	5	6	7	8
Average Rank	2.85	2.48	2.10	1.80	1.61	1.46	1.5	1.36	1.5
Frequency	79,465	69,474	33,792	11,981	3,982	1,056	209	39	2

The above table shows that not engaging in any trade in a game where all players are willing to make mutually beneficial trades, reduces the rank of the player to 2.85 which is less than the expected rank of 2.5. On the other hand, players who successfully carry out at least 1 trade tend to normalize to the expected rank. The frequency of occurrence shows that it gets more difficult to be able to pass up higher numbers of successful deals. However, in cases where a player succeeds to get 3 or more deals, a player can consistently expect to be in the top 2 players.

Table 4: Turning Point Analysis

Total Games	50,000 games
Turning point not reached	14,645 games
Average Turning Point	133rd turn overall, 33rd turn for the player

In this analysis, the number of turns after which all the properties are owned, hereinafter referred to as turning point, was determined. This data provides an insight on when a player can expect to change strategies for the Bus Ticket, Go to jail and stay in Jail, and the Mr. Monopoly strategy. In the data obtained from 50,000 simulated games, the turning point is not reached for 14,645 games, suggesting that for 30 percent of the game, the player can expect not to change strategies. Furthermore, the average turning point of 133rd overall roll, or 33rd turn for the player, indicates when it is advisable for players to change strategies for Mr. Monopoly, Go to Jail and Stay in Jail and Bus Ticket.

Results

Based on the results of Figure 2, we re-ran the genetic algorithm to identify effective strategies for the Mega Edition game including Bus Tickets, Auctions & Property Trading. The updated simulation algorithm assumes that players will now use an available Bus Ticket to skip the third side of the board and the fourth side prior to the Short Line. The termination criterion was

set to terminate at 70 generations, as in the initial test run; this length was sufficient to reach a plateau in the evolutionary algorithm's fitness progress.

Figure 2. Expected rent paid for each square for three scenarios Bus Ticket and Dice Roll (BD), Dice Roll and Bus Ticket (DB), and Dice Roll and Dice Roll (DD) over a million simulated games.

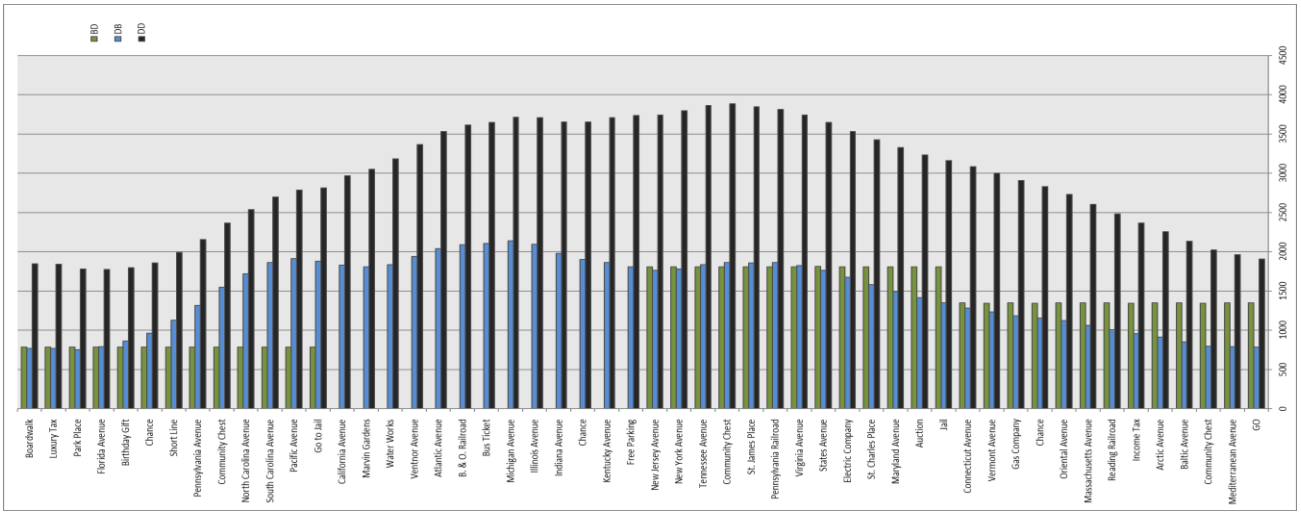
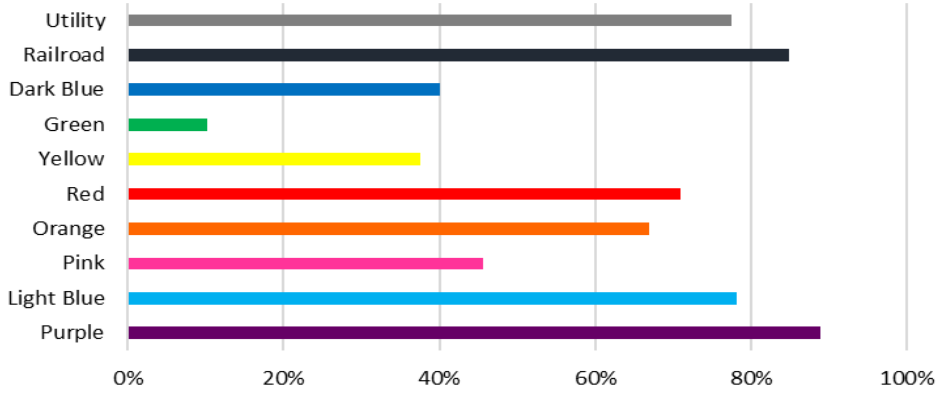


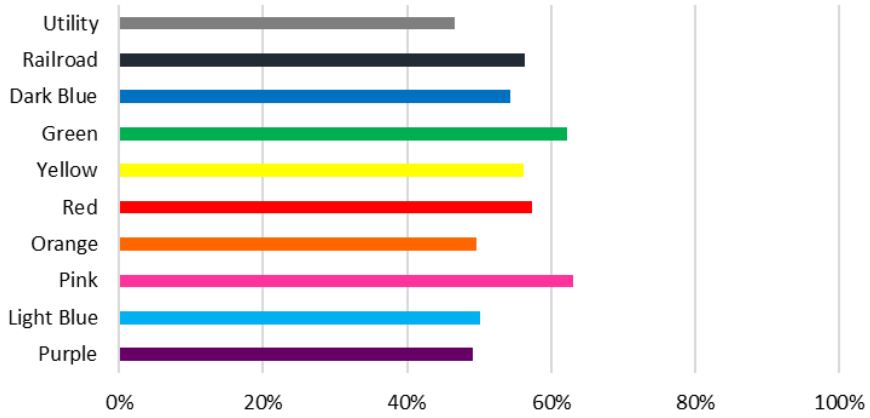
Figure 3 shows results for the average perceived value of purchase and upgrade for each property group. It can be seen that the dark blue, green and yellow property groups are devalued compared to other property groups. This is expected since due to the presence of Go to Jail in the third corner of the board, between the yellow and the green property groups, players are more likely to move to the Jail square instead of traversing the fourth side of the board where the green and dark blue property groups reside. Furthermore, the Bus Ticket can also be used to skip these properties to move to the Go to Jail square and then to Jail, where the player gets to wait until a double is rolled or 3 turns to get out of the Jail.

Figure 3. Average of the perceived value of purchase for each property group across 70 generations of optimal candidates.



Higher probabilities represent a higher likelihood of the player buying the property upon landing on the respective square. Smaller probabilities represent a lesser likelihood of purchase. For a detailed analysis of how the perceived value of purchase evolves over generations, please refer to Appendix I. In addition to the perceived purchase values, the average perceived value of upgrade for each property group is presented below in Figure 4.

Figure 4. Average of the perceived value of an upgrade for each property group across 70 generations of optimal candidates. Higher probabilities represent a higher likelihood of the player upgrading the property if owned. Smaller probabilities represent a lesser likelihood of upgrade. For a detailed analysis of how the perceived value of purchase evolves over generations, please refer to Appendix II.



Even though the dark blue, green and yellow properties were devalued for purchase, which was the expected effect of the Bus Ticket, Figure 4 shows us that these properties are still worth upgrading. Due to higher rental values for these properties, it may be that upon upgrading, these properties generates higher rent than other properties even if they are not visited as frequently as other property groups. Note that, as Bus Tickets are not guaranteed, players may still visit properties on the third and fourth sides of the board, albeit less frequently than in the original version of Monopoly.

Higher perceived values of purchase and an average perceived value of upgrades for the Railroad, Utility and Purple property groups imply that these properties can serve as a consistent source of rent. Since these properties come at relatively lower costs for both purchase and upgrade, these produce a higher return on investment than the other properties.

Conclusions

The Mega Edition of the classic Monopoly game offers more strategic freedom to the player than the original version. Although genetic algorithms are at an early stage of development, this method has been used to optimize complex systems with a large number of variables such as Monopoly. Frayn (2005) successfully analyzed the original version of the Monopoly game based on the genetic approach using simulations. Using a similar genetic approach for analyzing the Mega Edition with multiple additional elements of complexity we have taken into account additional rules like Bus Tickets, Speed Die, Auctions, doubles and triples as well as property trading, not previously addressed in the literature. We were able to verify that certain property groups like dark blue, green and yellow are significantly devalued whereas property groups like railroads, utilities and purple have a larger average perceived value, implying they have a higher return on investment under these conditions. However we anticipate further work will be necessary incorporating rules such as mortgages which provide strategic cash management opportunities, which are yet to be accounted for in our simulated version of the game due to the complex nature of such rules. Further insight into aspects that involve elements of strategy and skill rather than pure luck should lead to a richer understanding of real world property trading and investment decisions.

Notes

1. This research was supported by The Edward and Stella C. Van Houten Memorial Fund through the Student-Faculty Collaborative Scholarship Program at Rollins College.

References

Abbott, Stephen D., and Matt Richey. 1997. 'Take A Walk On The Boardwalk'. *The College Mathematics Journal* 28 (3): 162-171.

Ash, Robert B., and Richard L. Bishop. 1972. 'Monopoly As A Markov Process'. *Mathematics Magazine* 45 (1): 26-29.

Frayn, Colin. 2005. 'An Evolutionary Approach to Strategies for the Game of Monopoly.' *CIG*.

Janik, Carolyn. 2008. *The Monopoly Guide To Real Estate*. New York: Sterling Pub.

Johnson, Roger W. 2003. 'Using Games to Teach Markov Chains'. *PRIMUS* 13 (4): 337-348.

Koury, Ken. 2012. *Monopoly Strategy: How To Win The World's Most Popular Board Game.* CreateSpace Independent Publishing Platform.

Lucas, Simon M., and Graham Kendall. 2006. 'Evolutionary computation and games.' *Computational Intelligence Magazine, IEEE* 1 (1): 10-18.

McNally, Matt, 'MONOPOLY®: The MEGA Edition: Winning Tips to Own It All'. 2007.

Murrell, Paul. 1999. 'The Statistics of Monopoly'. *Chance* 12 (4) 36-40.

Orbanes, Philip. 2013. *Monopoly, Money, and You: How to Profit from the Game's Secrets of Success*. McGraw-Hill.

Orbanes, Philip. 2007. *The Monopoly Companion: The Players' Guide*. Sterling Publishing Company.

Shrestha, Shree R., Richard A. Lewin, and Jennifer Seitzer. 2015. 'Analyzing Real Estate Implications of Monopoly Mega Edition through Simulation'. *AEF Papers & Proceedings*, Vol.39.

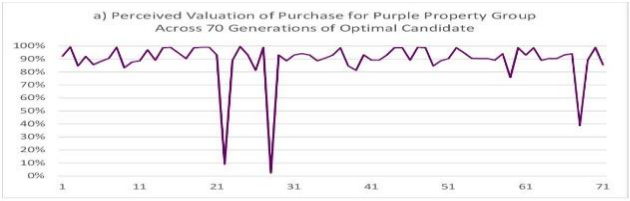
Shrestha, Shree R., Daniel S. Myers and Richard A. Lewin. 2016. 'Towards An Optimal Strategy For Monopoly: The Mega Edition Using Genetic Algorithms & Simulations'. *AEF Papers & Proceedings*, Vol.40.

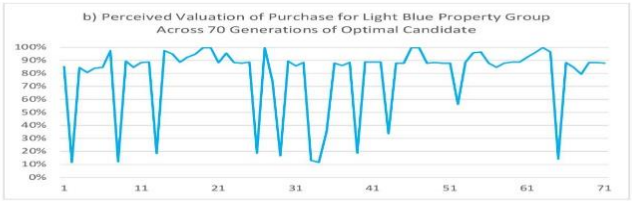
Stewart, Ian. 1996a. 'How Fair is Monopoly? '. *Scientific American.* 274 (4): 104-5.

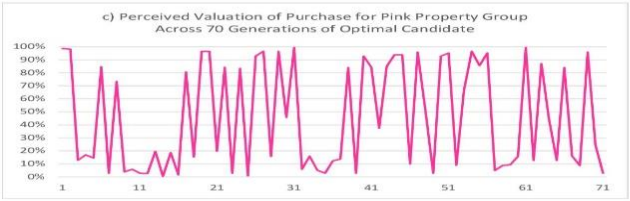
Stewart, Ian. 1996b. 'Monopoly Revisited'. *Scientific American.* 275 (4): 116-119.

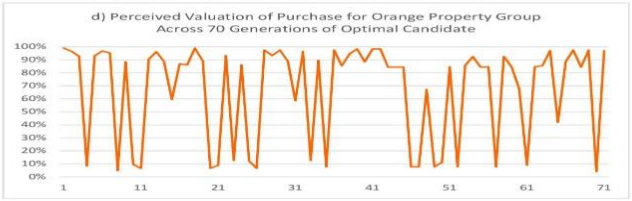
Appendices

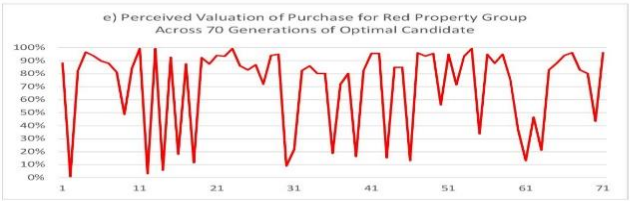
Appendix 1: The figures below show how the perceived value of purchase for each color group changes across 70 generations of optimal candidates. Figure a. through j. are color coded with their respective color group with the exception of Railroads and Utility which are coded as Dark Gray and Light Gray respectively.

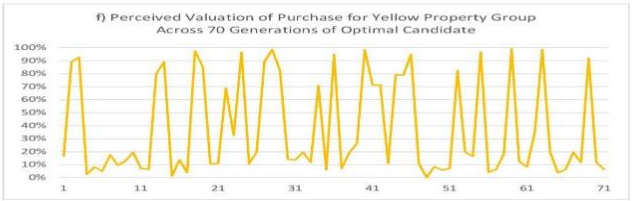


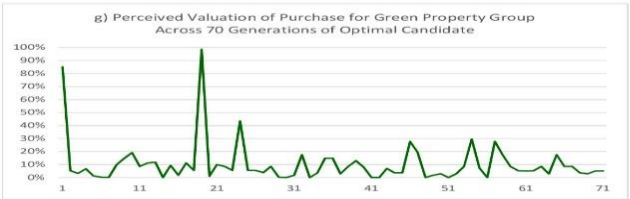


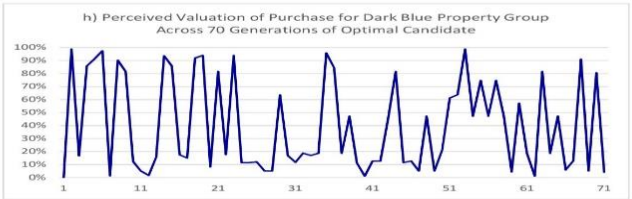


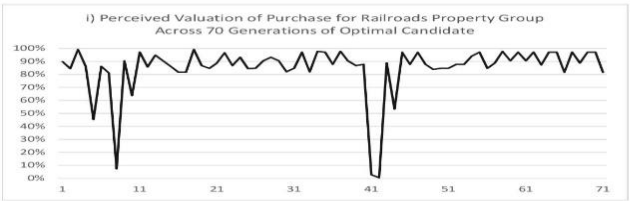


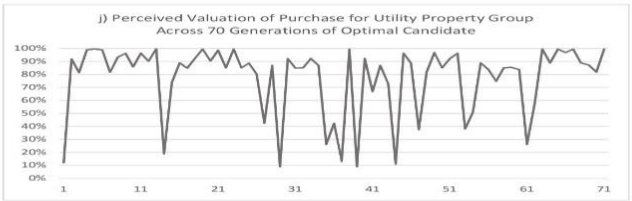




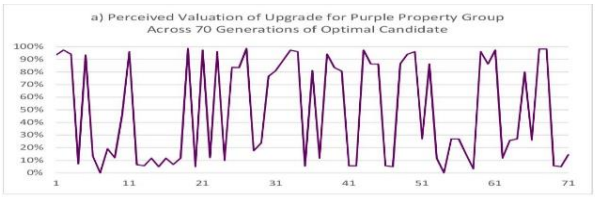


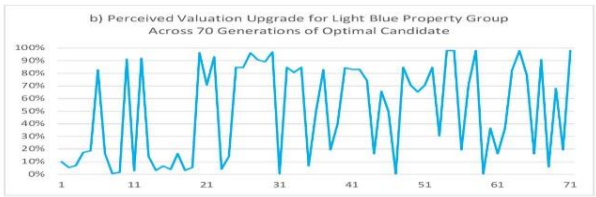


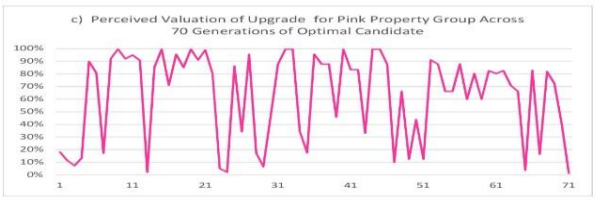


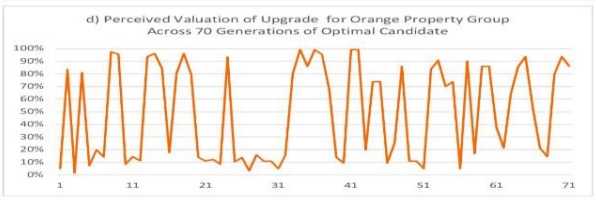


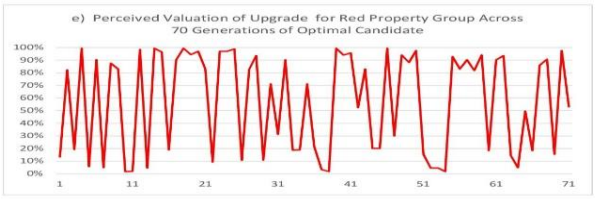
Appendix 2: The figures below show how the perceived value of upgrade for each color group changes across 70 generations of optimal candidates. Figure a through j are color coded with their respective color group with the exception of Railroads and Utility which are coded as Dark Gray and Light Gray respectively.

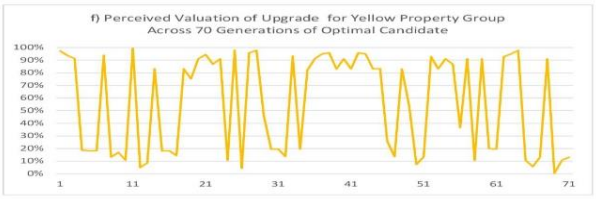


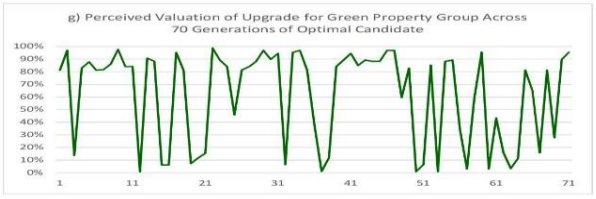


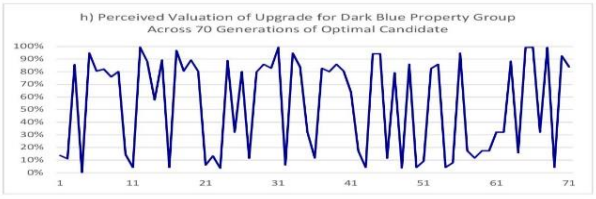


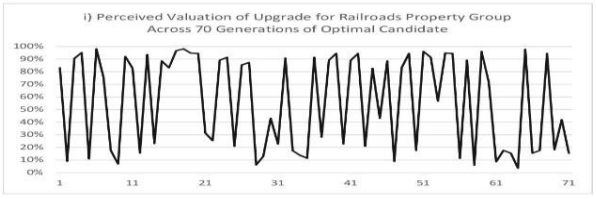


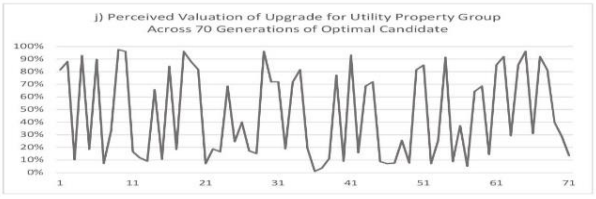












Appendix 3: Please visit the following link to refer to the complete list of python scripts used for the analysis.
https://github.com/shreerajshrestha/Monopoly_Mega_Edition_Evolutionary_Optimization

Appendix 4: The following link includes the Java files used in the simulation for the preliminary research where we analyzed the Mega Edition using Markovian analysis. The statistics that were generated are also included in this repository.
https://github.com/shreerajshrestha/Monopoly_Mega_Edition_Markovian_Optimization

Appendix 5: Visiting Probability Analysis

	3%
literranean Avenue	0%
mmunity Chest	2%
ic Avenue	3%
tic Avenue	2%
me Tax	4%
ding Railroad	3%
ssachusetts Avenue	3%
ental Avenue	3%
nce	4%
Company	2%
mont Avenue	3%
necticut Avenue	3%
	3%
tion	0%
yland Avenue	4%
Charles Place	2%
tric Company	3%
es Avenue	1%
ginia Avenue	3%
nsylvania Railroad	4%
ames Place	3%
mmunity Chest	3%
nessee Avenue	3%
York Avenue	3%
Jersey Avenue	4%
Parking	0%
tucky Avenue	3%
nce	0%
ana Avenue	3%
ois Avenue	0%
higan Avenue	0%
Ticket	3%
O. Railroad	3%
ntic Avenue	0%
tnor Avenue	3%
er Works	2%
vin Gardens	3%
fornia Avenue	3%
o Jail	3%
fic Avenue	0%
th Carolina Avenue	2%
th Carolina Avenue	2%
mmunity Chest	3%
nsylvania Avenue	0%
rt Line	0%
nce	0%
hday Gift	3%
ida Avenue	3%
Place	0%
ury Tax	4%
rdwalk	0%
ail	3%

How Retirees Prioritize Socializing, Leisure And Religiosity - A Research Based Study For Planners.

Aman Sunder, Swarn Chatterjee, Lance Palmer, Joseph W. Goetz, University of Georgia

Abstract

Happiness and well-being are among the two most important goals for individuals approaching retirement and often form the basis of total income requirement during retirement. These goals are almost met when the household expenses are paid for, individual health is taken care of, and there is residual money available for social, leisure, and religious activities. Time spent on socializing, leisure, religiosity, and financial services is the main concern. We use the *American Time Use Study* (ATUS) *2013 Well-Being Module* to examine the predicted probabilities of retirees' time spent in each activity of interest based on various demographic and socioeconomic factors.

Introduction

Financial planning has come a long way and has deep roots embedded in the theories of household economics, finance, law, psychology, and communication studies. It has expanded its horizon from consulting and research to counseling and therapy. Individuals seeking the help of professional financial planners include individual investors, financially constrained small business owners, retirees, and transitioning young adults. However, all of these individuals who seek expert financial advice, knowingly or unknowingly, have similar goals. These goals include financial well-being and the judicious use of available resources, including money and time.

The use of time and money can be seen as dependent on each other. For example, increased time spent on activities such as shopping can lead to increased purchases. On the other hand, time spent on household activities such as lawn mowing can save money that is otherwise spent on these services. However, paying for these services can lead to a saving in personal time and add to one's convenience. We attempt to use the time use diary included in the *Well-Being Module* of the *American Time Use Survey 2013* (ATUS) to identify the predictors of the choice to spend time on activities such as socializing, leisure, and religiosity.

The results of our study indicate that time spent with family is positively associated with the time individuals spent on activities related to socializing and financial services. The time spent on financial services is highly associated with individual stress and well-being and negatively related to time spent with spouse. Income is positively associated with religiosity. The findings of this study provide new information for individual investors, financial planners, and scholars of aging and retirement-related issues. Policy implications of the key findings of this study are also discussed.

Methodology

Data

This study uses the *American Time Use Survey Well-Being Module 2013*. The ATUS data constitute a nationally representative self-reported time diary of the American population. The data include comprehensive information on the use of time by the people surveyed. Individuals 15 years of age or older are selected. Monthly telephone interviews are conducted continuously throughout the year after random selection. For the *Well-Being Module* (WB), three activities of five minutes or more are randomly selected, and respondents are asked to rate their general feeling about these activities on a defined scale. Incomplete responses are not included in the WB module. ATUS data include demographic variables such as race, age, education, and family income. ATUS is sponsored by the *United States Bureau of Labor Statistics*, *US Department of Labor,* and conducted by the *US Census Bureau*.

The time use variables of interest employed in this study include socializing, expensive leisure, religiosity, and financial services. These constructed variables are publicly available in the ATUS data. Detailed definitions and the construction methodology used for these variables are available in the ATUS code book (Bureau of Labor Statistics, 2016).

Analysis

The first analysis attempts to plot the predicted time spent on activities related to socializing, leisure, religious, and personal financial management. These activities have different implications for the nest egg needed for retirement and influence the happiness, stress, and well-being of individuals during old age. Next, time use ratios are defined as a unit of available time to plot the predicted probabilities of high time use. In the end, an attempt is made to study the implications of these activities for happiness, stress, and well-being of the elderly.

Variables

Table 1: Summary of the key variables with *Time Available* >0

Variable	***N***	***Mean***	***Std. Dev***	***Min***	***Max***
Socializing	11385	75.76	113.63	0	944
Socializing Ratio	11357	0.2	1.33	0	118
High Social	11357	0.16	0.36	0	1
Financial Services	11385	0.54	4.79	0	120
Financial Services Ratio	11357	0	0.02	0	1
High Finance	11357	0.01	0.09	0	1
Religiosity	11357	13.57	47.79	0	680
Religiosity Ratio	11357	0.03	0.1	0	0.98
High Religious	11357	0.03	0.17	0	1
Leisure (Expensive)	11385	24.46	49.67	0	1055
Leisure Expensive Ratio	11357	0.06	0.11	0	0.95
High Leisure	11357	0.1	0.3	0	1
Leisure (Inexpensive)	11357	236.05	204.56	0	1439
Happy	11357	4.38	1.62	0	6
Wellbeing	10353	7.11	2.02	0	10
Stress	11357	1.29	1.74	0	6
Time Available	11357	525.3	258.3	1	1440

1. *Socializing, Socializing Ratio, and High Socializing Ratio (0, 1): Socializing* includes time spent on events, meeting people, attending events of household and non-household children, time spent with these children, religious services, and time spent planning these activities. The *Socializing Ratio* is calculated by dividing *socializing time* by *time available*. *High Socializing Ratio* is defined as the 4th quartile of the *Socializing Ratio*. High is coded as 1 and others as 0.
2. *Financial Services, Financial Services Ratio, and High Financial Services Ratio (0, 1): Financial Services* includes time spent on banking and other financial services and the time spent waiting for and traveling for these activities. *Financial Services Ratio* is calculated by dividing *financial services time* by *time available*. *High Financial Services Ratio (0,1)* is defined as the 4th quartile of *Financial Services Ratio*.
3. *Religiosity, Religiosity Ratio, and High Religiosity Ratio (0,1): Religiosity* includes time spent participating in attending religious activities, waiting and traveling times related to these activities, and watching religious programs on television. *Religiosity Ratio* is calculated by dividing *religiosity time* by *time available*. *High Religiosity Ratio* is defined as the 4th quartile of religiosity ratio. High is coded as 1 and others as 0.
4. *Leisure (Expensive), Expensive Leisure Ratio, and High Expensive Leisure Ratio (0,1): Expensive Leisure* includes time spent shopping for non-grocery and grocery items, gas, comparison shopping, and traveling and waiting related to these activities. *Expensive Leisure Ratio* is calculated by dividing *expensive leisure time* by *time available*. *High Expensive Leisure Ratio (0,1)* is defined as the 4th quartile of *Expensive Leisure Ratio*.

5. *Leisure (Inexpensive): Inexpensive Leisure* includes time spent resting, watching TV, listening to music, playing games on a computer, arts and crafts, reading or writing for personal interest, attending performing arts, watching movies in a theater, and similar activities.
6. *Happy* is recoded from WUHAPPY of the *Well-Being Module*; 0 means than an individual reported that he/she is not happy at all and 6 means he/she is happy.
7. *Well-being* is self-reported by the individuals based on how they perceive their situation, from a low of 0 to a high of 10.
8. *Stress* is self-reported by the individuals based on how they perceive their stress, from a low of 0 to a high of 10.
9. *Time available* is calculated by subtracting time spent on sleeping, sleeplessness, relaxing, household activities, and inexpensive leisure from the total of 1,440 minutes per day. An individual engages in several activities over which he/she has free choice and others that do not involve free choice. The attempt is to seclude the time over which an individual has free choice. For the purpose of this study, the data were restricted to *Time Available* >0.

Middle and High Income Retirees

Financial planners are usually interested in clients who can provide them with assets under management (AUM) of $1 million or more. Finke, Huston, and Winchester (2011) suggest that individuals with half a million dollars in assets are more likely to seek financial advice. Other studies have found that middle-income individuals with greater stock of human capital are also more likely to seek financial advice (Harness, Chatterjee, & Salter, 2014). We use the approach suggested by Kalenkoski and Oumtrakool (2014) of connecting net worth with income. The authors suggest that retirees with assets of half a million dollars or more have income of $62,332. Therefore, the minimum family income for this study is $60,000. We define retirees as those who are 55 years of age or older and report that they are retired. In essence, our study is restricted to middle- and high-income retired individuals, who are the primary respondents in the survey. .

Results

Socializing, expensive leisure, religiosity and financial services for the middle and high income retirees.

Table 2: Regression results

Variable	Socializing		Exp. Leisure		Religiosity		Financial Serv.	
	Coef.	**P>|z|**	**Coef.**	**P>|z|**	**Coef.**	**P>|z|**	**Coef.**	**P>|z|**
Age	0.370	0.447	0.120	0.476	- 0.030	0.817	0.003	0.396
Sex	0.480	0.981	1.490	0.851	1.920	0.768	0.070	0.613
Whites	- 39.290	0.626	14.290	0.253	- 7.400	0.570	0.040	0.890
Hispanics	- 98.120	0.228	11.870	0.495	- 4.100	0.783	-0.070	0.852
Blacks	- 10.960	0.908	17.140	0.283	-13.540	0.315	0.625	0.176
Education Level	3.960	0.417	- 0.660	0.703	0.320	0.808	0.050	0.202
Family Income Level	- 6.070	0.534	- 1.300	0.687	**6.410***	0.100	-0.040	0.504
Time spent with spouse	- 0.080	0.368	- 0.020	0.485	0.030	0.212	**-0.002***	0.063
Time spent with family	**0.120***	0.054	0.020	0.354	0.010	0.465	**0.002***	0.070

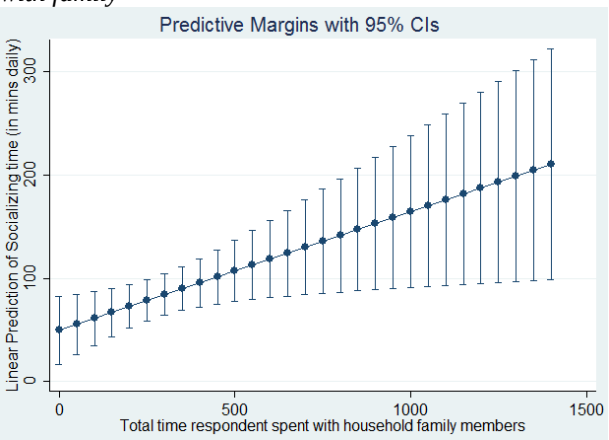
Significant at * 10% level, **5% level and ***1% level

Table 2 shows the ordinary least squares (OLS) regression with the independent variables being demographic variables such as age, gender, race, level of education, and household income. We also include a few social attributes such as presence of spouse or partner and time spent with spouse or partner and family. The time spent with family and spouse is included as a control because marital status can influence an individual's choice of time used for other activities (Passias, Sayer, & Pepin, 2015). We restrict the regression to middle- and high-income retirees. The results from Table 2 indicate that family income is positively associated with religiosity, whereas time spent with spouse or partner is negatively associated with time spent on

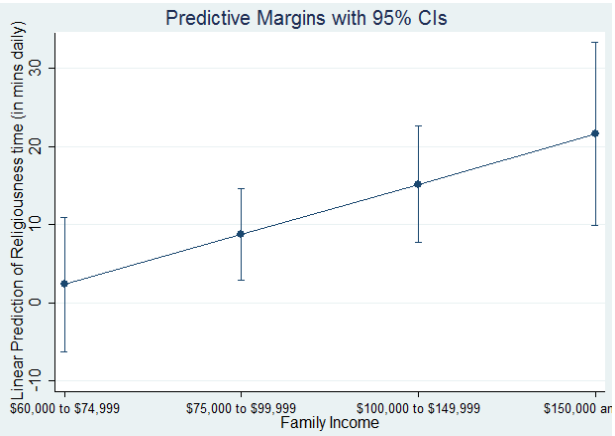
financial services-related activities. However, time spent with family is positively associated with time spent on financial services activities.

Figure 1: Margins plots of linear predictions:

Linear predictions for socializing times *by time spent with family*

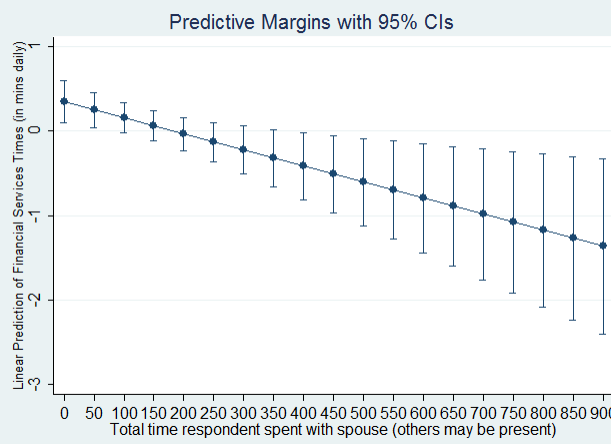


Linear Predictions for Religiosity times by family income.

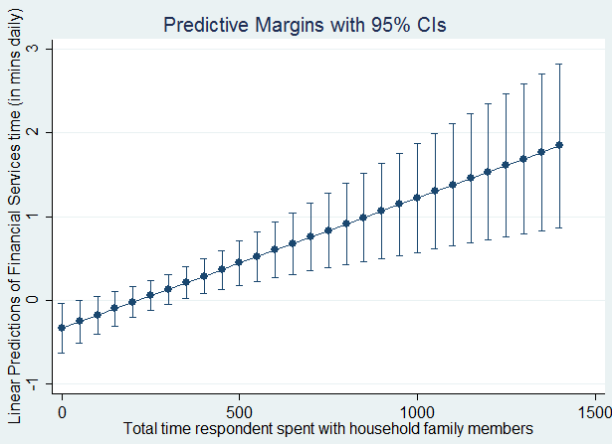


Linear predictions for the time spent on financial services

By time spent with spouse.



By time spent with family



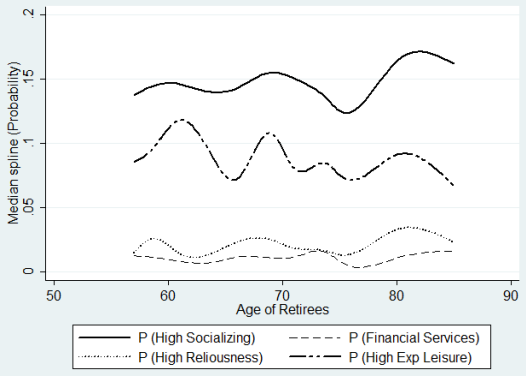
Probability plots of high socializing, high religiosity, high expensive leisure and high financial services time-use.

The plots of predicted probabilities of the dependent variables on some of the independent variables of interest are shown in Figure 2. These plots are generated from the logistic regression on the binary variables created, such as *High Socializing Ratio* (0 and 1), *High Financial Services Ratio* (0 and 1), *High Religiosity Ratio* (0 and 1), and *High Expensive Leisure Ratio* (0 and 1), using the same independent demographic and social variables.

Figure 2. Predicted probabilities of High Socializing Ratio, High Financial Services Ratio, High Religiosity Ratio, and High Expensive Leisure Ratio.

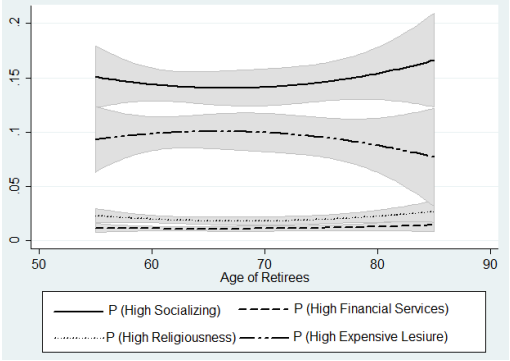
Median splines

Predicted probabilities based on age of retirees

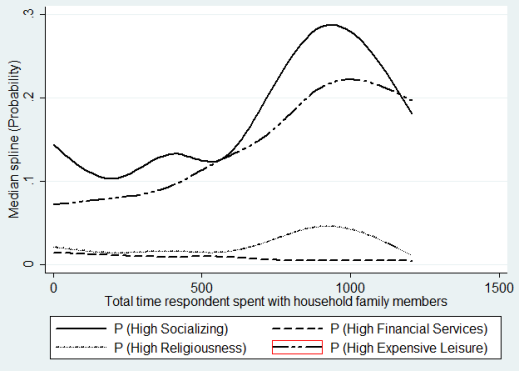


Fitted values with 95% CI

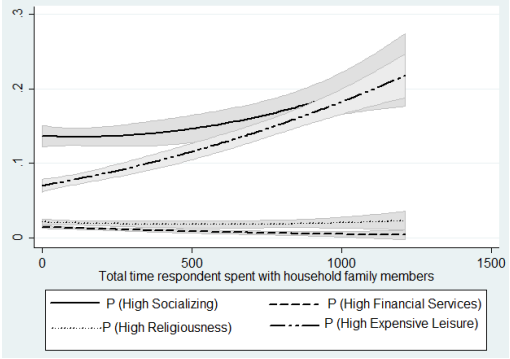
Predicted probabilities based on age of retirees



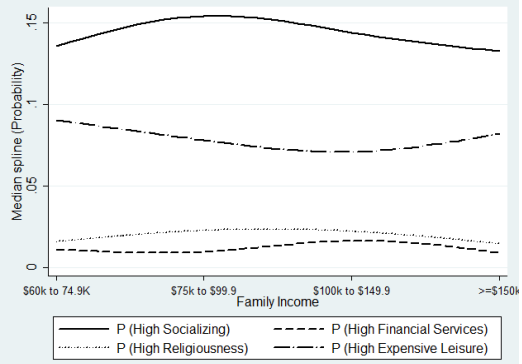
Predicted probabilities based on family time



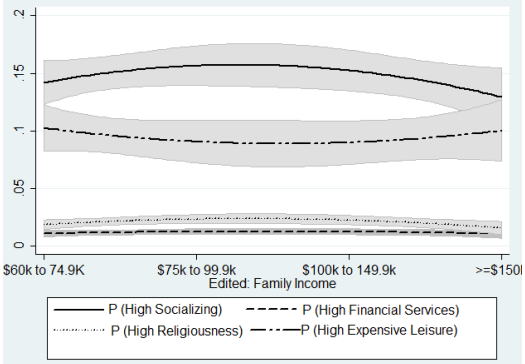
Predicted probabilities based on family time



Predicted probabilities based on family income



Predicted probabilities based on family income



Finding happiness, wellbeing and stress in socializing, expensive leisure, religiosity and financial services.

The results from table 3 show that the self-reported individual stress and well-being are highly associated with the time spent on financial services. Inexpensive leisure and religiosity are negatively associated with happiness.

Table 3: Regression results

	Happy		***Stress***		***Well being***	
Variable	***Coef***	**P>|z|**	***Coef.***	**P>|z|**	***Coef.***	**P>|z|**
Socializing	0.0009	0.460	-0.0011	0.510	0.0004	0.840
Financial Services	-0.0704	0.250	**0.2950*** **	**0.000**	**0.3430*** **	**0.001**
Religiosity	**-0.0102***	**0.060**	0.0076	0.250	-0.0048	0.420
Leisure (Expensive)	-0.0057	0.240	0.0039	0.470	-0.0014	0.720
Leisure (Inexpensive)	**-0.0015****	**0.050**	0.0010	0.270	-0.0005	0.710
Age	-0.0057	0.350	0.0018	0.810	-0.0028	0.750
Sex	0.3177	0.240	0.2799	0.400	0.1831	0.610
White	-0.1498	0.850	-1.0314	0.420	0.8914	0.470
Hispanic	-0.2886	0.750	-2.0220	0.140	0.4720	0.770
Black	0.2441	0.770	-1.8501	0.160	0.0843	0.950
Education	-0.0976	0.110	0.0660	0.280	0.0612	0.420
Family Income	-0.0369	0.760	0.2044	0.190	0.1472	0.340
Time spent with Family	-0.0003	0.590	0.0001	0.870	-0.0012	0.160
Time spent with spouse	0.0004	0.670	-0.0010	0.260	0.0002	0.870

Significant at * 10% level, **5% level and ***1% level

Conclusion

The findings from this study indicate that time spent on financial services is positively associated with well-being at retirement. This may indicate that retired people feel more content when they handle their own finances. This may also indicate that contented wealthier retirees are inclined to participate in financial activities such as investments and estate planning. In that case, these findings are consistent with Browning et al. (2016), who report that wealthier retirees reinvest the required minimum distributions (RMD) from their retirement account into financial assets (Smith & Love, 2007) and spend much less money than expected in old age. The time spent on financial services activities is positively associated with the time spent with the family and with well-being during retirement, possibly showing care for the family and satisfaction associated with it. However, time spent on financial services activities is negatively associated with the time spent with spouse, possibly because married or partnered households divide the work, thus requiring a lesser amount of time spent on activities such as personal financial management.

Another surprising result is that the time spent on financial services is highly associated with self-reported stress. This may be due to either a pre-existing association with participation in financial activities or the amount of time spent on financial activities increasing stress for some retirees. Although more research is necessary to understand this linkage between stress and time spent on finances, previous studies have found that financial education can improve the perceived financial wellness of individuals (Garman et al., 1991). Policies that encourage greater community involvement for the elderly might play a role in improving the retirees' overall sense of well-being and reduce the stress associated with the performance of different activities, including time spent on personal finances. Participation in a financial education program may reduce retirees' lack of information and the stress associated with managing one's finances. Also, teaching the importance of time spent with spouse, partners, and family members should be made part of educational programs that prepare individuals approaching retirement, as preparation for the next stage of their life.

The increase in time spent on religiosity is positively associated with higher family income. However, there is a negative association between religiosity and self-reported happiness. Examining the causality of whether unhappy retirees are turning to religiosity or whether time spent on religiosity makes people unhappy is outside the scope of this study, but the

significant association between religiosity and unhappiness found in this study warrants further examination. Another possibility is that different religious activities have different effects**.** For example, donations may make some wealthy retirees feel content and others feel unhappy. Similarly, driving to church may bring happiness to some and inconvenience to others. Inexpensive leisure that includes sedentary and sometimes solitary activities such as resting, watching TV, listening to music, playing games on a computer, arts and crafts, reading or writing for personal interest, attending performing arts, watching movies in a theater, and similar activities is negatively associated with self-reported happiness. More research is needed to understand this association. The findings from this study provide a basic understanding for the readers of this journal, scholars of household finance, and policy makers that time spent on different activities is associated with happiness, stress, and the perceived sense of well-being that an individual may have. More research is needed to develop this line of work to better understand the nuances associated with the time spent on various activities and the financial and mental well-being of households.

References

Brown, Alexandra Milligan. "The Meta Marriage: Links between Older Couples' Narratives and Marital Satisfaction." PhD diss., Auburn University, 2014.

Browning Chris, Guo, Tao, Cheng Yuanshan, Michael S Finke. "Spending in Retirement: Determining the Consumption Gap", *Journal of Financial Planning* 29, No. 2 (2016)

Bureau of Labor Statistics (2016). ATUS survey documentation. Available at: http://www.bls.gov/tus/#documents

Finke, Michael S., Sandra J. Huston, and Danielle D. Winchester. "Financial advice: Who pays." *Journal of Financial Counseling and Planning* 22, no. 1 (2011): 18.

Garman, E. Thomas, Jinhee Kim, Constance Y. Kratzer, Bruce H. Brunson, and So-Hyun Joo. "Workplace financial education improves personal financial wellness." *Journal of Financial Counseling and Planning* 10, no. 1 (1999): 81.

Harness, Nathaniel, Swarn Chatterjee, and John Salter. "Financial Planning for the Middle Market." *Journal of Financial Planning* 27, no. 4 (2014): 48-54.

Kalenkoski, Charlene M., and Eakamon Oumtrakool. "How Retirees Spend Their Time: Helping Clients Set Realistic Income Goals." *Journal of Financial Planning* 27, no. 10 (2014).

Marhankova, Jaroslava Hasmanova. "'Women are just more active'–gender as a determining factor in involvement in senior centres." *Ageing and Society* 34, no. 09 (2014): 1482-1504.

Passias, Emily, Liana C. Sayer, and Joanna R. Pepin. "Who Experiences Leisure Deficits? Mother's Marital Status and Leisure Time." (2015).

Smith, Paul A., and David A. Love. "Measuring Dissaving Out of Retirement Wealth." *Available at SSRN 968431* (2007).

Early Child Health Intervention and Subsequent Academic Achievement

Bhavneet Walia, Syracuse University

Abstract

Low birthweight (LBW) status is a prominent health concern in the care of infants. The World Health Organization estimates that 15.5 percent of infants are LBW (i.e., less than 5.5 pounds) and that LBW status contributes to 60 to 80 percent of neonatal deaths. Given the prevalence of and risk factors associated with LBW status, hospitals in the United States and in many other parts of the world typically maintain a special treatment protocol for LBW infants. By comparing National Longitudinal Survey of Youth 1979 Children and Young Adult cohort infants below and above the LBW threshold, we find a significant effect of LBW treatment upon scholastic achievement. Controlling for birthweight—a variable that is found to contribute significantly to scholastic achievement—and other variables, LBW treatment is estimated to add between 0.24 and 0.49 years to an individual's subsequent years of schooling. Given that the return on an additional year of schooling is typically estimated to be an 8 to 13 percent wage premium, we estimate proportionally that the LBW treatment leads to a wage premium of between (roughly) 2 and 6.5 percent. This effect is dramatic and is equivalent in effect to the addition of at least 2.5 pounds in infant birthweight.

Introduction

The effect of early health interventions impact upon subsequent child development is an important question that has economic, educational, and human developmental implications. In aggregate, significant resources are utilized to improve early childhood health. However, large disparities (e.g., in access to early health interventions) remain both between and within countries. In the United States, for example, maternal and neonatal stays represent 27 percent of hospital costs for patients insured through Medicaid, 15 percent of costs for privately-insured patients, and 4 percent of costs for uninsured patients. These numbers suggest both substantial resource allocation toward maternal and neonatal care, as well as potentially large disparities in access to infant care (Agency for Healthcare Research and Quality, 2014). The Center for Disease Control (2004) suggests that access to quality neonatal care likely differs by race in the U.S.

The *World Health Organization* estimates that 15.5 percent of infants—nearly 1 in 6—are low birthweight or LBW (< 5.5 pounds at birth) and that LBW status contributes to between 60 and 80 percent of neonatal deaths. While an estimated 96.5 percent of these births occur in developing countries, there remain a large number of such births (approximately 700,000 per year) in developed countries. Given the extent of early health interventions, it is important to establish their long-term effects of early health interventions upon child development.

In a study of very low birthweight (VLBW) infants born in the U.S., Almond *et al*. (2010) find significantly lower mortality rates among those just below the VLBW birthweight threshold (1,500 grams) relative to those just above. This result is important, given that mortality risk tends to decrease with birth weight. Almond et al. conclude that extra care at hospital is causing this unexpected result. Those just below the threshold spend incur 10 percent higher hospital bills than those just above. Bharadwaj *et al*. (2013) adopt a similar in studying VLBW infants born in Chile and Norway approach and expand the findings of Almond et al. (2010). They address the effect of improved neonatal health care on mortality and long run academic achievement.

In the U.S., hospitals typically establish treatment protocols not only for VLBW infants (<1,500 grams) but also for LBW infants (<2,500 grams) (Stanford Children's Health, 2016). These protocols often include additional guidelines for oxygen therapy, cardiovascular monitoring and care, fluid and electrolyte regulation, nutrition, trophic (gut stimulatory) feedings, blood glucose regulation, and prevention of infection (UCSF Children's Hospital, 2004). In addition to substantial short-term benefits in terms of higher LBW infant survival rates, these treatment protocols may have unintended but desirable child development benefits to both VLBW and LBW infants in the U.S.

Within the present study, we utilize birth and subsequent scholastic outcomes from the National Longitudinal Survey of Youth 1979 Children and Young Adult Cohort data set to estimate the long-term effects of LBW treatment protocols upon an individual's subsequent years of schooling. By comparing NLSY infants below and above the LBW threshold, we find a significant effect of LBW treatment upon scholastic achievement. Controlling for birthweight—a variable that is found to contribute significantly to scholastic achievement—and other variables, LBW treatment is estimated to add between 0.24 and 0.37 years to an individual's subsequent years of schooling. This effect is dramatic and is equivalent in magnitude to the addition of more than 2.5 pounds in infant birthweight. That is to say, a 5.49-pound (LBW) infant is expected to obtain

significantly more schooling than a 5.5-pound (non-LBW) infant and at least as much schooling as a 7-pound infant due to the treatment effect.

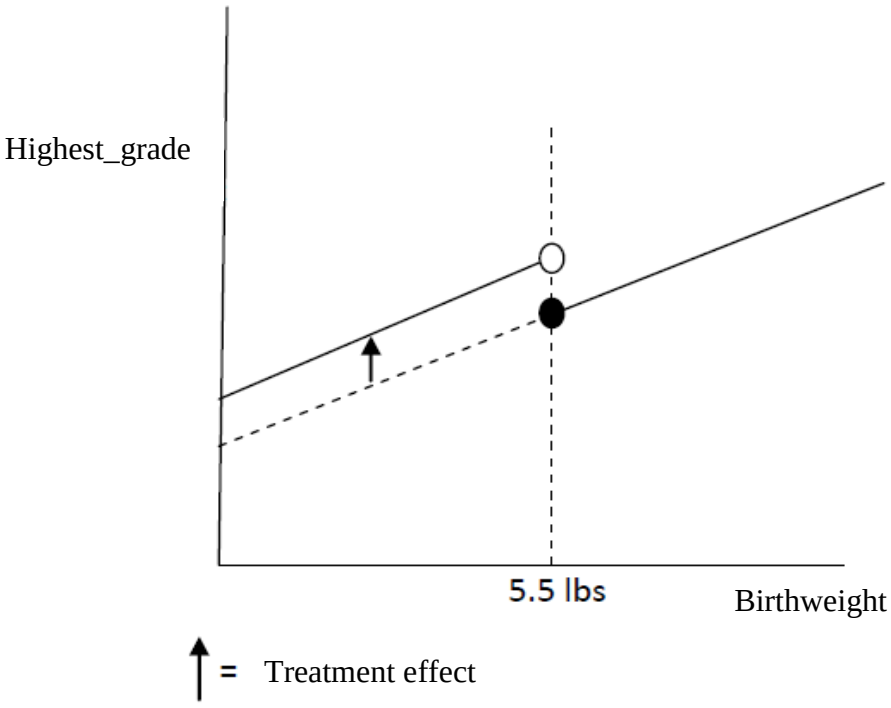
Empirical Model and Data Analysis

The study uses data from the NLSY79 Children and Young Adult cohort data set, a nationally represented sample of 11,512 Americans born to NLSY79 mothers. Birthyears in the data set range between 1970 and 2009 with a median birthyear of 1985. The data was collected from interviews and surveys completed by NLSY79 mothers and their offspring, where offspring completed surveys on an annual basis.

To estimate the effect of additional neonatal treatment for LBW children upon scholastic achievement, we consider the independent variable *highest_grade* (highest schooling grade completed). To model variation in this variable, we conduct series of stepwise ordinary least squares regressions that feature a) month-of-birth fixed effects and b) region-of-residence fixed effects. The set of control variables for these regressions include *m_age* (mother's age at time of subject's birth), *child_female* (whether or not the subject is female, *birth_order* (one greater than the number of older siblings the subject has), *m_earn* (mother's annual income prior to subject's birth), smoke (whether the mother or another household member smoked in the 12 months prior to subject birth), *alcohol* (whether the mother or another household member drank alcohol in the 12 months prior to subject birth), *num_ill* (number of times subject was ill in first year of life), *m_school* (mother's highest schooling grade completed), *prenatal* (whether subject received a prenatal checkup), and a *race* dummy variable that separates white and non-white subjects.

Independent variables of interest include *birthweight* (birthweight in ounces) and *LBW* (a dummy variable indicating whether the child is low birthweight or below 5.5 pounds at birth). Given past literature, we expect birthweight to contribute positively to educational attainment. Given U.S. hospital protocols for LBW infants, however, we also expect a compensatory treatment effect, whereby the trend line for LBW infants shifts upward in response to cognitive developmental returns from additional neonatal treatment. Figure 1 displays the hypothesized effects of birthweight and LBW treatment.

Figure 1: Hypothesized effect of birthweight upon subsequent scholastic achievement



The figure hypothesizes that those infants just below the LBW threshold (e.g., those infants at 5.49 pounds) may have a higher expected educational attainment than those just above the threshold (e.g., those infants at 5.5 or even 6 pounds). The possibility

of such a result depends upon the presence and magnitude of a positive treatment effect among LBW infants. We present summary statistics for the dependent variable and for the independent variables of interest below.

Table 1: Summary Statistics for Dependent Variable and Key Independent Variables

Variable	Mean	Std. Dev.	N
Highest_grade	12.22	2.27	6096
Birthweight (ounces)	116.15	22.82	10289
LBW	0.083	0.275	10289

These summary statistics reveal that approximately 8.3 percent of sampled infants are categorized as low birthweight. The average birthweight in sample is 116.15 ounces or approximately 7.25 pounds. The average highest grade attained is 12.22. In other words, the average level of educational attainment within the sample is a fraction of a year more than high school graduation. We now report two sets of stepwise regression results. The first set of ordinary least squares regression models feature month-of-birth fixed effects. The second set of OLS regression models feature region-of-residence fixed effects. For each set of regressions, a Hausman test specifies the use of fixed effects over random effects. Regression results are provided in Tables 2a and 2b in the Appendix.

These tables present evidence that birthweight positively and significantly influences subsequent scholastic achievement. Specifically, each additional ounce of birthweight is expected to increase a subject's highest grade achieved by between .004 and .006 years, *ceteris paribus*. That is, each additional pound of birthweight is expected to increase a subject's highest grade achieved by between .064 and .096 years, *ceteris paribus*. Despite this biological effect, the LBW treatment effect is consistently significant and substantial, accounting for between 0.24 and 0.49 additional years of schooling. This effect is dramatic and is equivalent in effect to the addition of at least 2.5 pounds in infant birthweight across the 20 models tested. The long-term LBW treatment effect upon scholastic achievement, then, is substantial. Given that the return on an additional year of schooling is typically estimated to be an 8 to 13 percent wage premium (Kolesnikova, 2010), we can make the proportional estimate that the LBW treatment leads to a wage premium of between (roughly) 2 and 6.5 percent, where 2 percent (6.5 percent) is approximately equal to an 8 percent premium (13 percent premium) multiplied by a 0.24 year increase (0.49 year increase) in years of schooling.

Conclusion

By controlling for demographic and family characteristics upon scholastic achievement, we are able to isolate the respective effects of birthweight and additional hospital treatment for low birthweight infants upon scholastic achievement. Birthweight positively and significantly influences subsequent scholastic achievement. Despite this biological effect, the low birthweight treatment effect is consistently significant and substantial, accounting for between 0.24 and 0.49 additional years of schooling. This effect is dramatic and is equivalent in effect to the addition of at least 2.5 pounds in infant birthweight across the 20 models tested. We further estimate that LBW treatment leads to a wage premium of between (roughly) 2 and 6.5 percent. Future research might consider the effect of low birthweight treatment upon other important characteristics, such as subsequent child health.

References

Almond, D., Doyle Jr, J. J., Kowalski, A. E., & Williams, H. (2010). Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns. *The Quarterly Journal of Economics*, *125*(2), 591.

Bharadwaj, P., Løken, K. V., & Neilson, C. (2013). Early Life Health Interventions and Academic Achievement. *American Economic Review*,*103*(5), 1862-91.

Kolesnikova, N. (2010). The return to education isn't calculated easily. *The Regional Economist*, *18*(1), 12-13.

Moore, B., Levit, K., Elixhauser, A. (2014). Costs for Hospital Stays in the United States, 2012. *Agency for Healthcare Research and Quality*. Statistical Brief #181.

Stanford Children's Health (2016). Low Birthweight. Available online at: http://www.stanfordchildrens.org/en/topic/default?id=low-birthweight-90-P02382

UCSF Children's Hospital (2004). Intensive Care Nursery House Staff Manual. Available online at: https://www.ucsfbenioffchildrens.org/pdf/manuals/20_VLBW_ELBW.pdf

World Health Organization. (2016). Care of the preterm and/or low-birth-weight newborn. Available online at: http://www.who.int/maternal_child_adolescent/topics/newborn/care_of_preterm/en/

Appendix

Table 2a: Stepwise OLS Regression with Month-of-Birth Fixed Effects

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
LBW	0.349***		0.489***	0.371***	0.372***	0.306**	0.238*	0.239*	0.263**	0.264**	0.272**
	(3.221)		(3.533)	(2.818)	(2.836)	(2.340)	(1.846)	(1.856)	(2.023)	(2.035)	(2.021)
birthweight		0.004***	-0.000	0.004**	0.004**	0.005***	0.006***	0.006***	0.005***	0.005***	0.004**
		(2.726)	(0.171)	(2.125)	(2.249)	(3.202)	(3.392)	(3.379)	(2.768)	(2.715)	(2.294)
m_age				-0.147***	-0.148***	-0.148***	-0.123***	-0.122***	-0.131***	-0.132***	-0.166*
				(24.374)	(24.646)	(24.700)	(19.792)	(19.544)	(20.496)	(20.629)	(24.009)
race					-0.416***	-0.411***	-0.331***	-0.331***	-0.351***	-0.345***	-0.174**
					(6.059)	(6.033)	(4.899)	(4.911)	(5.070)	(4.974)	(2.384)
child_female						0.488***	0.494***	0.495***	0.494***	0.492***	0.501***
						(8.656)	(8.884)	(8.901)	(8.794)	(8.749)	(8.681)
birth_order							-0.341***	-0.344***	-0.316***	-0.312***	-0.211***
							(12.566)	(12.615)	(11.310)	(11.178)	(7.101)
m_earn								-0.000	-0.000	-0.000	-0.000**
								(1.137)	(1.473)	(1.477)	(2.312)
smoke									-0.420***	-0.415***	-0.224***
									(6.603)	(6.528)	(3.369)
alcohol									0.368***	0.365***	0.284***
									(6.278)	(6.225)	(4.709)
prenatal										0.747***	0.855***
										(3.008)	(3.326)
m_school											0.161***
											(12.874)
N	6096	5604	5604	5604	5604	5604	5604	5604	5439	5437	4958

* p<0.10, ** p<0.05, *** p<0.01

Table 2b: Stepwise OLS Regression with Region-of-Residence Fixed Effects

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Birthweight	0.004***	-0.000	0.000	0.004**	0.004**	0.006***	0.006***	0.006***	0.004**
	(2.883)	(0.006)	(0.029)	(2.380)	(2.418)	(3.365)	(3.360)	(3.253)	(2.389)
LBW		0.478***	0.486***	0.348***	0.348***	0.284**	0.285**	0.290**	0.311**
		(3.466)	(3.453)	(2.602)	(2.612)	(2.142)	(2.146)	(2.190)	(2.293)
Num_ill			-0.034	-0.010	-0.006	0.013	0.013	0.002	-0.080
			(0.556)	(0.179)	(0.102)	(0.221)	(0.227)	(0.031)	(1.356)
m_age				-0.150***	-0.152***	-0.151***	-0.151***	-0.152***	-0.185***
				(24.394)	(24.654)	(24.758)	(24.621)	(24.705)	(28.713)
race					-0.394***	-0.382***	-0.382***	-0.372***	-0.174**
					(5.256)	(5.124)	(5.127)	(4.988)	(2.243)
child_female						0.487***	0.488***	0.483***	0.483***
						(8.485)	(8.488)	(8.408)	(8.282)
m_earn							-0.000	-0.000	-0.000*
							(0.331)	(0.355)	(1.760)
Prenatal								0.949***	0.945***
								(3.691)	(3.610)
m_school									0.194***
									(16.269)
N	5569	5569	5362	5362	5362	5362	5362	5355	4885

* p<0.10, ** p<0.05, *** p<0.01