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

PARKER
COLLEGE OF BUSINESS
GEORGIA SOUTHERN UNIVERSITY
Department of Economics
11935 Abercorn Street
Savannah, GA 31419

THE ACADEMY OF ECONOMICS AND FINANCE JOURNAL
VOLUME 10, 2019

ACADEMY of ECONOMICS and FINANCE JOURNAL

AEF

VOLUME 10 2019

editors
DR. CLAY MOFFETT
DR. MICHAEL TOMA

ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 10 2019

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 The submission of work to the journal will imply that it contains original work, unpublished elsewhere, that is not under review, nor a later candidate for review, at another journal.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 10:

Volume 10 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-Fifth Annual Meeting of the AEF in St. Petersburg, FL. from February 6-9th, 2019. The program was arranged by Dr. Olivier Maisondieu Laforge, University of Nebraska Omaha.

ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 10 **Table of Contents** 2019

How Presidential Tweets Affect Financial Markets

***Sasha Cove and Marc Sardy*, Rollins College**

Abstract

This paper examines the effect of President Trump's Tweets and their impact on financial markets as measured by the Market proxy of the S&P 500. Data has been collected on Trumps tweets since his presidency began until fall 2018, Market data is based on daily S&P results for the daily returns. Our findings are that Trumps tweets have an impact on financial markets even at the level of daily returns.

Introduction

The manner in which information is disseminated to the masses has evolved with the introduction of social media platforms. Individuals are now receiving their daily news in the form of Facebook updates, Instagram posts, and Twitter 'tweets.' The average Twitter user can see around 1,000 tweets in the fourteen minutes that they spend on Twitter in a day, meaning that users often spend only seconds on a given tweet (Pancer, E., & Poole, M. 2016). Twitter, in particular, came to the foreground during the 2016 presidential election, as candidates used the platform to increase their media coverage and garner public support. This method proved extremely successful for the current president of the United States, Donald Trump, who achieved unprecedented media coverage through his Twitter activity. His current position gives his tweets the power to influence public sentiment, and thus the performance of the financial markets.

Social media use is becoming integral to politics, not only as a method of keeping up with political news, but also as a mechanism through which individuals' political views and involvement are being impacted. A survey conducted by the Pew Research Center found that 25% of social networking site (SNS) users said that the sites are 'very important' or 'somewhat important' to them in discussing or debating political issues with others, 25% of users say that they have become more active in a political issue after discussing or reading about it online, and that 16% of users say that they have changed their views about a political issue after discussing or reading about it online (Rainie, L., & Smith, A. 2014). Furthermore, Social media's use as a media outlet is will continue to grow in importance in coming years, as social media is most used by the younger generations. As this generation ages, they are beginning to comprise a growing portion of the voting population, indicating that social media will maintain its relevance in politics. The use of social media in politics has become a method through which the younger generation can be engaged; there is a strong positive correlation between social media use and youth political engagement in users between the ages of 16 and 29 (Michael Xenos, Ariadne Vromen & Brian D. Loader 2014).

The emergence of the new hybrid media system has changed the 'currency' of the media system as a whole. It is no longer the ability to communicate to mass audiences, but rather the ability to attract attention from them in a world where the average individual is constantly inundated with massive amounts of information on a day to day basis. In our new 'attention economy,' media organizations and platform compete for views, as do the issues, actors, causes, and factions that comprise our political sphere (Zhang, Y., Wells, C., Wang, S., & Rohe, K. 2018). According to Zhang et al. in "Attention and amplification in the hybrid media system: the composition and activity of Donald Trump's Twitter following during the 2016 presidential election," attention's power is derived from its four characteristics. It is necessary to change or mobilize the opinions of an audience, as it gives an actor access to the 'socialized communication' that lends an audience to its depiction. It can be a mechanism for translating communication's power into action; politically, it can be converted into civic engagement. It is a transferrable currency, as the holder of attention has the power to direct said attention to another. Finally, it is powerful because others perceive it as powerful; it holds a 'popularity bias,' as attention tends to beget attention. Zhang et al. also introduced the idea of social media amplification, which are the actions of social media users that increase measures of engagement or audience metrics, the likes, @replies, follower number, and retweets a Twitter account receives. Twitter's software uses these measures to engagement to determine what is 'trending,' what to feature on the platform's homepage and what will appear at the top of a user's newsfeed. In this sense, the attention given to a tweet is exponential; with each retweet the tweet is disseminated to the followers of the account that retweeted it, causing more retweets, more likes, more attention.

The amount of attention an individual receives on social media has become an indicator of social status, of 'worthiness,' and this 'worthiness' is often translated to news-worthiness (Zhang, Y., Wells, C., Wang, S., & Rohe, K. 2018). As a tweet begins to 'trend,' news platforms feel required to cover it, if not only to beat their competitors to the punch. As such, the tweets of socially relevant individuals have been increasingly covered in the news, providing the tweets with validity, and further increasing the attention the tweets receive.

Social media activity levels can be predictive of stock performance. A study researched the correlation between intra-day responses of the stock market and the spread of news on Twitter. Unusually high spikes in Twitter activity regarding 96 firms

on the NASDAQ over 193 trading days were statistically compared with Yahoo! Finance data of the related firms over the forty minutes following the spike. The study found a spike in tweets per minute resulted in a spike in trading activity in the following forty minutes (Tafti, A., Zotti, R., & Jank, W. 2016). For an individual, tweeting is fast and effortless but has no immediate consequences on financial markets. However, if said individual is the president, the act of tweeting remains fast and effortless but has the potential to cause immediate financial consequences. It becomes less necessary that there be a spike in activity because the amount of people that are viewing presidential tweets is significantly larger, as is the content of the tweet's potential implications on the markets.

The concept that not everyone on Twitter holds the same influence on market sentiment is explored in a study examining the influence of a financial community, a group of users whose interests align with the financial markets. (Yang, S., Mo, S., & Liu, A. 2015) The community comprised a central group of 50 well-recognized investment experts, 25 influential traders and 25 accounts associated with the seven financial news providers, and their followers. By determining the sentiment of the central group's tweets, disseminated by their followers, a robust correlation was determined between the group's sentiment and financial market movement measured by lagged daily prices ((Yang, S., Mo, S., & Liu, A. 2015). It can be argued that because users are harvesting information from these influential sources in order to make their daily trading decisions, the financial community was exerting influence on the financial markets. It can be rationally inferred that the influence on public sentiment that the financial community had can be compared to the influence that Trump's Twitter has as a single central group, amplified by both his supporters and opposers, paired with the resulting media coverage of his tweets.

As President, Trump's tweets not only garner massive attention, but also have the potential to have significant global implications. If an average individual were to tweet that increasing tariffs would be good for the country, the tweet would have no relevance to the markets. However, if Trump were to tweet the a similar statement, there is a significant chance that tariffs may in fact be increased, which would have a significant impact to the markets. We propose a correlation between when Trump tweets about the economy and movements in the financial markets.

H1. Tweets posted by Trump that are related to the economy are positively correlated to movements in the S&P 500 and the VIX.

Early research on predicting financial markets was centered around the efficient market hypothesis (EMH), which argues that financial market valuations incorporate all existing, new, and hidden information because investors act as rational agents who seek to maximize profits (Mao, H., Counts, S., & Bollen, J. 2011). The collective investment decisions of investors use the wisdom of crowds to determine stock market prices, driven mainly by new information, such as news, rather than present and past prices. EMH has been challenged by behavioural finance, as research has shown that stock prices do not follow a random walk and can be predicted to an extent using early indicators such as Twitter feeds (Bollen, J., Mao, H., & Zeng, X. 2010). EMH is further challenged by behavioural finance research which emphasizes the role of behavioural and emotional factors in financial decision making (Mao, H., Counts, S., & Bollen, J. 2011).

Studies have determined that public mood state can be tracked from the content of large-scale Twitter feeds, and that changes in public mood correlate to shifts in the Dow Jones Industrial Average (DJIA) values, which occur with a lag of three to four days. (Bollen, J., Mao, H., & Zeng, X. 2010). The correlation between the positive and negative mood of the masses on Twitter and stock market indices, including VIX, DJIA, S&P 500, and NASDAQ have also been examined. Collective emotions were categorized by 'mood words' such as fear, worry, and hope, and it was found that both positive and negative moods were positively correlated with all four indices (Zhang, Fuehres, & Gloor. 2011). This result is relevant because VIX performance is strongly negatively correlated with the DIJA, S&P 500, and NASDAQ, implying that people use more emotional words during times of uncertainty, whether it is positive or negative in a financial context. The idea of public mood as a financial market predictor can be furthered by looking at the sentiment toward specific stocks rather than general public sentiment. A study tracked public sentiment towards 30 DIJA companies over a 15 month period and found a strong statistical correlation between stock price returns and Twitter sentiment and volume of tweets regarding the companies (Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. 2015).

Research has found that negatively worded tweets are liked and shared more often than positively worded tweets, information which arguably furthered Trump's media coverage during the election. A study conducted by Pancer et al. found that during the 2016 election, posts relating to policy such as the economy, the environment, or labour, decreased the post's popularity by about 1,800 likes. Conversely, those concerning inflammatory rhetoric increased popularity related to a 2,600 like increase, which they termed the content valence effect. The study found that the popularity of Trump's tweets was improved by mentioning disasters and religion, but decreased when mentioning social issues, and that the content valence effect was present for Trump, but not for Clinton during the election (Pancer, E., & Poole, M. 2016). This had a corruptive effect on the content being tweeted at the time, an effect which has extended into the presidential term.

Reaching a larger audience is made possible by posting socially inflammatory content, and this has led to tweets which have the power to cause shocks in the global financial markets. We propose a correlation between the sentiment of Trump tweets relating to the economy and movements in the financial markets.

H2. The sentiment of Tweets posted by Trump that are related to the economy is positively correlated to movements in the S&P 500 and the VIX.

Data

Our analysis is conducted on the six month time period of July 1, 2018 to December 31, 2018. In the analysis we investigate the relationship between market data and Twitter data. The details of both are given in the remainder of this section. The market and twitter data is available at https://doi.org/10.6084/m9.figshare.7679363.v1.

Market data

The first source of data contains information on the performance of the financial markets, utilizing two stock market indices, the S&P 500 and the CBOE Volatility Index (VIX), indicators of financial market performance. We compiled the day-to-day data of both indices, each day's open, close, and percent change, for the six month period. This data is publicly available and can be downloaded for various Internet sources, such as the Yahoo! Finance website (i.e. https://finance.yahoo.com/quote/%5EGSPC?p=^GSPC for the S&P 500). As the markets are closed on the weekends, and Trump tweeted over the weekends, Saturdays and Sundays were counted as a cumulative 'day,' with their market open being that Friday's close, and their market close being that Monday's open; the percent change over the weekend was then calculated. The same method was utilized for the five Market holidays that occurred during the sample time period, using the close of the day before and the open of the day after in order to calculate the percent change over the holiday.

Twitter data

The second source of data is from Twitter, and consists of the Tweets posted by Donald Trump from his personal twitter account @realDonaldTrump, and the sentiment of each Tweet. The data was collected from Export Tweet (https://www.exporttweet.com/), and contained the date, content, retweets, and likes. In the 6 month period of study, Trump tweeted a total of 1,989 times. Saturdays and Sunday tweets were combined into one 'day' to account for the lack of weekend market data. First, we determined total amount of tweets made per day. Then, each tweet determined to be related to the economy was identified, a total of 284 economy-related tweets, and the amount of economy-related tweets made per day was determined. The percent of tweets per day that were economy-related was then calculated.

A Tweet sentiment was determined for each of the economy-related tweets. Sentiment was evaluated according to the that of the author, Trump; each tweet was assigned one of three sentiment labels, positive, neutral, or negative, determined by whether he had worded it in a positive, neutral, or negative tone. The daily tweet sentiment was then calculated by assigning each sentiment label with a score; positive: +1, neutral: 0, negative -1, and summing the sentiment of each economy-related tweet of the day. The more positive tweets made in a day, the higher the score, and the more negative the lower; neutral statements had no effect on the day's cumulative sentiment score.

Methodology

We performed two regression analyses in order to determine the relationship between market performance and Trump's economy-related tweets.

Figure 1: Correlation between S&P performance and the % of economy related tweets

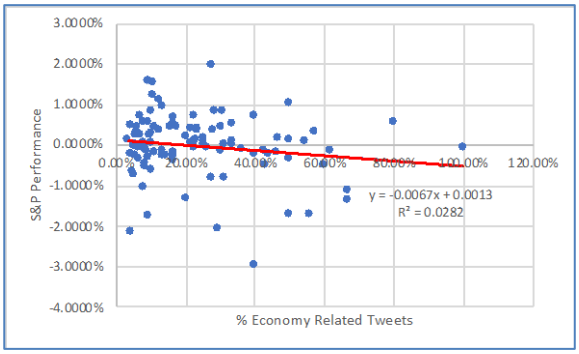
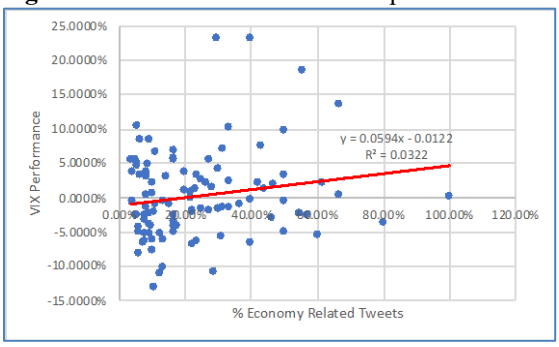


Figure 2: Correlation between VIX performance and the % of economy related tweets



Sentiment Analysis

In order to examine how sentiment was viewed we used a measure of S&P movements and VIX movements:

Table 1: The net impact of daily tweet sentiment on S&P 500 movement

Sentiment	Total	S&P Up Days	S&P Down Days	% Up	% Down
Positive	75	42	33	56.00%	44.00%
Negative	20	9	11	45.00%	55.00%
Neutral	9	2	7	22.22%	77.78%

Table 2: The net impact of daily tweet sentiment on VIX movement

Sentiment	Total	VIX Up Days	VIX Down Days	% Up	% Down
Positive	75	34	41	45.33%	54.67%
Negative	20	9	11	45.00%	55.00%
Neutral	9	4	5	44.44%	55.56%

Results

Market moves do not seem to be independent of Trumps comments with bot the S&P Moves and VIX moves showing low significance with regard to independence. So there seems to be some effects that are significant as shown in our regression analysis of tweets vs S&P and VIX. The equation of S&P moves vs Trumps tweets about the economy is:

$$\text{S\&P (move)} = 0.013 - .0067^{*} \text{ (Tweets) with an } R^2 = .0282 \quad (1)$$

While the explanatory power for the tweets are low there is a significant but negative effect on Trump tweets and the effects on the S&P 500. There are several reasons as to why the results may not be very large. Firstly, we were using daily returns and the instantaneous tweet information. There are a lot of things that happen on a daily basis that can have an impact on the S&P. The effects of tweets can be diffused over the course of the day. Future study will look at more accurate timing of market moves with regard to the timing of the tweets. Furthermore, it seems that every time trump tweets up about the economy the S&P has a 56% change of positive movement and a 45% chance of the S&P declining. With regard to negative Tweets, there is a 55% (10% margin of error) chance of a down move in the S&P and a 45% (10% margin of error) of an up move.
The VIX data regression shows similar results with the VIX moving positively with Trump's Economic tweets.

$$\text{VIX (Move)} = -0.0122 + 0.0594 \text{ (tweet) With an } R^2 \text{ of } 0.0322 \quad (2)$$

While the explanatory power for the tweets with respect to the VIX are low there is a significant but positive effect of Trump tweets and the effects on the VIX. There are several reasons as to why the results may not be very large. Firstly, once again we

were using daily returns and the instantaneous tweet information. There are a lot of things that happen on a daily basis that can have an impact on the VIX. The effects of tweets can be diffused over the course of the day. Future study will look at more accurate timing of market moves with regard to the timing of the tweets. Furthermore, it seems that every time trump tweets up about the economy the VIX has a 45% (10% margin of error) chance of positive movement and a 55% (10% margin of error) chance of the VIX declining. With regard to negative Tweets, there is a 55% chance of a down move in the VIX and a 45% (10% margin of error) of an up move.

Discussion/Conclusion

Much of our findings are limited by the nature of daily S&P 500 returns. There is a gross mismatch between the timing of the tweets and the daily results. If we had more frequent S&P results we would be more directly able to see the instantaneous effect of the tweets on the market. That said, we do actually see an impact of tweets on the market but it is considerably smaller then we might see with more micro-structured data.

Our results showed that the tweets related to economic issues have an impact on the market. Up tweets raise the market 56% ±10% of the time and down tweets move the market 55%±10%. In addition to moves in the S&P, the VIX is also affected by the tweets. The VIX is up when President Trump tweets about the economy. The market Volatility goes up by 45%±10%. when President Trump tweets up about the economy. The market volatility also changes when President Trump tweets down about the economy The VIX potentially moves down by 55%±10%. or up by 45%±10%. Overall there seems to be a direct impact of tweets and market moves that are significant and impactful.

References

Bollen, Johan ; Mao, Huina ; Zeng, Xiao-Jun, (2011), Twitter mood predicts the stock market, Journal of Computational Science, 2(1), March 2011, Pages 1-8

Himelboim, Itai ; Smith, Marc A ; Rainie, Lee ; Shneiderman, Ben ; Espina, Camila, (2014) Classifying Twitter Topic-Networks Using Social Network Analysis Social Media + Society, February Vol.3(1)

Mao, Huina ; Counts, Scott ; Bollen, Johan , (2011),Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data, IDEAS Working Paper Series from RePEc,

Pancer, Ethan ; Poole, Maxwell, (2016) The popularity and virality of political social media: hashtags, mentions, and links predict likes and retweets of 2016 U.S. presidential nominees' tweets, Social Influence, 01 October Vol.11(4), p.259-270

Rainie, L., & Smith, A. (2014, February 18). Politics on Social Networking Sites. Retrieved from http://www.pewinternet.org/2012/09/04/politics-on-social-networking-sites/

Ranco, Gabriele ; Aleksovski, Darko ; Caldarelli, Guido ; Grčar, Miha ; Mozetič, Igor (2015),The Effects of Twitter Sentiment on Stock Price Returns, PLoS One, Sept, Vol.10(9), p.e0138441

Tafti, Ali ; Zotti, Ryan ; Jank, Wolfgang, (2016) Real-Time Diffusion of Information on Twitter and the Financial Markets.(Report) PLoS ONE, August 9, 2016, Vol.11(8), p.e0159226

Vromen, Ariadne; Xenos, Michael A. ; Loader, Brian, (2014) Young people, social media and connective action: from organisational maintenance to everyday political talk, Journal of Youth Studies, 01 July, p.1-21.

Yang, Steve Y.; Mo, Sheung Yin Kevin ; Liu, Anqi, (2015), Twitter financial community sentiment and its predictive relationship to stock market movement, Quantitative Finance, 03 October Vol.15(10), p.1637-1656

Zhang, Xue; Fuehres, Hauke ; Gloor, Peter A,.(2011) Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear", Procedia - Social and Behavioral Sciences, Vol.26, pp.55-62

Zhang, Yini; Wells, Chris; Wang, Song; Rohe, Karl, (2018) "Attention and amplification in the hybrid media system: The composition and activity of Donald Trump's Twitter following during the 2016 presidential election, New Media & Society, September 2018, Vol.20(9), pp.3161-3182.

Breakeven Point Analysis:
An Alternative Breakeven Point Analysis Paradigm

David P Echevarria, University of North Carolina Wilmington

Abstract

The objective of this study is to define a process by which we may determine the operating breakeven points and the magnitude of sales that maximizes operating profitability. A new paradigm assumes a straight-line total revenue function and a curvilinear total cost function. Total costs are modelled as a quadratic function of sales relative to the breakeven point estimate. Analysis of a small set of industrial companies suggests that sample firms perform below the optimal level of sales. Moreover, data indicates that implied capacity utilization is inversely related to operating profitability.

Introduction

The focus of this inquiry is to examine the operating profitability of firms over time with respect to the behavior of operating profit margins as firms operate above their breakeven points. These two measures represent in large measure the operating aspects of the production function of the firm. Firms come to the marketplace with three competencies or factors; their production function, product mix, and marketing strategy. The first factor, the production function, is key to the firm's operating profitability. The production function reflects the mix of labor and capital which the firm uses to produce its goods and/or services at the lowest possible cost consistent with the desired level of quality. The focus of this inquiry is not in the traditional Cobb-Douglas production function that seeks to determine the magnitudes of labor and capital to achieve some desired level of output. The principal objective of this inquiry is to develop a system of analysis to determine with reasonable accuracy the optimal level of output/sales which maximizes operating profitability as measured by operating profit margins. The secondary objective is to determine the locus of the upper breakeven point; the level of output at which the increase in total costs relative to total revenues (sales) results in zero operating profits. Firms' managements are motivated to maximize profits. This requires the firm's managers to have some idea of the level of sales (output) that will maximize their operating profits. Experience suggests that the level of sales that maximizes the firm's operating profits is an elusive objective. The observed behavior is that the managers set sales goals in expectation of greater levels of profitability irrespective of whether the new sales targets are optimal. If managements had a better process for determining the level of output that maximizes operating profitability, the ought to be able to make better decisions relative to scale of plant and the magnitude of optimal production.

Issues with Standard Production Function Analytical Models

Rasmussen reminds us that there is no given mathematical functional form for a production function. "All the functional forms that have been used to describe the production have historically been based on more or less subjective choices. The best known of these function forms is the so-called Cobb-Douglas…" (Rasmussen, p17) The Cobb-Douglas production function has two variable inputs; labor and capital.

Issues with the Cobb-Douglas Production Function

The Cobb-Douglas relationship of output to the mix of labor and capital is well-established. Equation (1) represents the relationship.

$$Q = AL^{\alpha}K^{\beta} \tag{1}$$

Where:

Q = Total Output (in units or dollar value equivalents)
L = Labor input (person-hours or the total cost of labor inputs)
K = Capital (the value of all plant and equipment)
A = Total Factor Productivity
α and β are the output elasticities of Labor and capital, respectively.

If $\alpha + \beta = 1$ then constant returns to scale are observed. If $\alpha + \beta < 1$then decreasing returns to scale are observed. Finally, if $\alpha + \beta > 1$, then increasing returns to scale are observed. Returns to scale are important in that it tells us something about the nature of the firm's production function. In constant returns to scale, an increase of 10% in inputs should result in a 10%

increase in outputs. In decreasing returns to scale, a 10% in inputs will result in less than 10% increase in outputs. Accordingly, a 10% increase in inputs resulting in greater than 10% increase in outputs describes increasing returns to scale. It is possible that all three situations may be observed, although not simultaneously over time, as output goes from the breakeven point through that point of maximum profitability and onto the upper breakeven point. From the lower breakeven point to that point of maximum profitability the firm's production function can be characterized by increasing returns to scale. At that point of maximum profitability output, the production function should exhibit constant returns to scale. Passing the point of maximum profitability, the firm's production function should be characterized by decreasing returns to scale.

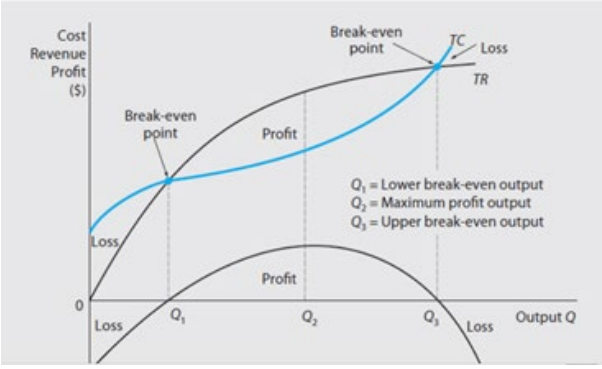
A significant amount of research suggests that the elasticities of labor and capital may not equal 1.00. A recent working paper by Gechert, et alia (2019) examines 121 studies. They find that the mean elasticity reported in the literature approaches 0.9. After correcting for various biases, they find that the elasticity approaches 0.5 and may run as low as 0.3. Moreover, their study notes that estimates are very sensitive to the level of aggregation; i.e., sector-level, state-level, or country-level. They define elasticity as "as a percentage change of the ratio of two production factors divided by the percentage change of the ratio of their marginal products." (See their Equation (1) on page 5.) The research procedure that follows does not aggregate by sector. While it is easy to assume that companies producing the same goods (for example: petroleum refining, pharmaceuticals) may not have identical production functions especially where trade secrets may be at work.

Identifying the labor input poses problems: It is not readily known how many of the employees of a company are involved as direct labor and how many are involved in the activities of selling, general support, and administrative functions. Moreover, we do not know the wage rate for direct labor. The second issue involves capital. At the lowest level consider the investment in plant and equipment, exclusive of the property (land), in terms of invested capital. The issue at hand regards the use of gross or net plant and equipment. While it is true that these assets are depreciated on the schedules approved by the Internal Revenue Service, what is not true necessarily is that the depreciation of plant and equipment do not necessarily suggest that they cannot maintain production rates for a significant period. A related issue addressed in Gechert, et alia, is the "rental price" of capital, assuming they refer to the cost of capital. It may be safely assumed that firms in the same sector may face different costs of capital as driven by their capital structure and investor perceptions of firm-specific risk. Aside front the issues addressed in Gechert, et alia, it remains that the principal issue with the Cobb Douglas production function is that it cannot tell us very much about the breakeven points and much less about the level of inputs at which the firm maximizes its profitability.

Issues with Breakeven Point Analysis

Breakeven Point (BEP) analysis defines that level of sales where operating profits are zero. BEP's are calculated using income statement data. The BEP is quotient of the firm's *fixed* overhead expenses (Selling, General, and Administrative expenses [SGA]) divided by the gross profit margin (GPM). An implicit assumption is made that the gross profit margin and total overhead expenses (SGA) remain the same from the BEP to where the firm is currently operating. While it is highly probable that SGA expenses may remain fixed in the short run, the same cannot be said for gross profit margins as output proceeds from the BEP level to a level consistent with expected sales. Consider the typical textbook representation of the total revenue and total cost schedules in Exhibit 1:

Exhibit 1: Managerial Economics BEP Model (McGuigan, Et al, p. 318)



Analysis of Exhibit 1 implies increasing returns to scale for operating income (and operating profit margins [OPM]) as output increases from Q1 (BEP) to Q2 (optimum level of output that maximizes OPM's). At output level Q2, the firm experiences constant returns to scale. The slope of the operating profit line is zero. As output increases above Q2 to Q3, the firm experiences decreasing returns to scale (OPM's decline). Operations above Q3 results in negative operating profits (OPM < zero). Less obvious are the implications of a curvilinear Total Revenue (TR) function. The curvilinear characterization

suggests that selling prices must vary from Q1 to Q3. Empirical observation suggests that prices (manufacturer's list prices) remain stable while output rates may vary from month to month or quarter to quarter. The total revenue function is the reflection of Price times Quantity where price remains constant, at least in the short run. The P times Q characterization yields a straight-line TR function. The same is cannot be true for the total cost function. Even if SGA (fixed expenses) remain fixed in the short run, the same is cannot be true for variable expenses. Variable expenses (Labor and Materials) must exhibit increasing returns to scale from Q1 to Q2. Workers are more productive in units of output as the firm approaches Q2.

An analogy to automobiles is relevant. Automobiles of recent manufacture are designed to achieve their best performance, as measured by miles per gallon, at say 65 miles per hour. Operating at speeds below or above 65 mph result in lower miles per gallon. Production operations are designed in similar fashion, especially those that can operate at slightly varying rates of output depending on the rate of expected demand without a meaningful change in the number of workers. Consider the additional per unit labor costs when employees work overtime. The Fair Labor Standards Act (FLSA, 1938, amended 2014) requires that non-exempt employees (hourly workers) be paid time and a half when they work more than 40 hours in a given pay week. The usual practice is to start time and a half when work exceeds 8 hours per day. In addition, hourly pay premiums are typical for second and third shifts. A firm experiencing increased demand may require overtime work or the addition of a second or even a third shift to meet demand. Assuming per unit labor costs will vary, the cost function is modeled as a polynomial function of order 2. The polynomial of order 2 will yield two solutions; the lower and the upper breakeven points. Setting the first derivative to zero fixes the optimal level of output at which operating profits are maximized. In micro-economic terms, the slope is zero (constant returns to scale) or the point at which we observe the maximum operating profits.

Modeling the Alternative Breakeven Point Analysis

The traditional methods for analyzing the efficiency of a firm's operations focus primarily on its operating cost structure. The focus on operating cost structure necessarily limits the inquiry to measures of gross and operating profit margins, which are the building blocks of traditional breakeven point analysis.

There are two ratios to measure the efficiency of production operations. The first ratio is *gross profit margin* (GPM). The gross profit margin measures how much of each sales dollar remains after covering the cost of labor and materials. Accounting refers to these expenses as the cost of goods sold (COGS). From a financial analysis perspective, we call the cost of goods sold *direct operating* expenses.

The second ratio is *operating profit margin* (OPM). The operating profit margin measures how much of each sales dollar remains after covering the cost of goods and the expenses incurred in the administrative support functions, selling, and marketing activities, and any of the other general (SGA) expenses incurred to support the production effort. The financial analyst refers to SGA expenses as *indirect operating* expenses. The economist would refer to these as fixed overhead expenses, at least in the short run. The OPM becomes the variable of interest since it captures both elements of total costs; Variable costs (COGS) and the overhead costs (SGA) or "fixed costs".

There are two related items that can be readily identified at each reported level of the firm's output in terms of sales. The level of current sales can be represented as a multiple of the breakeven point (BEPM). In addition, the firm's operating profit margin is directly related to its distance from the breakeven point. The relationship is assumed to be non-linear as we must allow for two breakeven points. The new BEP analysis paradigm is modelled as a nonlinear relationship and is represented as a quadratic equation shown below in Equation (2):

$$OPM_{i,t} = a_i\, BEPM_{i,t}^2 + b_i\, BEPM_{i,t} + C_i \qquad (2)$$

Where: $OPM_{i,\,t}$ = the OPM for company i in year t
$BEPM_{i,t}$ and $BEPM_{i,t}^2$ = breakeven point multiplier for company i in time t
a_i and b_i = regression estimates for company i
C_i = regression constant for company i

The quadratic model is relevant because it has two roots. The task remaining is to determine three points: The values of BEPM that make OPM equal to zero and the value of BEPM that maximizes operating profits or point Q_2 in Figure A above. Point Q2 is determined by setting the first derivative, f '(x), to zero.

Data Sample and Research Methodology

The typical data sample for many research inquiries uses large numbers of observations over time for large numbers of subjects. This allows for the generation of means, variances, and regression coefficients that describe the common

characteristics across the sample. The problem with using this approach is that like fingerprints, production functions are essentially unique to each company and may vary slightly within an industrial classification. Accordingly, the data sample for this inquiry uses a small number of companies (4 to 6) in the same standard industrial classification (SIC) with 20 years of operating data. The principal cross-sample common scale-free variables are the BEPM and OPM measures.

Yearly operating profit margins and BEP multiples are calculated for each company. Results are grouped by the firm's primary SIC. There are 4 industries in the sample data set: SIC 2834 – Pharmaceutical Preparations (5), SIC 2911 – Petroleum Refining (5), SIC 3312 – Steel Coke Ovens (6 domestic, 4 Foreign), and SIC 3674 – Semiconductor Manufacturers (4). The data covers the 20-year period from 1998 to 2017. These industries where chosen because they tend to be high volume output operations capable of varying their output rates in response to expected changes in demand.

Results

Three very relevant observations are made regarding the results of this inquiry. The first and most important is the sign of the $BEPM^2$ regression coefficient (a). If the sign is negative, then the expected production function can be modelled: We observe a lower and upper breakeven point as well as the output level that maximizes profitability. If, however, the sign is positive, the result is an inverted parabola that does not cross the zero-profit line. There is no suitable explanation for this observation. The best example of this was Nucor Steel. One potential explanation is in the impact of 2 or 3 outliers. Splitting the data sample occasionally yields a 4- to 10-year periods when that term has a negative sign.

The second observation pertains to the regression implied magnitude of the lower breakeven point as multiple of sales: BEPM(L). The *a priori* expectation is that BEPM(L) should be equal to 1. This is the value implied by the sales breakeven point formula: BEP = SGA / GPM. Twenty-three mean values of the BEPM(L) were computed. They range in magnitude from -3.695 (Chevron, 2001-2008) to +3.9186 (NUE, for the period 2011-2017). Nine (9) of the BEPM(L) estimates fall between 0.9000 and 1.1000. The mean BEPM(L) across the combined SIC groups was 0.8224. These observations suggest that the standard measure of the breakeven point as described in the literature may be incorrect. Curiously the most extreme outlier incidence occurs in just two companies; Chevron in both sub-periods, and Nucor in both sub-periods.

The third observation is the wide dispersion of implied average capacity utilization. Assume that the regression implied point of maximum profitability [$BEPM_{max}$] represents 100% efficiency. The average implied capacity utilization of the company is estimated by dividing the mean BEPM over the sample time-period by the $BEPM_{max}$. Sample estimates run from a low of 17.44% (MT) to a high of 107.33% (MUR, 1998-2007). The cross-sample average is 63.06%.

The term "implied capacity utilization" reflects the observation that these values are drawn from financial data. It is very possible that the utilization rates of any designed physical capacity may vary in much the same way as the velocity at which an automobile may be operated. What will change are the operating costs which are themselves very sensitive to the level of output activity as well as the impact of inflation and effects of supply and demand on the prices of factor inputs.

All the results are contained in Table 1, Panels A and B. The sample companies are grouped by their primary SIC designation. The grouping by SIC reflects the fact that the production functions of oil and gas exploration companies will be dramatically different from the production functions of steel producers and pharmaceutical companies. Several companies (Chevron, ExxonMobil, Murphy Oil, Carpenter Technology, and Nucor) appeared to have sub-periods of time where the dispersion of data points seem to suggest significant changes in their production functions.

The 4th column from the right displays the unadjusted r-square for the periods shown in column 3 (Data Years). The next three columns indicate if the regression coefficients were statistically significant at the 0.05 level of better: "ns" indicates not statistically different from zero.

Appendix A contains graphs of a several companies. The graphs include correlation results. The BEPM variable was treated as the causal variable and the OPM the effect variable. BEPM and OPM variables exhibit concomitant variation. There is no intent to suggest one causes the other.

Conclusions

This study finds that a new paradigm is available to determine the lower and upper breakeven points as well as an estimate of the level of sales/output that maximizes operating profitability. It also strongly suggests that the current breakeven point calculation model may not accurately describe the (lower) breakeven point of sales resulting in zero operating profits. The most intriguing observation is that most firms operate, on average, at levels below the optimal levels of sales/output that would maximize operating profitability. The preceding observation is accompanied by the observation that firms utilize less than 100% of implied capacity. This might suggest that firms are expanding capacity ahead of sales/output expansion and may be the most critical factor contributing to firms operating below optimal levels. These observations suggest further study is required in order to understand why these conditions seem to persist of long periods of time.

An infrequent observation of extreme outliers affecting regression results that preclude the graphing of lower and upper breakeven points as well as the optimal level of output expressed as a multiple of the breakeven point in sales must be further investigated. The two industries most affected by this phenomenon or SIC 2911 (Petroleum) and SIC 3311 (Steel Coke Ovens). 20-year period volatility is observed and maybe one reason for the existence of several significant outliers. The steel industry on the other hand, has been subjected to severe external shocks in the form of foreign steel competition which has affected their ability to generate profits. The possibility exists that there may have been some creative accounting or incorrect estimations of true production costs which result in the wide dispersion of observed operating profit margins in the concomitant sales levels expressed as a multiple of the breakeven point.

The alternative production function paradigm developed in this study will provide the analyst with a much-improved estimate of how operating profit margins will be affected by an increase or decrease in sales as reported by management in their operations review sessions with the analysts. It will also alert the analyst to be watchful of the presence of outliers and the implication for reported profits.

Further research is indicated in the attempt to explain why the relationship between operating profit margins and breakeven point multiples result in inverted parabola in which operating profits continue indefinitely in either direction from the perigee as implied by the regression coefficients.

Acknowledgement

I want to thank the many suggestions and comments from participants, especially those of Chad Kwon (UTRGV), at the 2019 Academy of Economics and Finance as well as the anonymous reviewer. Any remaining errors are the author's.

References

Gechert, Sebastian; Havranek, Tomas; Irsova, Zuzana; Kolcunova, Dominika (2019): Death to the Cobb-Douglas Production Function? A Quantitative Survey of the Capital-Labor Substitution Elasticity, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg.

McGuigan, James R., R. Charles Moyer, Frederick H.deB. Harris. "Managerial Economics: Applications, Strategies and Tactics." 13th Ed. Figure 9.5 Generalized Break-Even Analysis, Pg. 318.

Rasmussen, Svend. "Production Economics: The Basic Theory of Production Optimization." 2nd Edition. 2013. Springer, Heidelberg. ISBN 978-3-642-30199-5.

Table 1 Panel A: Production Function Analysis: Implied Lower and Upper Breakeven Points, Optimal Profitability

Company Name	SIC	Data Years	Mean GPM	Mean OPM	Mean BEPM	BEPM(L)	BEPM(U)	BEPM Max	Max OPM	Average Utilization	R-sqr	Regr. Coeff. Significance ax2	bx	c
Abbott Laboratories	2834	1998-2017	62.92%	27.88%	1.849	1.201	2.766	1.983	30.35%	93.24%	0.859	0.01	0.01	0.01
Bristol Myers	2834	1998-2017	75.73%	28.42%	1.620	0.832	5.670	3.250	52.60%	49.85%	0.934	ns	ns	ns
Johnson & Johnson	2834	1998-2017	74.99%	30.61%	1.694	0.968	3.575	2.271	38.28%	74.57%	0.977	0.01	0.05	0.05
Merck (Excl 2003-04)	2834	1998-2017	71.07%	32.38%	1.903	1.227	2.499	1.863	37.42%	102.13%	0.952	0.001	0.001	0.001
Pfizer	2834	1998-2017	86.37%	36.69%	1.866	0.796	4.622	2.710	49.70%	68.85%	0.901	ns	0.0564	ns
SIC 2834 Averages			74.22%	31.20%	1.7862	1.005	3.826	2.415	41.67%	77.73%				
Chevron	2911	2001-2008	19.71%	15.87%	5.450	-3.695	23.896	10.100	18.12%	53.96%	0.559	ns	ns	ns
Chevron	2911	2009-2017	21.99%	18.16%	6.343	-2.419	31.021	14.301	23.95%	44.35%	0.964	ns	0.05	0.05
Exxon-Mobil	2911	1998-2007	22.43%	15.88%	3.688	-0.047	13.346	6.649	20.44%	55.47%	0.964	ns	ns	0.05
Exxon-Mobil	2911	2009-2017	19.14%	14.30%	4.117	0.385	9.823	5.104	15.44%	80.67%	0.750	ns	ns	ns
Murphy Oil	2911	1998-2017	26.04%	18.89%	4.758	1.541	9.090	5.315	26.10%	89.50%	0.443	0.01	0.01	ns
Murphy Oil	2911	1998-2007	18.54%	13.90%	4.330	0.870	7.199	4.035	15.47%	107.33%	0.366	ns	ns	ns
Murphy Oil	2911	2008-2012	13.81%	11.72%	6.742	0.476	19.056	9.766	13.37%	69.03%	0.929	ns	ns	ns
Murphy Oil	2911	2013-2017	53.26%	36.04%	3.628	0.956	13.191	7.073	67.36%	51.29%	0.999	0.01	0.02	0.01
Valero Energy	2911	1998-2017	6.57%	5.31%	6.403	-0.384	20.283	9.950	6.58%	64.36%	0.335	ns	ns	ns
Holly-Frontier	2911	1998-2007	11.49%	9.12%	5.258	1.455	11.744	6.599	11.24%	79.68%	0.590			
SIC 2911 Averages			21.30%	15.92%	5.07176	-0.086	15.865	7.889	21.81%	69.56%				
Analog Devices	3674	1998-2007	65.74%	31.88%	1.958	0.980	5.192	3.086	45.41%	63.46%	0.930	ns	0.05	ns
Intel	3674	1998-2007	71.37%	40.82%	2.369	Nuc	4.237	2.744	45.37%	86.34%	0.765	0.01	0.01	0.01
Micron Technology	3674	1998-2007	45.34%	25.24%	2.512	0.858	9.732	5.295	45.68%	47.44%	0.934	0.01	0.01	0.01
Texas Instruments	3674	1998-2007	59.32%	32.17%	2.231	0.780	8.727	4.753	54.81%	46.94%	0.983	0.01	0.01	0.01
Qualcomm	3674	1998-2007	65.74%	34.85%	2.143	1.288	3.962	2.625	42.22%	81.63%	0.684	0.02	0.01	0.02
SIC 3674 Averages			61.50%	32.99%	2.24256	0.976	6.370	3.700	46.70%	65.16%				

Table 1 Panel B: Production Function Analysis: Implied Lower and Upper Breakeven Points, Maximum Profitability

Company Name	SIC	Data Years	Mean GPM	Mean OPM	Mean BEPM	BEPM(L)	BEPM (U)	BEPM Max	Max OPM	Average Utilization	R-sqr	Regr. Coeff. Significance ax2	bx	c
AK Steel	3312	1998-2017	9.22%	4.73%	2.1537	1.050	13.011	7.031	16.23%	30.63%	0.922	0.02	0.01	0.01
Commercial Metals	3312	1998-2017	13.36%	5.63%	1.7486	0.993	5.464	3.225	10.32%	54.22%	0.890	ns	0.01	0.01
Carpenter Technology	3312	1998-2017	24.74%	14.22%	2.4783	0.543	10.045	5.290	22.61%	46.85%	0.709	ns	ns	ns
Carpenter Technology	3312	1998-2007	27.85%	15.77%	2.4836	0.918	6.137	3.530	20.98%	70.36%	0.826	ns	0.05	ns
Carpenter Technology	3312	2008-2017	21.64%	12.68%	2.4731	0.901	9.744	5.320	22.25%	46.49%	0.968	ns	0.01	0.05
Nucor	3312	1998-2017	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM	NM
Nucor	3312	1998-2008	19.44%	15.83%	5.35	2.246	16.629	9.440	23.99%	56.67%	0.880	ns	ns	ns
Nucor	3312	2011-2017	13.38%	10.60%	4.8218	3.919	5.377	4.648	13.19%	103.74%	0.594	ns	ns	ns
Steel Dynamics	3312	1998-2017	20.95%	15.65%	4.0601	1.163	18.674	9.920	29.46%	40.93%	0.680	ns	ns	ns
US Steel	3312	2002-2017	10.26%	6.28%	2.821	1.059	14.177	7.618	15.64%	37.03%	0.934	0.01	0.01	0.01
SIC 3312 Averages			17.87%	11.27%	3.15446667	1.421	11.029	6.225	19.41%	54.10%				
Gerdau S/A (Brazil)	3312	2003-2016	23.74%	16.48%	3.2173	1.999	7.356	4.677	25.46%	68.79%	0.916	0.05	0.01	0.01
AcelorMittal (Luxem.)	3312	1998-2017	16.42%	11.69%	3.7914	1.044	42.440	21.740	47.08%	17.44%	0.921	ns	0.01	0.05
Posco (S. Korea)	3312	1998-2017	24.17%	19.00%	4.4086	1.095	24.399	12.747	40.11%	34.59%	0.939	ns	0.01	ns
Co. Siderurgica Nacional (Brz)	3312	1998-2017	46.95%	31.41%	3.6426	0.891	10.628	5.760	45.63%	63.24%	0.936	0.01	0.01	0.01
SIC 3312Averages			27.82%	19.65%	3.764975	1.257	21.206	11.231	39.57%	46.01%				
Cross Sample Averages			35.26%	20.00%	3.4640	0.809	11.931	6.376	30.09%	63.06%				

NM = Not Meaningful, ns = not statistically significant at the 5%

Appendix A

Panel A: Case of the 6 Domestic Pharmaceutical Preparations Firms SIC 2834

Table 2A contains the averages for Capacity Utilization, Max [Profit] BEPM, and Maximum Profit at optimum sales (output) levels. Table 2B displays the expected positive correlation between $BEPM_{max}$ and the maximum Operating Profit Margin (OPM_{max}). The negative correlation between the average implied capacity utilization and OPMmax was an unexpected result. The a priori expectation should be that increases in capacity utilization should be positive correlated with maximizing OPM.

Table 2A: Pharmaceutical Preparations

SIC 2834 Mean Operating Results			
Ticker'	Avg Util	$BEPM_{max}$	OPM_{max} %
ABT	91.02%	1.983	30.35%
JNJ	74.57%	2.271	38.28%
MRK	101.35%	1.886	38.32%
LLY	64.64%	2.487	47.19%
PFE	68.85%	2.710	49.70%
BMY	49.85%	3.251	52.59%

Table 2B: Pharmaceutical Preparations

SIC 2834 Correlation Matrix			
SIC2834	Avg Util	BEPMmax	OPMmax %
Avg Util	1		
BEPMmax	-0.9448	1	
OPMmax %	-0.8256	0.8856	1

Exhibit 2: Johnson & Johnson OPM vs. BEPM

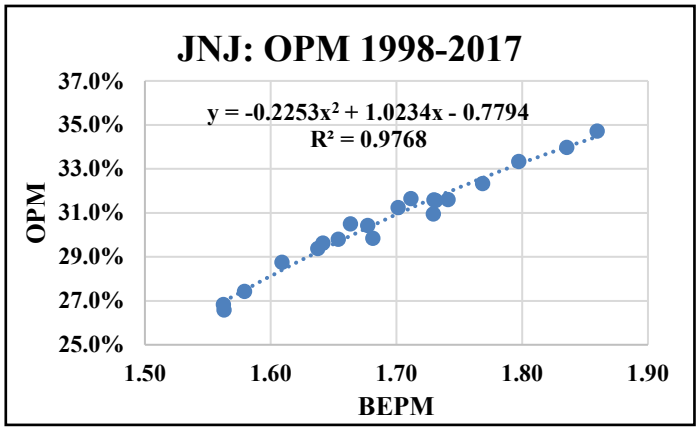
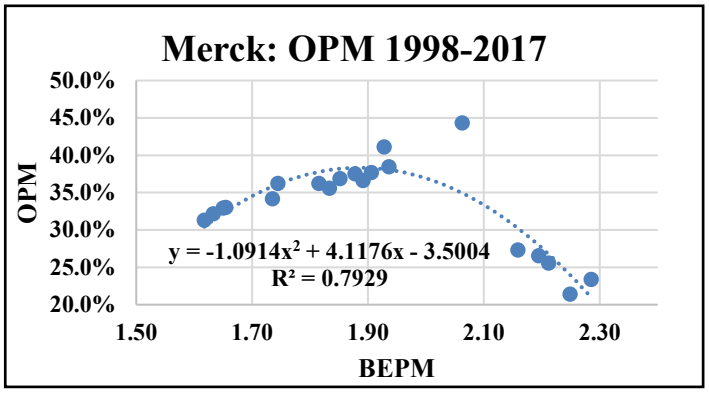


Exhibit 3: Merck: OPM vs. BEPM

Merck was the only SIC2834 firm to operate above 100% implied capacity utilization for 10 of 20-year sample.



Appendix A

Panel B: Case of the 6 Foreign Pharmaceutical Preparations SIC 2834

Note the near perfect positive correlations between $BEPM_{max}$ and OPM_{max}. The greater the BEPMmax value the greater the maximum Operating Profit Margin. The unexpected observation is the inverse correlation between the Average [Capacity] utilization and $BEPM_{max}$ and OPM_{max}. The reasonable expectations are for all three metrics to be positively correlated.

Table 3A contains the averages for Capacity Utilization, Max [Profit] BEPM, and Maximum Profit at optimum sales (output) levels. Table 3B displays the expected positive correlation between $BEPM_{max}$ and the maximum Operating Profit Margin (OPM_{max}).

Table 3A: Pharmaceutical Preparations

SIC 2834 (Foreign) Operating Results			
Ticker	Avg Util	BEPMmax	OPMmax %
NVS	82.66%	1.897	31.58%
TEVA*	95.85%	2.159	31.99%
GSK	83.38%	2.058	39.94%
RHHBY	76.68%	2.369	42.67%
SNY	45.12%	4.065	65.84%
AZN	39.72%	4.148	72.40%

* TEVA only foreign company to exceed 100% Capacity Utilization during sample period.

Table 3B: Pharmaceutical Preparations

SIC 2834 (Foreign) Correlation Matrix			
	Avg Util	BEPMmax	OPMmax %
Avg Util	1		
BEPMmax	-0.9586	1	
OPMmax %	-0.9780	0.9783	1

Exhibit 4: Roche Holing AG OPM vs. BEPM

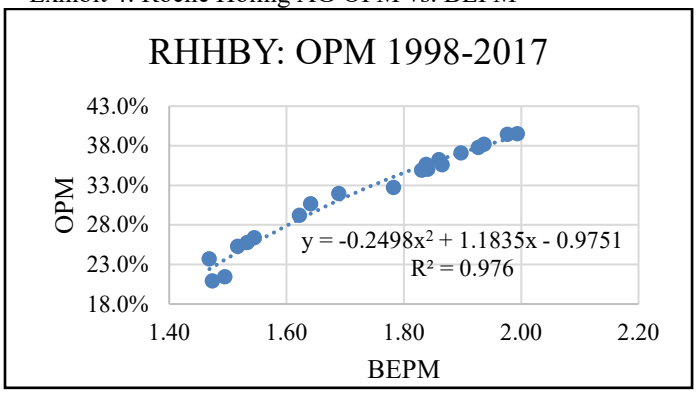
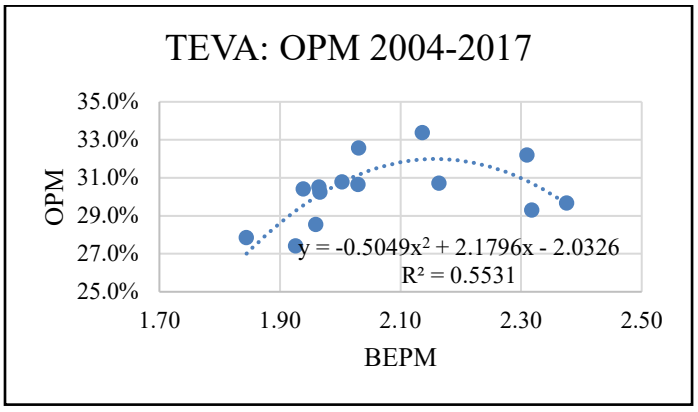


Exhibit 6: Teva Pharmaceutical Industries OPM vs BEPM



Do Microfinance Institutions Help with Business Growth? Evidence from Rural Nepal

Shankar Ghimire, Western Illinois University
David Clifford and Grace Costa, Maryville College

Abstract

This paper studies the role of microfinance institutions in promoting business growth in the rural parts of Nepal. The paper uses a combination of quantitative and qualitative data obtained through a survey of 129 local businesses. The empirical findings show a weak correlation between microfinance loans and increased business revenue. However, the case studies of select successful businesses indicates a near-unanimous public support for microfinance investment in the community. The implications of the results are that microfinance institutions have the opportunity to identify and enable business ventures to start, expand, and catalyze their growth in rural areas facing financial constraints.

Introduction

This paper presents the findings from a survey administered in rural Nepal to study the economic impact of micro loans to small businesses. Specifically, the paper analyzes the role of microfinance institutions in businesses growth in Ward No. 4 of Baiteshwor Rural Municipality, the then Kabre Village Development Committee ("Kabre" hereafter) of the Dolakha District. This is an especially relevant topic given the immense expansion of microfinance institutions (MFIs), savings and credit cooperatives in the area since the mid-2000s. Microfinance has been considered a tool for poverty alleviation in rural areas of developing countries, despite mixed results across the globe. Where the positive results are obtained, the source of the poverty reduction is not clear. To that end, we postulate that microfinance institutions provide the much needed financial capital for business growth. This would provide increased opportunities for businesses, create more jobs, raise incomes, and help reduce the number of people living in poverty. The examination of the relationship between microfinance institutions and the communities they serve in terms of their role in local business growth is the main contribution of this paper, unexplored hitherto.

This research evaluates the impact of financial development in rural economies specifically through the lens of business growth. While the impact analysis of microfinance is not a new topic, existing literature has focused on issues such as poverty reduction (Khandker, 2005; Morduch & Haley, 2002; Weiss & Montgomery, 2005; Quinones & Remenyi, 2014, etc.), women empowerment (Ashe, 2002; Cheston & Kuhn, 2002), employment generation (Ghimire et al., 2017), etc. How these micro loans in the rural community have helped (or not) the small shops and farms to grow and generate more revenue is missing in the literature and this paper fills that void. The study is contextualized at several levels from the perspective of rural development, particularly significant in a nation where 80% of the population lives in rural communities (Gautam and Anderson, 2016).

The data collection took place in Kabre, a rural village in Nepal. This location is suitable to study the issue especially due to a significant presence of MFIs in the region and an overwhelmingly high rate of participation of the local population and businesses in the microfinance programs. The study was structured to capture the demographic information of the local businesses and their participation levels in microcredit needed to conduct statistical analysis on business performance. The methodological approach included questionnaires, interviews, and case studies. These approaches provide insight into both borrower and lender perspectives with respect to business growth due to micro loans used to finance a business. Kabre village is also ideal for observing this dynamics because of the diversity of businesses and substantial number of established microfinance institutions in the vicinity.

In a broader context, this research contributes to the efforts made in developing countries to find economic strategies that encourage investment and improve the overall sense of economic opportunity for aspiring entrepreneurs – something well established in the literature (Beck, et al., 2015; Lindvert, et al., 2015; Igwe et al., 2018). The study in Kabre fits into that context by representing the voices and real-life situations of rural business owners in Nepal. This region of Nepal was impacted by the blockade of the southern Nepal border by India—by far Nepal's largest trading partner, as well as the enormous earthquake in 2015 that devastated the Dolakha region and inflicted large scale damage and destruction to the region's building's and infrastructure. Both of these factors had negative impacts on the level of business growth in the region, evidenced by business owners making reference to them during our survey. The presence of microfinance institutions came as a relief for many. Their struggles and successes provide an example that can be replicated and improved upon by other rural communities in developing world. The presence of microfinance is important for rural development because the ease of access to microcredit can stimulate agricultural and market-based economies, which most rural areas are built upon.

Our findings show that microfinance institutions have a positive role in business growth but the magnitude of this relationship varies. Certainly, it seems that factors such as education, year of establishment, and particularly if they had received a loan from a microfinance institution, were relevant determinants towards the respondent having a generally positive attitude towards the microfinance industry. However, the responses tested are not directly significant to the loan's ability to help expand business, but more geared towards the local businesses views towards the loans themselves. We can clearly see that the loan recipients had the most positive opinion of the microfinance industry, with education level also proving to be a significant indicator of a positive opinion towards MFIs. Whether the loans truly created a positive impact on the businesses surveyed is difficult to determine from the above data alone, for various reasons explained later in the paper.

The remainder of this paper is structured in the following order: a review of literature specifically focused on the role, impact, and effectiveness of microfinance in Nepal is presented in the next section, followed by a description of data and methodology, empirical analysis and the discussion of results, and conclusions.

Literature Review

Formal research on microcredit is now decades old and there is a significant body of work that documents the progress as well as shortcomings on the impact of microcredit in impoverished areas where poor borrowers have had substantial barriers to accessing credit. With time, the effects of microcredit have become substantial enough to analyze, and questions have arisen of whether microcredit is beneficial. Looking at factors such as investment spending, employment generation, disposable income, growth of savings, women's empowerment, quality of life, and business development, researchers across the globe have sought to explain the system of microcredit in the context of rural households, businesses, and families.

Banerjee (2013) overviews studies that have been conducted since microcredit research began twenty years ago. Banerjee explains the complexities of microcredit as it adapts to meet the needs of specific community dynamics. One of the highlighted practice around the world is group lending. Group lending as a mechanism of microfinance was initially seen as innovative way to maintain low default rates due to the joint-liability of every member in the group. In this manner, women's group-based lending in Nepal was established in order to meet both the business development objectives of microfinance and empower women seeking credit to improve their autonomy in the domestic and public sector (Cheston & Kuhn, 2002). Similar study is presented in Ashe (2002) and Akoi & Pradhan (2013). Overall, group-lending programs can help borrowers within the group feel more secure in credit consumption and be more likely to make investments with the confidence of the rest of the group and thus help promote business expansion. Nepal's specialized microfinance institutions include women's group-based microcredit programs, savings and credits, and microfinance accommodated to the needs of farmers.

In the case of the women's group lending in the Terai region of Nepal (the southern plains), Sharma (2007) examines the correlation between participation in group lending and the empowerment of women. The contribution of microfinance on women's social and financial independence was apparent in household decision making, access to economic resources, growth of social networks, and freedom of mobility. The paper comes to this conclusion through field surveys, group discussions with credit groups, and thorough interviews with the MFI administrators. The sampling process was based on criteria for the MFIs and distinguishable characteristics that could be used to categorize the MFIs. On the individual level, survey questionnaires and group discussions with borrowers facilitated the researcher's understanding of household factors that could be used to determine independent decision making in various household characteristics.

This paper uses similar analysis methods presented in Sharma (2007) which included determining what factors affected the relationship between MFI loans and business growth. Our research included questionnaires and one-on-one interviews with established microfinance institutions in the village and these questions were oriented to find specific indicators of successful investments in business rather than address specific social consequences.

In Jeffrey Ashe's 2001 analysis of the Pact's Women's Empowerment Program (WEP), women's empowerment was again measurably improved via the microfinance industry. Ashe finds that WEP operated through 6,500 individual groups in Nepal and mobilized $2,000,000 in assets in under 3 years. By many metrics, it was found that for women in the program experienced an increase in empowerment: primarily with increased decision making roles in family planning, children's marriage, buying and selling property, and sending their daughters to school (Ashe, 2002; Cheston & Kuhn, 2002).

In undertaking the task of researching a fluctuating indicator such as business expansion, researchers must be careful to consider which variables and analytical methods might overstate the benefits of microfinance, due to correlated unobservable factors. Rajbanshi et al. (2014) use a dataset from eastern Nepal to study the effect of microfinance on first-time borrowers against previously recorded experimental studies to determine the possible discrepancies between the two due to falsely correlated variables. The study refers to two distinguishable factors as responsible for most discrepancies in experimental studies on microfinance. The first factor is the timing with which a borrower takes a microfinance loan. The second factor is the experimental error of documenting the effects of microfinance after a population has initially been exposed to credit institutions, thus underestimating the most substantial impact of microfinance borrowing observable only in the initial stages.

This issue is subdued in our method because we have a census like sample which has both borrowers and non-borrowers before and after the event took place.

The second factor that can influence data recording on microfinance borrowing is whether the subjects took loans when microfinance first became available or at a later time. The authors detail the significant discrepancy that can occur if the data is taken from the latter group because these later borrowers are less likely to have a threshold of optimistic economic opportunity than the first round of borrowers. Thus, the second group of borrowers will not realize the same level of success as the first group because the second group's evaluation of economic opportunity is less. Economic opportunity is what motivates individuals to invest in their well-being and business, so having less motivation can be correlated to less productivity. Our study takes this concept into account by including those individuals who claimed to take microfinance loans as primary sources of capital for starting a business. Those who decided to start a business based on other forms of capital and take loans later on were not counted as the treated population because the impact would not be the same. This argument is consistent with the literature that highlight the impact evaluation methods (Copestake, et al., 2001; Amendariz & Morduch, 2005; Banerjee et al., 2015).

Data and Methodology

The survey was administered to collect information applicable for both quantitative and qualitative analysis: qualitative in the interviews and case studies of both representative businesses that had benefited greatly from microfinance loans, and microfinance institutions operating in the study area. The survey compiled information from 129 individual respondents of local business operators. The survey collected a combination of demographic data, business performance data, and the respondents' relationship with the microfinance industry—whether they had taken a microfinance loan or not. The survey instrument is included in Appendix A. This data was then contextualized, analyzed for descriptive statistics from the respondents, and run through a series of statistical analyses to determine the significance of the findings and their relation to the business expansion measured by the difference in the revenue of the business between their starting period and the current period. Our analysis closely follows Ghimire et al. (2017) who use the same dataset to analyze the impact of microfinance on employment generation within the same sample.

Survey Methodology

As shown in Appendix A, the questionnaire was divided into six sections that targeted areas such as business type and establishment, personal information, microfinance awareness, sources of capital, and business performance. The questions were printed in English and the answers recorded by researchers. A Nepali translator facilitated the questionnaire process by reading and explaining the research objectives. Once giving consent, the respondents proceeded to answer the questionnaires verbally, which were translated and then documented by the researchers. These responses were used to statistically analyze the data with respect to the hypothesis that respondents who take microfinance loans to start their businesses see a greater improvement in business growth than respondents who do not take a loan.

In order to analyze business growth from the perspective of microfinance institutions, we administered similar questions to credit cooperative administrators in the form of interviews. In addition to opinion-based questions about determining whether microcredit improves business performance, the interview questions focused on loan default rates and growth of microfinance members since establishment. The interviews were also translated and recorded on video for later reference. Analysis of the perspectives of the credit cooperatives contributes to understanding the relationship between borrowers and lenders as both seek increases in overall economic opportunity.

While the questionnaires and interviews investigated variables that might indicate general trends in overall business expansion, there were four case studies conducted to study the individual impact of microfinance loans on a personal scale. The case studies were recorded on video and used the same approach as the questionnaires. One researcher administered the questions, which were translated for the respondent. Both the general survey questions and personal, individualized questions were asked in a sequence to create well-rounded stories of business origin and growth since taking a loan. These case studies contributed to the study of what variables may or may not affect uninterrupted business growth when taking a loan as start-up capital. Their success indicates that microfinance certainly has the capacity to inspire both business growth and employment generation. The respondents explained their challenges and benefits as they experienced them over the term of borrowing credit.

Descriptive Statistics

The survey captured a wide array of data from the owner of the business, including their performance, as well as their interaction with the local microfinance institution. Out of the 129 respondents who owned a business, 53% were male and 47%

female; 60% of respondents were between 20-40 years old; an overwhelming 90% of the business owners were married; education levels varied widely, and correlated closely to their age. Older participants—particularly women—often had no formal education or only reached the level of 4th grade or below, while younger respondents typically had reached 10th grade or higher. Shops constituted 50% of businesses surveyed. These shops offered a range of goods, from food to electronics. Survey subjects were almost exclusively (99%) sole proprietors or their businesses, and what's more, 60% of the businesses contacted had begun in the last 15 years, aligning with the overall trend of rapid growth of microfinance in the area.

The survey participants were well aware of the local microfinance institutions, and had broad approval of their operation in their community. Eighty-eight percent of respondents felt that 'microfinance institutions' encouraged business growth; 90% of respondents felt that they encouraged entrepreneurship, 77% that they generated more employment. Of the 129 respondents the research team surveyed, 70% reported that their business 'directly benefited' from a microfinance institution, in the form of credit and savings activities. When asked about how they were able to start their business, 86% replied that some personal savings were used, while 53% of the total businesses said they turned to microfinance for the remaining start-up capital. On average, the *initial* loan they took out to start their business was 316,000 Nepalese rupees (or, roughly $3,160 @ 1USD = Rs. 100 in 2015). The average repayment period was 3.5 years, and 74% of respondents reported that they had been able to repay the loan within the specified time, or were on track to.

Potential Concerns

While we followed a systematic method to collect and analyze the information to a great detail, a conclusion of whether or not the loans truly created a positive impact on the businesses surveyed have some caveats for multiple reasons. First, during the survey period the economy of Nepal was still affected from the massive earthquake that devastated the country in April 2015. During the surveys, many respondents answered that decreases in revenue since initial establishment were often due to both consumers and producers still recovering from the earthquake. In the wake of the natural disaster, disposable income and consumer investment spending fluctuated, which impacted the otherwise uninterrupted growth of businesses. Many Nepali people who suffered losses from the earthquake faced long delays in receiving relief from the government and the donor agencies who committed assistance in the form of foreign aid, so businesses who provided goods other than essentials were understandably at risk of lower revenue. In the months after the earthquake, Nepal experienced an economic blockade in the south, which is a main port of entry for goods from India. The government and the Nepali people were still recovering from this economic shock when we conducted our study. The blockade lasted from September 2015 until February 2016, only a few months prior to our study. Petroleum shortages sent the economy reeling as all sectors of production—agriculture, transport, manufacturing, tourism, to name a few—were suffering. The country could not even support international flights as the fuel for aircrafts was in such short supply. With these caveats, next, we present our main findings.

Empirical Analysis and Results

The research team embarked on this trip in the interest of studying microfinance in the rural developing world, with the research hypothesis that microfinance institutions have a positive impact in rural Dolakha in both business expansion and employment generation. To do this, we collected a combination of qualitative and quantitative data: qualitative in their case studies of specific businesses that benefited from an MFI loan, as well as local microfinance operating in Dolakha; and quantitative in the form of surveys of 129 local businesses, gathering basic demographic information, business performance, and relationship with the local microfinance industry. From this quantitative data, the team then ran a series of statistical analyses to determine if their findings were significant, and if their hypothesis was proven true. Four different statistical analysis tests are performed: correlation coefficient test, two sample t-test, probit regression for the likelihood of a business taking a microfinance loan for startup capital, and propensity score matching to find out if those businesses that took a loan had higher revenue differential between the initial phase of the business operation and the current period. The findings from these tests are listed and contextualized below.

Correlation Coefficient

The purpose of the correlation test is to see the association between two variables. In this case, the two variables tested are the reported revenue increases of MFI borrowers and non-borrowers. The correlation coefficient indicating the correlation between receiving a loan and increase in revenue was not statistically significant. The coefficient result of -0.078 demonstrates a *negative* correlation between borrowing and increased revenue—however this coefficient is small. Based on the correlation coefficient alone, we cannot conclude that microfinance loans had an impact on business expansion.

Two Sample T-Test

Our objective is to see if the average outcomes in terms of revenue are different for two groups of businesses: the loan borrowers and non-borrowers. The two sample t-test failed to reject the null hypothesis that the difference in revenue for the borrowers and non-borrowers is the same (diff=0, when diff= mean(0)-mean(1)). The two values were not statistically different . Borrowers reported an increase in monthly income of 91,000 NPR, when compared to the income reported by non-borrowers. This difference, however, was found to be statistically insignificant.

Table 1: Two Sample t-test with Equal Variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	61	0.2090	0.1826	1.4264	-0.1563	0.5743
1	68	0.5272	0.1323	1.0910	0.2631	0.7913
Combined	129	0.3767	0.1114	1.2658	0.1562	0.5972
diff		-0.3182	0.2223		-0.7581	0.1217
diff = mean (0) - mean (1)						t = -1.4314
Ho: diff = 0					degrees of freedom = 127	
Ha: diff < 0			Ha: diff != 0		Ha: diff > 0	
Pr(T < t) = 0.774			Pr (|T| > |t|) = 0.1548		Pr(T > t) = 0.9226	

Propensity Score Matching

Finally, the Propensity Score Matching (PSM) method as suggested in Caliendo & Kopeinig (2008) was used to analyze the impact of MFIs on those businesses that borrowed a loan from one of the microfinance institutions. The method tests the counterfactual – what would have happened to those businesses if they had not taken a loan? To answer the question, we needed to find a close match from the group that did not take a loan. The survey consists of 68 businesses in the treatment group (i.e. the businesses that borrowed a loan to start their business) and 60 businesses in the control group (i.e. the businesses that did not borrow a loan to start their business). Once we find the matching sample based on the propensity scores, we compare the growth in revenue for the two groups: microfinance borrowers and non-borrowers. The results are presented below:

Table 2: Propensity Score Matching Borrowers vs. Non-Borrowers

ATT estimation with the Radius Matching Method				
N. Treatment	N. Control	ATT	Std. Err	t-stat
68	60	0.397	0.24	1.649

Note: the numbers of treated and controls refer to actual matches within radius

The results show that the average treatment on the treated (ATT) was 0.397 more than the control group. This means, those businesses that borrowed a loan to supplement their startup capital, were able to expand their business by 39.7% more in terms of their revenue compared to those that did not borrow a loan. The result is statistically weakly significant at 10% significance level as shown by the t-statistic of 1.649.

In sum, while we see a tenuous correlation between taking microfinance loans, and the two sample t-test does not show a difference between the two groups, the propensity matching method shows that those who took the loan indeed outperformed the non-borrowers, albeit at a weaker statistical significance.

Perceptions towards Microfinance Institutions

To support our results from the propensity score matching methods, the case studies and the face-to-face interaction with MFI members seemed to indicate an overall positive association between loan beneficiaries and the growth of their businesses. To understand the perceptions of the local people, we asked individuals what they thought of the role of microfinances on business growth, entrepreneurship and employment generation. One of the questions on the survey asked if people thought that MFIs helped with business growth in the local areas. The respondent answered “yes” or “no” and the response was coded “1” or “0” respectively. We performed a probit regression where the dependent variable is the binary “1” or “0”. The explanatory

variables used in the model were the individual characteristics such as age, sex, education, caste, and the year that they business was established. From the regression results, we can see that the coefficient on "Borrowers" is positive with a p-value of 0.024. This means that the probability of the respondents reporting the business growth because of MFIs was higher if they took a loan from the local microfinance. This is a statistically significant result at less than 5% significance level. The probit regression model revealed that the opinions of credit cooperative loan beneficiaries were consistently positive towards the presence and work of microfinance in their area.

Table 3: Probit Regression Analysis (replication as in Ghimire et al., 2017).

	Business Growth Perception
Borrower Dummy (Borrowed == 1)	0.7944**
	(0.351)
Age	0.0035
	(0.015)
Education	0.08166*
	(0.0438)
Gender Dummy (Male==1)	0.0115
	(0.359)
Observations	129

Standard errors in parentheses. *$p<0.1$, **$p<0.05$, ***$p<0.01$

We also surveyed about their business performance. Business performance according to the surveyed owners was growing, if modestly. Sixty percent (60%) of respondents reported their revenue increasing from the time they started their business. Despite this, 60% of participants also reported their business performance as 'average', while 26% reported 'above average' performance, 3% 'excellent', and only 1% said their business was performing as 'below average', compared to their competitors. The survey data indicated that increased revenue was far more common than increased employment: only 28% of participants responded that employment increased between the start of the operation and the current employment. When analyzing the data as it pertains to business expansion, it seems that businesses certainly expanded, though what responsibility microfinance loans have for this growth is unclear.

Case Studies

We conducted four randomly selected case studies. Through this lens, the research team was able to observe anecdotal evidence of success stories utilizing microfinance loans. The research team observed a handful of different case studies, randomly chosen, and in each of the cases both business expansion and employment generation was observed. The cases were of two cardamom farmers, an herbal plant nursery, and a buffalo farm. Research team members individually interviewed the proprietors of the selected operations, and each one reported that they had used microfinance to start their operation, and had enjoyed a high level of success in both business performance and employment generation. These case studies demonstrated—at the very least—that microfinance has demonstrated a clear capacity to enable business growth and economic expansion.

Of the four, each had a remarkable story of success and business prosperity that began with a credit cooperative loan. The first category of case studies examined the thriving business of a group of cardamom farmers who each took out individual loans. Their farm was years in the making and growing with each harvest. One farmer agreed to allow the research team to view a portion of the farm and show just how much the business had progressed since its establishment. He had a positive outlook on the expansion of his farm in the future and felt like the credit cooperative services and accessibility were essential to his financial security.

One of the cardamom farmers had taken a loan from a credit cooperative to fund his cardamom nursery. After the initial two years it took to cultivate the plants, the nursery farmer reported that his business was stable and profitable enough to allow him to hire five more external hires. Though he complained about the interest rate of his loan, the nursery farmer expressed that he would not otherwise have had sufficient capital to start the nursery and was grateful for microfinance in that regard. The next cardamom farmer we interviewed had a similar experience with the microfinance in that she needed a larger pool of capital from which to invest in farm. The microcredit loan provided her with the funds she needed to make that investment and keep a competitive edge in the cardamom market.

Building off of the cardamom farmers' success in agricultural pursuits after taking a loan, the fourth case study covers a diversified herb farmer who took a loan to meet the demand for his goods. The herb famer began his business with the intention to grow pine trees with capability of treating cancer in some cases. His mother had passed away from cancer, which inspired

him to make contact with a local NGO for agricultural training and took out a micro loan to finance his farm. Though, he explained that there were still many improvements to be made on his business, the farmer's business relied on the flexibility and financial support of the credit cooperative to continue expanding.

The last case study was conducted on the case of a buffalo farmer who used his credit cooperative loan to turn his business into a profitable source of income and employment. After taking the initial loan and establishing the buffalo farm, the farmer was able to hire external employees to facilitate the expansion of his business.

Perspective of the Microfinance Institutions

From the perspective of the microfinance, there was a unanimous assertion that microfinance contributed to the development of business growth, employment, and entrepreneurship. For example, the women's group credit cooperative located in the village responded that their presence notably increased women's initiative to invest in business or establish one themselves. The microfinance institution thoroughly discussed their efforts to make credit as accessible and secure for their members as possible. To counterbalance the high interest rates of 13-14%, microfinance institutions offered savings accounts at rates that technically cut the interest in half over time. Between the success stories examined in the case studies and the optimism of microfinance, there is sufficient evidence that microfinance can be immensely beneficial to business owners and the community as a whole.

Conclusions

The main objective of this paper is to study if the microfinance loans in rural Nepal have a positive influence on rural business development, particularly in starting and growing these businesses. The research used a combination of both quantitative and qualitative methods. We surveyed 129 local businesses, which asked respondents for their personal and business information, as well as their relation to the microfinance industry. We also conducted multiple qualitative case study interviews with local microfinance institutions, as well as farmers who had benefited from the microfinance loans they were able to access through the aforementioned institutions. Our main findings show that those businesses that borrowed a loan to supplement their startup capital were able to expand their business by 39.7% more in terms of their revenue than those that did not borrow a loan. However, these results are only weakly significant.

The research team's qualitative interactions with four randomly selected entrepreneurs were illuminating in that they showed the potential effectiveness of the microfinance loans. In each of the four cases—two of cardamom farms, one of an herbal plant nursery, and the final one of a buffalo farm, were instances in which the businesses all used microfinance loans to start their business, and the results that ensued were remarkable; they had experienced both large business expansion as well as increased employment. The case studies each displayed the capacity for microfinance to be a positive agent for change.

We analyzed the survey data using various quantitative methods: t-tests, correlation analysis, and propensity score methods. While the data and statistical analysis across various methods were not consistent, there was significance in the answers to opinion-based survey questions. Loan beneficiaries overwhelmingly answered positively to questions that addressed whether respondents thought MFIs benefited business growth, employment generation, small-scale entrepreneurship, and their personal business. These responses show that there may at least be some accreditation to MFIs for boosting the overall sentiment among their members that microcredit has the potential to benefit business growth.

While the results are interesting, there are some caveats to the study. The inconsistency in the results could be due to multiple factors such as the earthquake and economic blockade that the country faced in 2015. From the individual perspectives of the business owners, there were consistent remarks about the duplication of businesses in close proximity, which raised the level of competition faced by both new entrepreneurs and established business owners. If the market is to diversify in the coming years, there might be more promising results for recording revenue increases. Otherwise, a market saturated by the same or similar businesses can fluctuate according to the changes in consumer preferences. Factors such as business expansion are difficult to document during times of instability or periods of rapid growth because the social and economic consequences can affect data in unpredictable ways.

In conclusion, despite the weakly significant empirical results, we found many successful businesses as documented in the case studies and the overall optimism of loan beneficiaries towards the growth of their business. For example, one case study described a farmer's road to success from self-employment to a employing external hires on a hugely successful cardamom nursery. He began the business by receiving a loan from an agriculture-focused credit cooperative that was able to accelerate the growth of his business when he otherwise would have lacked sufficient capital. These success stories gave insight into the possibility for business growth for Nepal's aspiring entrepreneurs. Future studies should focus on highlighting these individual cases.

Acknowledgement

This research was supported by ASIANetwork under the Student-Faculty Fellows (SFF) program funded by the Freeman Foundation. The authors gratefully appreciate the support. The authors are also thankful to two anonymous referees whose comments significantly improved the paper.

References

Aoki, C., & Pradhan, P. K. (2013). Impact of microfinance services on socio-economic empowerment of the women's group in Nepal. *Geography of Governance: Dynamics for Local Development*, *1*, 01-10.

Armendáriz, B., & Morduch, J. (2010). *The economics of microfinance*. MIT press.

Ashe, J. (2002). PACT's women's empowerment program in Nepal: A savings-and literacy-led alternative to financial building. *Journal of Microfinance/ESR Review*, *4*(2), 8.

Banerjee, A. V. (2013). Microcredit under the microscope: what have we learned in the past two decades, and what do we need to know? *Annual Review of Economics*, *5*(1), 487-519.

Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, *7*(1), 22-53.

Beck, T., Lu, L., & Yang, R. (2015). Finance and growth for microenterprises: evidence from rural China. *World Development*, *67*, 38-56.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31-72.

Cheston, S., & Kuhn, L. (2002). Empowering women through microfinance. *Draft, Opportunity International*, *64*.

Copestake, J., Bhalotra, S., & Johnson, S. (2001). Assessing the impact of microcredit: A Zambian case study. *Journal of Development Studies*, *37*(4), 81-100.

Ghimire, S., Rigatti, T., & Sexton, N. (2017). Effect of credit cooperatives in employment generation: Evidence from rural Nepal. *Journal of Development Innovations*, *1*(1), 29-44.

Igwe, P. A., Ogundana, A. N. A. O. M., Egere, O. M., & Anigbo, J. A. (2018). Factors affecting the investment climate, SMEs productivity and entrepreneurship in Nigeria. *European Journal of Sustainable Development*, *7*(1), 182-200.

Khandker, S. R. (2005). Microfinance and poverty: Evidence using panel data from Bangladesh. *The World Bank Economic Review*, *19*(2), 263-286.

Lindvert, M., Yazdanfar, D., & Boter, H. (2015). Perceptions of financial sources among women entrepreneurs in Tanzania. *African Journal of Economic and Management Studies*, *6*(2), 197-218.

Morduch, J., & Haley, B. (2002). Analysis of the effects of microfinance on poverty reduction. *New York: NYU Wagner Working Paper*, *1014*.

Quinones, B., & Remenyi, J. (2014). *Microfinance and poverty alleviation: Case studies from Asia and the Pacific*. Routledge.

Rajbanshi, R., Huang, M., & Wydick, B. (2015). Measuring microfinance: Assessing the conflict between practitioners and researchers with evidence from Nepal. *World Development*, *68*, 30-47.

Sharma, P. R. (2007). Micro-finance and women empowerment. *Journal of Nepalese Business Studies*, *4*(1), 16-27.

Weiss, J., & Montgomery, H. (2005). Great expectations: microfinance and poverty reduction in Asia and Latin America. *Oxford Development Studies*, *33*(3-4), 391-416.

Appendix A: Survey Questionnaire

A. Business Information

1. ID Number:
2. Name of business:
3. Year of establishment
4. Name of the respondent:
5. Ownership
    a. Employee
    b. Owner
6. Type of Business
    a. Sole proprietorship
    b. Partnership
    c. Group monitoring
    d. Other (please specify)
7. What good or service does the business provide?
    a. Shop
    b. Restaurant/Lodge
    c. Manufacturing
    d. Tailoring
    e. Agriculture
    f. Poultry Farming
    g. Animal Husbandry

B. Respondent Information:

1. Age:
2. Sex:
3. Education:
4. Marital Status:

C. Microfinance Awareness

1. Do you know about the presence of MFIs in your area?
    a. Yes. How many?
    b. No.
2. Do you think MFIs encourage business growth?
    a. Yes
    b. No
3. Do you think MFIs encourage entrepreneurship?
    a. Yes
    b. No
4. Do you think MFIs help generate more employment?
    a. Yes
    b. No
5. Does your business directly benefit from MFIs?
    a. Yes. How?
    b. No

D. Initial Business Financing

1. Select the sources of initial capital for your business (circle all that apply)
    a. Micro finance institutions / Cooperatives
    b. Other financial institutions (commercial banks)
    c. Remittance: How many people?
    d. Savings/Other source of income
    e. If none of the above state:
2. Select the amount of initial capital to your business (provide all that apply)
    a. Micro Finance Institutions.
        i. < 20,000
        ii. 20,001 – 100,000
        iii. 100,001 – 500,000
        iv. > 500,000
    b. Commercial Banks.
        i. < 20,000
        ii. 20,001 – 100,000
        iii. 100,001 – 500,000
        iv. > 500,000
    c. Remittance
        i. < 20,000
        ii. 20,001 – 100,000
        iii. 100,001 – 500,000
        iv. > 500,000
    d. Own Savings/Other Source
        i. < 20,000
        ii. 20,001 – 100,000
        iii. 100,001 – 500,000
        iv. > 500,000
3. Do you know your interest rate?
    a. Yes. If so, what is it?
    b. No

E. Questions for Microfinance Borrowers:

1. What was your motivation for obtaining a loan from an MFI?
   a. Easy access to loan
   b. Easy loan repayment
   c. Good services
   d. Amount of loan
   e. Other (please specify):

2. How much money did you initially borrow from an MFI?
   i. < 20,000
   ii. 20,001 – 100,000
   iii. 100,001 – 500,000
   iv. > 500,000
3. Have you taken out additional loans from MFIs since then? If so, how much?
   i. < 20,000
   ii. 20,001 – 100,000
   iii. 100,001 – 500,000
   iv. > 500,000
4. What is the loan repayment period according to the conditions of the MFI?
5. Were you able to repay your loan within the specified?
   a. Yes
   b. No. If so, how long did it/will it take?
6. How do you rate the loan services from the MFIs?
   a. Poor
   b. Good
   a. Very good
   b. Excellent
7. How could MFIs improve your business performance?
   a. Increase the lending ceiling
   b. Offer training
   c. Lengthen loan repayment
   d. Lower the interest rate
8. What challenges, if any, do you face when trying to access loans from the MFIs?
   a. Lack of proper knowledge
   b. Lack of business security
   c. Conditions on repayment
   d. Other specify
9. Do you face any challenges as a loan beneficiary from the MFI's?
   a. Yes.
   b. No
10. If yes state:

F. Business/Employment Performance

1. How do you rate the performance of your business?
   a. Excellent
   b. Above Average
   c. Average
   d. Below Average
2. What was your average daily revenue at the start of your operation?
   i. <1000
   ii. 1,000 – 10,000
   iii. 10,001 – 20,000
   iv. > 20,000
3. What is your daily revenue currently?
   i. <1000
   ii. 1,000 – 10,000
   iii. 10,001 – 20,000
   iv. > 20,000
4. How many employees did you employ at the start of this operation?
   a. Family Member
   b. External Hires

5. How many individuals do you currently employ?
   a. Family Members
   b. External Hire

Testing for Freak Jumps in Exchange Rates

Scott Gilbert, Southern Illinois University Carbondale
Hio Loi, The University of Tennessee at Martin

Abstract

This paper considers a one-time jump in foreign currency exchange rates, whereby an exchange rate is hit by a permanent multiplicative shock at some date. Modelling the jump as a departure from the classic random walk and Brownian motion models of exchange rates, a jump can help to improve on the fit of classical models, while retaining the usefulness of the models post-jump. Focus on a single jump simplifies econometric analysis, in contrast to multi-jump frameworks in the literature. For economists, this simplified framework may usefully bridge familiar economic models of exchange rates and more complicated alternatives in the econometrics literature.

Introduction

This paper considers a jump in a foreign currency exchange rate, whereby the exchange rate is hit by a permanent multiplicative shock at some date. Such a jump may take place at a known date, as in the US dollar to Chinese Yuan exchange rate that China adjusted in year 1994. The present work focuses on a jump at a known date or, alternatively at an unknown date.

A perceived "jump" in an exchange rate, over time, is subjective but may be expressed as a violation of the classical random walk model, whereby the natural logarithm $x_t = ln(X_t)$ of exchange rate X_t takes the form:

$$x_t = x_{t-1} + \varepsilon_t, t = 1,2,\cdots \quad (1)$$

with errors ε_t that are independent and identically distributed normal over time t, each having zero mean and a common variance σ_ϵ^2. For example, X_t may be the number of Euros that can be traded for 1 U.S. dollar at date t, with dates being successive days during which Euros may be traded for dollars in active markets.

In the classic random walk model of exchange rates, the log-return $r_t = x_t - x_{t-1}$ is independent and identically distributed (i.i.d.) over time, a normal iid time series. This random walk model can generate some big changes in exchange rates, over time, yet the time path of the log-return tends to appear smooth if the error variance σ_ϵ^2 is small, as may be appropriate for exchange rates measured at a daily frequency. A "jump" in exchange rates may be manifest as a special kind of departure from the classical random walk model.

In the econometrics literature, exchange rate models are often specified in continuous time, rather than discrete time. The continuous-time version of the discrete-time classical random walk model is Brownian motion, and a continuous time series with one or more jumps may be manifest as a departure from the Brownian motion model. There is a well-known class of such alternative models, the most familiar one being an augmented Brownian motion model that includes "jumps" specified as a Poisson jump process. In the mixed Brownian motion + Poisson jump model, the series x_t has multiple jumps over time, arriving at a rate that matches the Poisson probability distribution. Another alternative is a mixed Brownian motion + Le$\acute{v}$y process model, in which the Le$\acute{v}$y process exhibits infinitely many jumps in each finite time interval. Combining these two alternatives, one arrives at a mixed Brownian motion + Poisson jump + Le$\acute{v}$y process. Applications of these continuous-time econometric models appear in Ait-Sahalia and Jacod (2009, 2010, 2011, 2012), Lee and Hannig (2010), Tsay (2010), Cont and Tankov (2004), Jian and Oomen (2005), Bollerslev and Todorov (2011), Drechler and Yaron (2011), and Erdemlioglu, Laurent, and Neely (2015). For example, in application to intra-day changes in exchange rates, Erdemlioglu et al. (2015) find that 85 percent of exchange rate changes are associated with a Brownian motion component of the relevant time series model, while 15 percent are associated with a jump process.

A hallmark of the recent econometric literature on time series with jumps is a high degree of modelling complexity. One reason for this complexity is that multiple jumps are considered, even an infinity of such jumps in every discrete time period. The present work considers a much simpler setting in which there is at most one jump in the time series. With a single jump, it is straightforward to describe it as an extension of the Brownian motion model, and to test for its statistical significance. Also, from an economic standpoint a single dramatic jump may have greater meaning or interest than a series of relatively small jumps. By introducing a jump into the model, it's possible to accommodate a violation of the model while also preserving its usefulness post-jump. That is, post-jump the Brownian motion model remains useful. In discrete time, the analog statement is that the random walk model of exchange rates remains useful post-jump, despite the jump's violation of the model overall. In this sense, jumps can challenge the convention of classical time series models, yet maintain their predictive usefulness, a curious

theme that is true here and in the more complex multi-jump models in the econometrics literature. In the multi-jump context, the econometric effect of jumps on the skewness, kurtosis, and other higher-order moments of the time series may also be of interest for financial economists interested in pricing foreign currency options and other derivatives. In a single-jump context there is no post-jump effect on higher order moments, yet the jump can be added to a model in which such moments are non-normal.

The proposed methods focus on a specific jump event in an historical time series, either at a known date or unknown date. The former setting is and especially simple and affords an exchange rate jump test based on a t statistic for pre-jump and post-jump levels of log-returns. In the latter setting, we base a jump test on the maximum or minimum of log-returns, over the historical time period, with especially dramatic values of max and min being evidence of a jump. Dramatic max and min values may also be generated by heavy-tailed log-returns, without a "jump", and we allow for heavy tails when testing for a jump.

In application to foreign currencies exchange rates between and 14 countries/areas (Australia, Brazil, Canada, China, Euro, India, Japan, South Korea, Malaysia, Mexico, New Zealand, Switzerland, Thailand, and United Kingdom), estimated jumps at unknown dates, based on max or min log-returns, are large and often statistically significant. This is true when jumps are added to the classical random walk model with normal errors, and when the model is generalized to include non-normal errors. Post-jump, the random walk model remains useful, and the jump itself adds some additional means of representing large historical exchange rate swings that are difficult to produce in familiar non-jump models. In other words, the econometric exercise of working with jumps may help buttress familiar models from criticisms of misspecification, while also supporting the utility of such models post-jump.

If an exchange rate has a single jump it may eventually have more, and our tests for up-jumps (based on a max statistic) and down-jumps (based on a min statistic) often show evidence of both in a single exchange rate series. To the extent that such evidence may point toward a model with multiple jumps, the simple econometric approach developed here may serve as a segue to the multi-jump models in the econometrics literature.

Jump model of exchange rates

The random walk model can be cast continuous time by letting X_t be a continuous-time driftless stochastic process that is defined at all dates $t \geq 0$, as follows:

$$X_t = X_0 e^{\sigma W_t} \tag{2}$$

With σ a volatility parameter and W_t a standard Brownian motion or Wiener process for $t \geq 0$, such that $W_0 = 0$.Taking logarithms, $x_t = ln(X_t) = x_0 + \sigma W_t$, in which case $x_t - x_{t-1} = \sigma(W_t - W_{t-1})$, and setting $\varepsilon_t = W_t - W_{t-1}$, x_t follows a random walk (12).

Consider the possibility that exchange rate X_t experiences a one-time multiplicative "jump" in value, as follows:

$$X_t = X_0(1+\kappa)^{j(t)} e^{\sigma W_t} \tag{3}$$

with κ a parameter in $(-1, \infty)$, σ a positive volatility parameter, and $j(t)$ a jump indicator:

$$j(t) = \begin{cases} 0 & if\ t < t^* \\ 1 & if\ t \geq t^* \end{cases} \tag{4}$$

for some fixed date t^*. This model posits that the process X_t starts out with a continuous sample path until date t^* at which it is hit by a permanent multiplicative increase or decrease in value, with multiplier $1 + \kappa$, dropping down or rising up in value by $100 \times \kappa$ percent, and thereafter has a continuous sample path again.

Figure **1** illustrates the geometric Brownian motion model, and figure **2** illustrates the model with a single added jump at mid-sample, as in equations (2) - (4), based on a sample of $n = 10000$ discrete-time observations of simulated time paths under the no-jump and single-jump alternatives, and parameter values with $\sigma = .0052828$ and $\kappa = 0.8$, calibrated to the U.S.-China exchange rate discussed in more detail later. With a jump represented by a permanent change of $100 \times \kappa$ percent in an exchange rate, let the null hypothesis be that there is no jump:

$H_0: \kappa = 0$

The alternatives are a jump down ($\kappa < 0$) or jump up ($\kappa > 0$), which together form the two-sided alternative:

$H_1: -1 < \kappa < \infty$

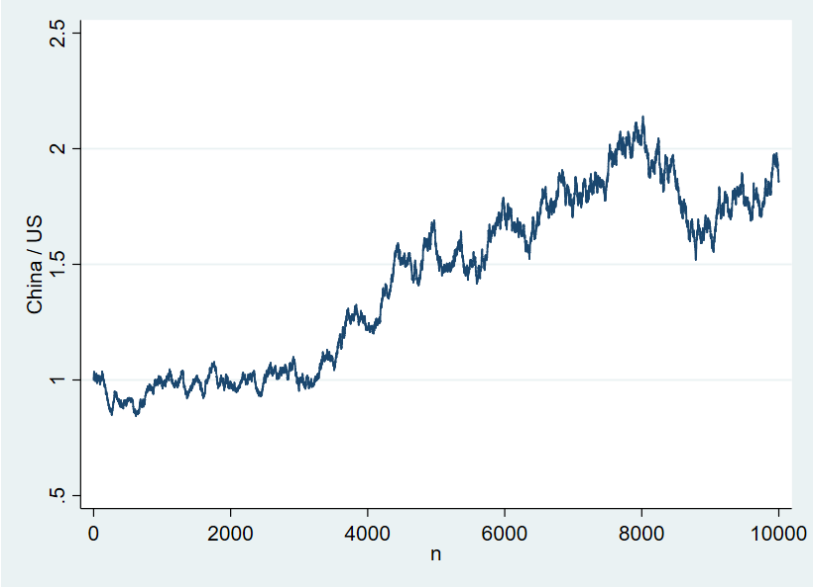


Figure 1: Continuous Path, Simulated Geometric Brownian Motion

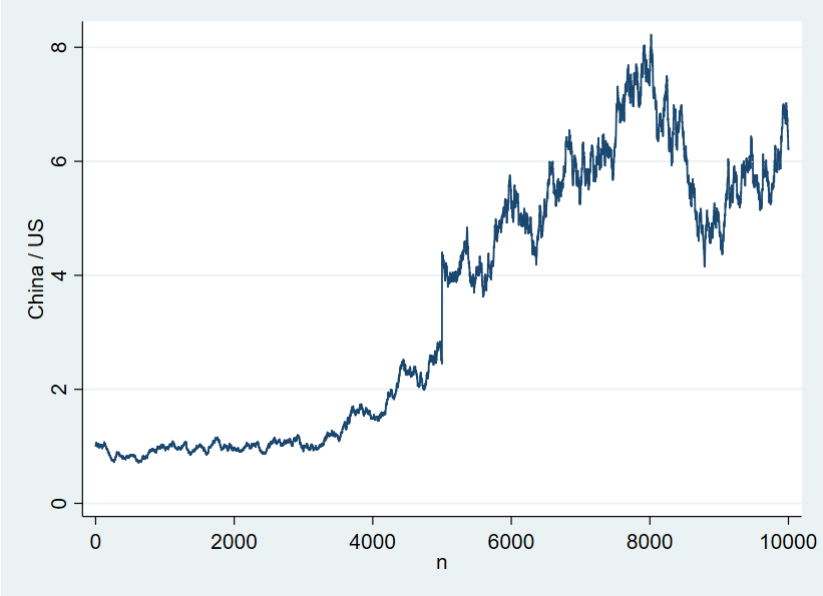


Figure 2: Path with Jump

If a continuous time record for price X_t were available then the existence of jump would be obvious. As in Figure 2, with probability 1 the jump would be manifest as a discrete change in the otherwise-continuous process X_t. Also obvious would be the date t^* at which the jump occurred. The value of κ would be also observable - with probability 1 - since:

$$ln(1+\kappa) = ln(X_{t^*}) - \lim_{t \uparrow t^*} ln(X_t) \tag{5}$$

in which case:

$$\kappa = e^{ln(X_{t^*}) - \lim_{t \uparrow t^*} ln(X_t)} - 1 \tag{6}$$

There is no continuous time series record for foreign exchange rates, so the existence of a jump cannot be determined with certainty. The following two sections of this paper consider the econometric inference for jumps.

Testing for a jump at a known date

In some situations a regime change at a known date may have caused a permanent shift in the exchange rate level. An example would be China's devaluation of the Chinese Yuan in year 1994. Figure 3 shows the US dollar to Chinese Yuan exchange rate, and a variety of others, plotted over time. The marked jump up in the Chinese Yuan per U.S. dollar exchange rate in 1994 coincides with efforts by the Chinese government to make Chinese goods cheaper and therefore more attractive to U.S. consumers. This particular devaluation has historical and economic importance.

To model a jump at a known date, given a discrete time series record of exchange rates let the sample be at dates $t_1, t_2, \cdots, t_n$ for which $t_i - t_{i-1} = \delta$ for some time span $\delta > 0$ and $i = 2, \cdots, n$. Under H_0, the exchange rate X_t is a geometric Brownian motion and log-returns $r_{t_i}, i = 1, \cdots, n$ are independent and identically distributed (iid) normal $N(0, \sigma^2\delta)$ random variables. Under H_a, log-returns are i.i.d. for all observations i except at one of them, call it i^* - for which t^* lies in $[t_i, t_{i+1}]$ - where log-return $r_{t_i^*}$ is normal $N(ln(1+\kappa), \sigma^2\delta)$.

Suppose that the jump date t^* is known. Then one can compute the log-return $r_{t_i^*}$ that includes the posited jump effect, and also test for the effect. If there is no jump then $r_{t_i^*}$ is normal $N(0, \sigma^2\delta)$ while if there is a jump then $r_{t_i^*}$ is normal $N(ln(1+\kappa), \sigma^2\delta)$.

If, in addition, the instantaneous volatility parameter σ has a known value then a simple test statistic would be available, the z statistic:

$$z = \frac{r_{t_i^*}}{\sigma\sqrt{\delta}} \tag{7}$$

Under H_0, the z statistic is standard normal, and provides one-sided tests of H_0 via suitable one-side normal critical values z_α at significance level α, as well as a two-sided test via the two-sided critical value $z_{\alpha/2}$.

In practice, the value of volatility parameter σ is usually unknown, but in the Brownian motion framework it can be estimated with high precision on a finite time interval using samples on which the time interval δ between observations is sufficiently small. As δ approaches 0, the sample standard deviation s of log-returns $r_{t_i}, i = 1, \cdots, n$, converges in probability to σ. Consequently, for the student's *t*-statistic:

$$t - statistic = \frac{r_{t_i^*}}{s\sqrt{\delta}} \tag{8}$$

the distribution of *t*-statistic converges to standard normal under H_0 as δ approaches 0. A test based on the student's *t*-statistic can the reject H_0 if the calculated value of *t*-statistic exceeds the relevant critical value from the normal distribution, with the interpretation that the critical value is asymptotically valid as the time interval δ approaches 0. Under the jump alternative, the sample standard deviation s includes one large deviation - the jump deviation - and for this reason it may be preferable to use instead of s the sample standard deviation s^* computed on all log-returns r_{t_i} except the one observed in the jump interval t_{i^*}. Denoting the modified *t*-statistic as t^*, t^* is again standard normal under H_0 as δ approaches 0.

Example 1: Let Z_t be the the US dollar to Chinese Yuan exchange rate which China reportedly devalued on date 1/3/1994. The maximum return $r_{t_i^*} = 0.40546$, $s = 0.00528$, and let time be measured in days, in which case $\delta = 1$. Then $t - statistics = 76.79167$, and this value exceeds the normal critical value for one-sided and two-sided tests of H_0, at significance level $\alpha = 0.01$. Similarly, $t^* = \cdots$ which likewise exceeds the same normal critical values.

In Example 1, there is a statistically significant upward jump in the US dollar to Chinese Yuan exchange rate on or near the date when China reportedly devalued its currency. The statistical significance relies on the geometric Brownian motion model of exchange rates, and the extension of the model to accommodate a jump. In particular, the test relies on the normality of log-returns r_t. A rejection of the null hypothesis H_0 of no jump could be due to a jump but could also be due to non-normality of r_t. Non-normality is an alternative to the geometric Brownian motion model, as is a jump in the time series path. These two alternatives are also related, in the context of continuous time stochastic processes. Specifically, a Brownian motion has increments that are normally distributed. It is possible to construct continuous-time processes that have increments that are non-normal and have heavier tails than normal, but these processes jump infinitely often in any finite time interval, with probability equal to 1. For this reason, to have non-normal fat-tailed i.i.d. returns r_t based on a continuous time model, the actual data generating process may have infinitely many jumps over time. If so, fat-tailed returns are not necessarily different than a jump possibility, but may suggest infinitely many jumps - rather than a single one.

In a discrete time framework, the null hypothesis H_0 can be restated as $E[r_{t^*}] = 0$. To test for a zero population mean in log-return $r_{t_i^*}$, suppose that under H_0, the time series $r_{t_i}, i = 1, \cdots, n$ is i.i.d. with some known continuous cumulative distribution function $F_0(\cdot)$ having zero population mean. Then, with q_p being the distribution's quantile at cumulative

probability p, under H_0 the statistic $r_{t_i^*}$ exceeds $q_{1-\alpha}$ with probability α, and this provides an upper-tailed one-sided test at significance level α. Similarly, under H_0, $r_{t_i^*}$ falls below q_α with probability α, providing a one-sided lower-tailed test. If log-returns are non-normal under H_0 then the actual distribution $F_0(\cdot)$ may not be known but can be estimated via a parametric model - such as the Student's t - distribution - or via nonparametric methods.

Testing for a jump at an unknown jump date

If the jump date is unknown then the best guess of where a jump has occurred is where the return r_{t_i} is either biggest jump or drop on price during the historical sample. To test for a jump in this setting, it suffices to compare the sample minimum log return $min\ (r_t)$ and maximum log return $max\ (r_t)$ to a relevant critical value under H_0. Since $r_{t_i}, i = 1, \cdots, n$ is i.i.d. normal underH_0, the distribution of the maximum is:

$$P(max\ (r_t) \geq x) = \big(P(r_t \geq x)\big)^n \tag{9}$$

To get a critical value c_α, for the statistic $max\ (r_t)$, set $P(max\ (r_t) \geq c_\alpha) = 1 - \alpha$, with α the test's desired significance level. Then, using equation (9), the critical value c_α solves:

$$c_\alpha = q_r\Big((1-\alpha)^{1/n}, \mu_r, \sigma_r\Big) \tag{10}$$

With q_r the quantile of the normal distribution of r, evaluated at probability $(1-\alpha)^{1/n}$, where $\mu_r = 0$ and σ_r are the mean and standard deviation of r.

It is easy to carry out this test if μ_r and σ_r are known, since normal quantiles q_r are easy to calculate. If the time step size δ is small then estimation error in the sample standard deviation of r is negligible, in which case the assumption that σ_r is known becomes reasonable. The assumption that μ_r is known may be more problematic, but for application to exchange rates it is reasonable to set μ_r equal to 0.

Application

As an illustration of the jump test, we apply 14 daily foreign currencies exchange rates (not seasonally adjusted) in this study, and all foreign exchange rates data are obtained from the Federal Reserve of St. Louis website. The exchange rates use the US dollar as a base currency, so a larger value of the exchange rate represents US dollar depreciation. Table 1 shows a summary statistic of the data. It includes the number of observations, mean, standard deviation, minimum value, maximum value, starting date and ending date for each foreign exchange rate. The sample sizes vary across different foreign exchange rates given some of it have different starting date, but the ending date (1/4/2019) is all the same. The South Korea currency is highly fluctuated since its standard deviation is the highest among the 14 foreign exchange rates. In contrast, the US dollar to UK pound exchange rate is the most stable one given the smallest value of standard deviation among others.

Table 1: Summary of U.S. dollars to foreign currencies exchange rates

Foreign exchange	Obs.	Mean	Std. Dev.	Min	Max	Starting Date	Ending Date
Australia/US	12,039	1.217	0.292	0.672	2.071	1/4/1971	1/4/2019
Brazil/US	6,029	2.209	0.803	0.832	4.202	1/2/1995	1/4/2019
Canada/US	12,052	1.220	0.166	0.917	1.613	1/4/1971	1/4/2019
China/US	9,486	6.171	2.136	1.526	8.741	1/2/1981	1/4/2019
Euro/US	5,025	0.846	0.128	0.625	1.209	1/4/1999	1/4/2019
India/US	11,538	32.159	19.640	7.190	74.330	1/2/1973	1/4/2019
Japan/US	12,040	160.250	72.911	75.720	358.440	1/4/1971	1/4/2019
South Korea/US	9,432	984.963	204.207	667.200	1960.000	4/13/1981	1/4/2019
Malaysia/US	12,024	3.014	0.632	2.105	4.730	1/4/1971	1/4/2019
Mexico/US	6,314	11.509	3.770	3.102	21.891	11/8/1993	1/4/2019
New Zealand/US	12,030	1.460	0.395	0.671	2.551	1/4/1971	1/4/2019
Switzerland/US	12,046	1.663	0.729	0.730	4.318	1/4/1971	1/4/2019
Thailand/US	9,465	31.328	6.608	20.360	56.100	1/2/1981	1/4/2019
UK/US	12,046	0.595	0.099	0.378	0.951	1/4/1971	1/4/2019

Data source: Federal Reserve Bank of St. Louis

Sample size is between 1/4/1971 to 1/4/2019, Daily, Not seasonally adjusted. Some foreign exchange rates are not available for the whole time period. Observations for Saturdays, Sundays and public holidays are excluded.

Figure 3 graphs the historical foreign exchange rates for all 14 countries. Some foreign exchange rates experienced large appreciation or depreciation in a short period of time. For example, Chinese Yuan to US dollar depreciated dramatically because the Chinese government tried to help the export of Chinese products. In addition, the Asia Financial Crisis in 1997 had caused larger exchange rates depreciation in South Korean, Malaysia, and Thailand. Besides some major historical events, exchange rates are still highly fluctuated in other times. Therefore, figure 3 shows that jumps could be a common phenomena in foreign exchange rates.

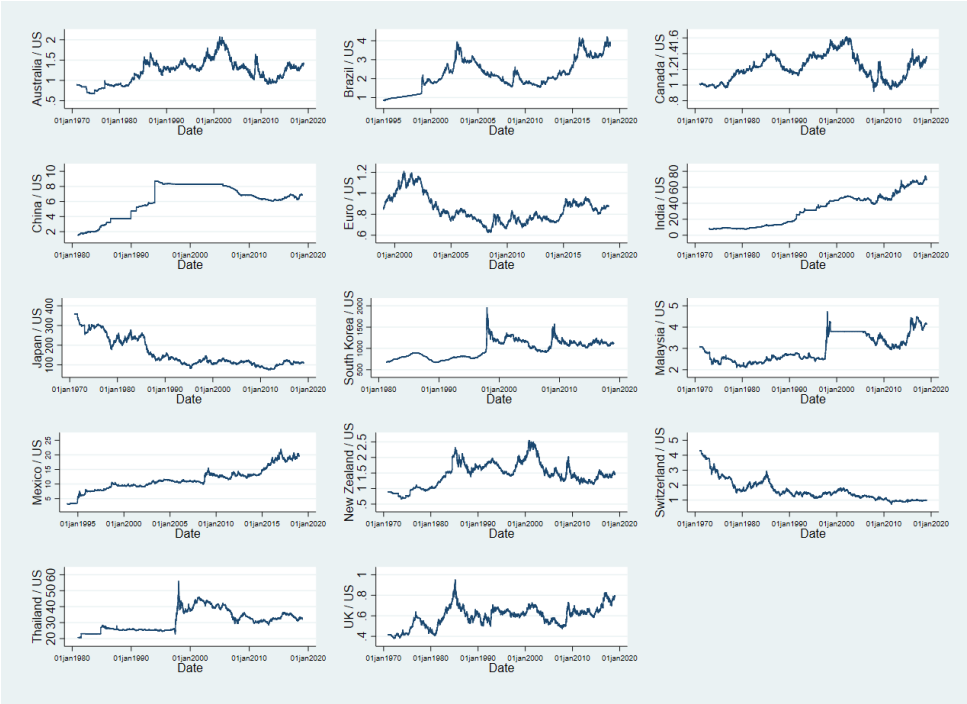


Figure 3: Graph the data of historical foreign exchange rates

Brownian motion simulation

Using equations (2) and (3), figures 4 and **5** graph the simulation of Brownian motion without and with a single added jump, assuming the jump parameter equals to 0.8 ($\kappa = 0.8$). In the simulations, the arbitrary jump is set to exit in the middle of the simulation where $n = 500$, and the jump parameter is a positive number, so there are upward jumps in figure **5**. The simulations use the first available observation for each foreign exchange rate as a starting value and the standard deviation of each exchange rate returns listed in table (2).

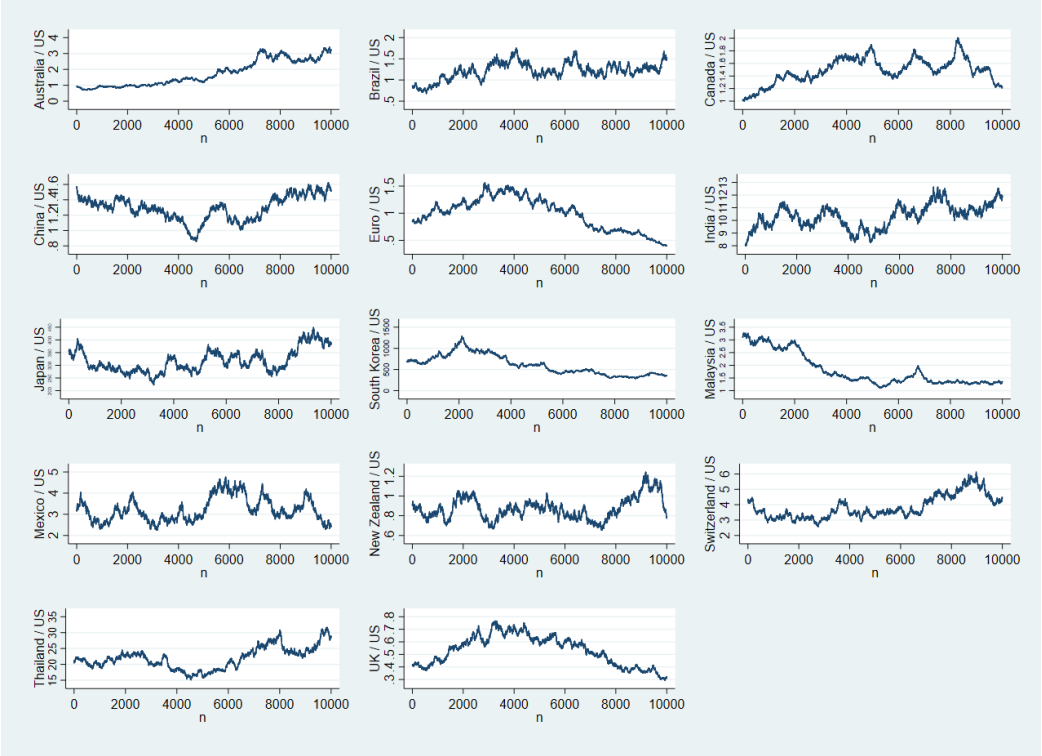


Figure 4: Graph the simulation of Brownian motion

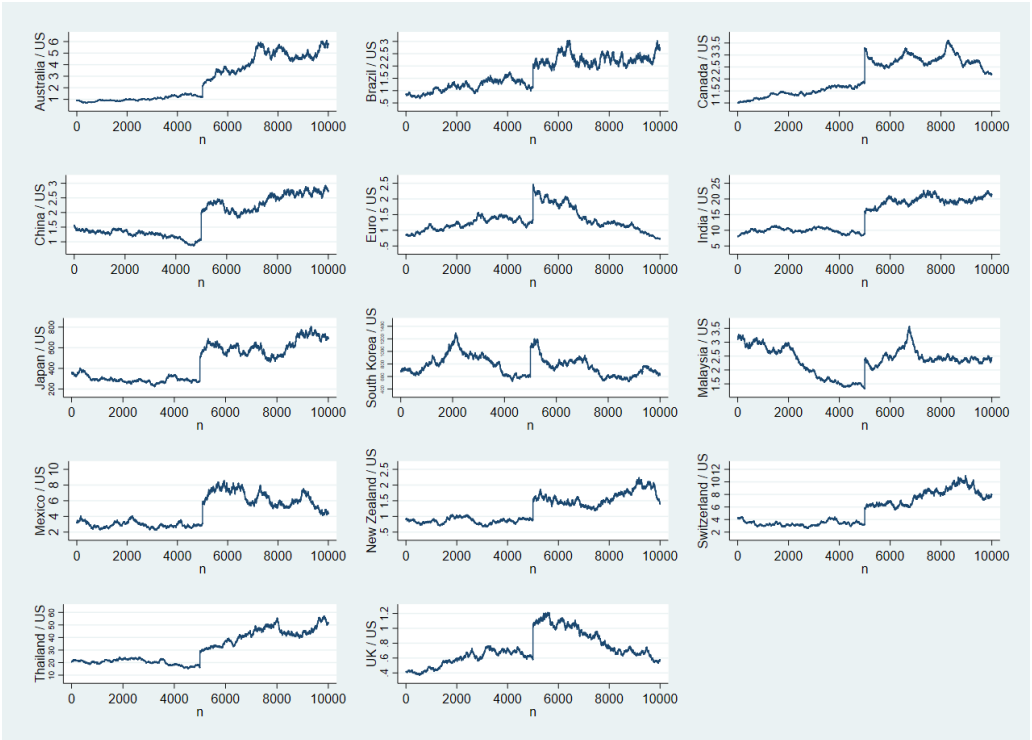


Figure 5: Graph the simulation of Brownian motion with a single jump exist

Testing for jumps empirically

Table 2: Summary of the returns of exchange rates

Return	Obs.	Mean	Std. Dev.	$min\ (r_t)$	Min. Date	$max\ (r_t)$	Max. Date
Australia/US	11,565	0.00004	0.00676	-0.06667	10/29/2008	0.19245	11/29/1976
Brazil/US	5,796	0.00015	0.00982	-0.08951	10/9/2008	0.11441	1/15/1999
Canada/US	11,582	0.00001	0.00408	-0.05072	10/29/2008	0.03807	11/20/2008
China/US	9,087	0.00013	0.00528	-0.02429	9/24/1984	0.40546	1/3/1994
Euro/US	4,831	-0.00003	0.00609	-0.04621	3/19/2009	0.03003	12/19/2008
India/US	11,080	0.00014	0.00480	-0.05435	3/15/1973	0.12811	7/3/1991
Japan/US	11,568	-0.00010	0.00633	-0.05630	10/7/1998	0.06256	1/7/1974
South Korea/US	9,032	0.00002	0.00656	-0.13222	10/30/2008	0.13645	12/23/1997
Malaysia/US	11,548	0.00001	0.00441	-0.09157	12/8/1997	0.07196	1/6/1998
Mexico/US	6,065	0.00024	0.00878	-0.17969	3/10/1995	0.19343	12/22/1994
New Zealand/US	11,554	0.00002	0.00715	-0.09295	9/10/1973	0.16280	8/11/1975
Switzerland/US	11,574	-0.00015	0.00710	-0.13022	1/15/2015	0.05827	11/1/1978
Thailand/US	9,045	0.00000	0.00543	-0.07408	7/23/1987	0.20769	7/2/1997
UK/US	11,574	0.00003	0.00587	-0.04589	3/27/1985	0.08169	6/24/2016

The return of foreign exchange is defined as $r_t = ln(X_t) - ln(X_{t-1})$.

The empirical test in this study focuses on the situation where the jump date is unknown. The best guess of an unknown jump is where the biggest changes/returns occur, and those changes/returns can be either positive or negative. Thus, the minimum and maximum returns are used for the empirical tests. Table (2) provides the statistic summary of the returns of exchange rates including the number of observations, means, standard deviations, minimum values, maximum values and the dates when the minimum and maximum returns occur. The returns are defined as the first difference natural logarithm of US dollar to foreign currency $\left(r_t = ln(X_t) - ln(X_{t-1})\right)$.

Table 3: The results of hypotheses tests (both lower-tailed and upper-tailed tests) under the unknown jump date situation.

	Lower-tailed test			Upper-tailed test		
Return	$min\ (r_t)$	**Critical value**	**Conclusion**	$max\ (r_t)$	**Critical value**	**Conclusion**
Australia/US	-0.06667	-0.03003	Reject H_0	0.19245	0.03003	Reject H_0
Brazil/US	-0.08951	-0.04216	Reject H_0	0.11441	0.04216	Reject H_0
Canada/US	-0.05072	-0.01813	Reject H_0	0.03807	0.01813	Reject H_0
China/US	-0.02429	-0.02320	Reject H_0	0.40546	0.02320	Reject H_0
Euro/US	-0.04621	-0.02590	Reject H_0	0.03003	0.02590	Reject H_0
India/US	-0.05435	-0.02127	Reject H_0	0.12811	0.02127	Reject H_0
Japan/US	-0.05630	-0.02811	Reject H_0	0.06256	0.02811	Reject H_0
South Korea/US	-0.13222	-0.02881	Reject H_0	0.13645	0.02881	Reject H_0
Malaysia/US	-0.09157	-0.01957	Reject H_0	0.07196	0.01957	Reject H_0
Mexico/US	-0.17969	-0.03776	Reject H_0	0.19343	0.03776	Reject H_0
New Zealand/US	-0.09295	-0.03178	Reject H_0	0.16280	0.03178	Reject H_0
Switzerland/US	-0.13022	-0.03154	Reject H_0	0.05827	0.03154	Reject H_0
Thailand/US	-0.07408	-0.02383	Reject H_0	0.20769	0.02383	Reject H_0
UK/US	-0.04589	-0.02606	Reject H_0	0.08169	0.02606	Reject H_0

The critical values on this table is assumed under a normal distribution with mean values equal zero $(\mu = 0)$ and $\alpha = 5\%$.

Table (3) shows the results of one-tailed hypothesis tests. The critical values are obtained through equation (10) with α and mean equal 5% and 0, respectively. The values of the number of observations and standard deviations depend on each exchange rates sample. Notice that the critical values are calculated based on a normal distribution, so the values are the same with just opposite sign for lower-tailed and upper-tailed tests. The results in table (3) show that all foreign exchange rates reject the null hypotheses for both lower-tailed and upper-tailed tests which mean significant downward and upward jumps exist in all observed exchange rates.

Table (3) assumes the return of all foreign exchange rates follow normal distribution; however, this assumption may not be realistic given the large fluctuations across different exchange rates. In fact, table (**4**) show the values of kurtosis for all

exchange rates are larger than 3, so the distributions are not normally distributed. Student's t-distribution is applied in order to improve the preciseness of the statistic inferences. Furthermore, the critical values may be influenced by the outlier observations. To address the issue, the top and bottom 1% observations are deleted from the sample size. Table (**4**) shows the kurtosis values from the trimmed sample. The calculation of critical value under student's t-distribution requires appropriate values for the degree of freedom. Thus, the excess kurtosis equation is used:

$$kurtosis = \begin{cases} 3 + \frac{6}{\nu - 4} & for\ \nu > 4 \\ \infty & for\ 2 < \nu \leq 4 \\ & otherwise\ unfined. \end{cases} \quad (11)$$

ν is the degree of freedom, in which case, ν equals:

$$\nu = \frac{6}{kurtosis - 3} + 4 \quad (12)$$

The values of degree of freedom for all foreign exchange rates are listed in table (4).

Table 4: Kurtosis, trimmed Kurtosis and degree of freedom

Return	Kurtosis	Kurtosis*	Degree of freedom**
Australia/US	86.51460	4.25681	8.77398
Brazil/US	17.88601	4.26683	8.73624
Canada/US	13.44743	4.32215	8.53806
China/US	4267.40290	13.62542	4.56468
Euro/US	5.35846	3.12871	50.61762
India/US	92.74268	5.98403	6.01071
Japan/US	8.78030	3.79234	11.57253
South Korea/US	102.90471	6.35787	5.78685
Malaysia/US	67.52071	6.67999	5.63044
Mexico/US	105.34316	4.24094	8.83504
New Zealand/US	42.01824	4.40061	8.28384
Switzerland/US	16.33246	3.47370	16.66620
Thailand/US	281.43889	7.57666	5.31100
UK/US	10.34834	3.54939	14.92116

* : The values of kurtosis are obtained from a trimmed sample where the top and bottom 1% observations are deleted.

**: The degree of freedom is obtained through the excess kurtosis equation under Student's t - distribution ($\frac{6}{\nu-4}$ for $\nu > 4$, ∞ for $2 < \nu \leq 4$, otherwise undefined, where ν is the degree of freedom).

Table 5: The results of hypotheses tests (both lower-tailed and upper-tailed tests) under the unknown jump date situation.

	Lower-tailed test			Upper-tailed test		
Return	$\boldsymbol{min\ (r_t)}$	**Critical value**	**Conclusion**	$\boldsymbol{max\ (r_t)}$	**Critical value**	**Conclusion**
Australia/US	-0.06667	-0.06184	Reject H_0	0.19245	0.06184	Reject H_0
Brazil/US	-0.08951	-0.08267	Reject H_0	0.11441	0.08267	Reject H_0
Canada/US	-0.05072	-0.03822	Reject H_0	0.03807	0.03822	Fail to reject H_0
China/US	-0.02429	-0.10853	Fail to reject H_0	0.40546	0.10853	Reject H_0
Euro/US	-0.04621	-0.02856	Reject H_0	0.03003	0.02856	Reject H_0
India/US	-0.05435	-0.06574	Fail to reject H_0	0.12811	0.06574	Reject H_0
Japan/US	-0.05630	-0.04750	Reject H_0	0.06256	0.04750	Reject H_0
South Korea/US	-0.13222	-0.09145	Reject H_0	0.13645	0.09145	Reject H_0
Malaysia/US	-0.09157	-0.06674	Reject H_0	0.07196	0.06674	Reject H_0
Mexico/US	-0.17969	-0.07366	Reject H_0	0.19343	0.07366	Reject H_0
New Zealand/US	-0.09295	-0.06882	Reject H_0	0.16280	0.06882	Reject H_0
Switzerland/US	-0.13022	-0.04468	Reject H_0	0.05827	0.04468	Reject H_0
Thailand/US	-0.07408	-0.08590	Fail to reject H_0	0.20769	0.08590	Reject H_0
UK/US	-0.04589	-0.03864	Reject H_0	0.08169	0.03864	Reject H_0

The critical values on this table is obtained under Student's t - distribution.

Table (5) shows the results of one-tailed hypothesis tests with the critical values obtained from Student's t - distribution. There are several cases fail to reject the null hypothesis. US dollar to Chinese Yuan exchange, US dollar to Indian rupee exchange and US dollar to Thai baht exchange fail to reject the null hypotheses in the lower-tailed tests which means these three exchange rates do not experience significant downward jumps. In contrast, only US dollar to Canadian dollar fails to reject the upper-tailed test which indicates this foreign exchange rate has no significant upward jump. For all other situations, the results are the same as table (3). Thus, most of the foreign currency exchanges rate face both downward and upward jumps.

Conclusion

This paper examines a one-time jump in 14 US dollar to foreign currency daily exchange rates over the time span during the period 1/4/1971 - 1/4/2019. The results suggest that jumps are statistically significant in both downward and upward directions, and these jumps appear in most foreign exchange rates.

This study discloses movements of foreign exchange rates using the jump test; however, one should not over interpret the results. The existence of significant jumps in foreign exchange rate does not imply anything about market failure. The fluctuation of the foreign exchange rate depends on both domestic and foreign economies, and also a country's economy could be influenced by a lot of factors. The large fluctuation of foreign currency exchange rate seems reasonable. Thus, the results in this study is a warning signal for investors who are interested in the foreign currency exchange market.

References

Aït -Sahalia, Yacine, and Jean Jacod. 2009. “Testing for Jumps in a Discretely Observed Process.” *The Annals of Statistics* 37: 184-222.

Aït-Sahalia, Yacine, and Jean Jacod. 2010. "Is Brownian motion necessary to model high-frequency data?" *The Annals of Statistics* 38: 3093-3128.

Aït-Sahalia, Yacine, and Jean Jacod. 2011. "Testing whether jumps have finite or infinite activity." *The Annals of Statistics* 39: 1689-1719.

Aït-Sahalia, Yacine, and Jean Jacod. 2012. "Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data." *Journal of Economic Literature* 50: 1007-50.

Bollerslev, Tim, and Viktor Todorov. 2011. "Estimation of jump tails." *Econometrica* 79: 1727-1783.

Cheah, Eng-Tuck, and John Fry. 2015. "Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin." *Economics Letters* 130: 32-36.

Cont, Rama, and Peter Tankov. 2004. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC Financial Mathematics Series.

Drechsler, Itamar, and Amir Yaron. 2010. "What's vol got to do with it." *The Review of Financial Studies* 24: 1-45.

Erdemlioglu, Deniz, Sébastien Laurent, and Christopher J. Neely. 2015. "Which continuous-time model is most appropriate for exchange rates?" *Journal of Banking & Finance* 61: S256-S268.

Jiang, George, and Roel Oomen. 2005. "A new test for jumps in asset prices." *Preprint*.

Johansen, Anders, Olivier Ledoit, and Didier Sornette. 2000. "Crashes as critical points." *International Journal of Theoretical and Applied Finance* 3: 219-255.

Lee, Suzanne S., and Jan Hannig. 2010. "Detecting jumps from Lévy jump diffusion processes." *Journal of Financial Economics* 96: 271-290.

Lo, Andrew W., and A. Craig MacKinlay. 2002. *A non-random walk down Wall Street*. Princeton University Press.

MacDonell, Alec. 2014. "Popping the Bitcoin bubble: An application of log-periodic power law modeling to digital currency." *University of Notre Dame working paper*.

Tsay, Ruey S. 2005. *Analysis of financial time series*. Vol. 543. John wiley & sons.

Øksendal, Bernt. 2003. *Stochastic Differential Equations: An Introduction.* 6th edition. Springer.

Interest Rate Pass-Through from the Effective Federal Funds Rate to Various Market Interest Rates

Natalie Hegwood, Mark Tuttle, and M. Douglas Berg, Sam Houston State University

Abstract

This paper examines the effective federal funds rate long-run pass-through for various short-term and long-term U.S. interest rates. Interest rate pass-through is the monetary policy transmission mechanism by which changes in a policy-controlled or targeted interest rate are reflected in retail interest rates. The degree and speed of this pass-through is therefore a measure of the effectiveness of monetary policy and responsiveness of bank operations. Previous research produces mixed pass-through results, finding either complete or incomplete pass-through. Our results suggest complete pass-through in seven of the eleven rates studied.

Introduction

Interest rate pass-through is the monetary policy transmission mechanism by which changes in a policy-controlled or targeted interest rate are reflected in retail interest rates. Those changes to retail rates then have an impact on the economic activity of households and firms. Understanding the timing and magnitude of interest rate pass-through is necessary in order to properly devise effective monetary policy. Ideally policy changes would be fully and quickly transmitted to retail rates, however in practice this has shown not to be the case. Hegwood and Tuttle (2017) find complete pass-through but with a long adjustment period. Payne and Waters (2008), Sander and Kleimeier (2004 & 2006) find incomplete pass-through. Hofmann and Mizen (2004), Becker, Osborn and Yildirim (2012), and Sander and Kleimeier (2006) have mixed results for different interest rates. This interest rate rigidity may be related to adjustment costs and asymmetric information in credit markets as discussed by Stiglitz and Weiss (1981). In addition, Sellon (2002) proposes that changes in financial institutions and the structure of financial markets, as have occurred in the U.S. in recent decades, may alter the timing and magnitude of retail interest rate responses to changes in monetary policy.

Most of the above mentioned studies for the U.S. have focused on a single pass-through relationship (for example: the effective federal funds rate to the prime rate, or the effective federal funds rate to 30 year mortgage rates.) However those that study non-U.S. rates have taken a broader approach and examined a cross section of retail interest rates. In this paper we examine interest rate pass-through from the effective federal Funds rate to a variety of primarily U.S. interest rates. We include both long and short-term rates and find somewhat mixed results. We also examine the pattern of adjustment to long run equilibrium and find that for many, the effective federal funds rate appears to be making a larger share of the adjustment which runs counterintuitive to the idea that policy changes are being passed through to retail rates.

Literature Review

In general the results for pass-through from a policy controlled or targeted interest rate to retail rates are mixed. When complete pass-through is found, it is usually accompanied by either a structural break or a lengthy adjustment period. Hegwood and Tuttle (2017) find complete pass through between the U.S. effective federal funds rate to long term mortgage interest rates, but with a half-life of adjustment measuring 20 months. Payne and Waters (2008) find incomplete pass-through from the effective federal funds rate to the U.S. prime rate unless a structural break is included. Sander and Kleimeier (2004) also include a structural break. They find that for EU countries, the pass-through remains incomplete but is improved by the addition of the structural break. Hansen and Welz (2011) find long run complete pass-through for all rates that they study for Sweden. However, others have found that the degree of pass-through and speed of adjustment vary across different types of retail interest rates. Hofmann and Mizen (2004) find complete pass-through in the long run from the U.K. base rate to deposit rates but not to mortgage rates. Becker, Osborn and Yildirim (2012) find complete pass-through from the U.K. base rate to the LIBOR but incomplete pass-through from the LIBOR to mortgage rates. Finally, Sander and Kleimeier (2006) find more complete pass-through for loan rates than deposit rates in the EU using a one-month money market rate as a proxy for a policy rate.

Data and Unit Root Tests

The Federal Reserve Bank of St. Louis provides interest rate data from various sources. Various interest rates were collected as monthly data from FRED (Federal Reserve Bank of St. Louis). The effective federal fund rate, various Treasury rates, and

the sixty-day AA non-financial commercial paper rate are collected from the Board of Governors of the Federal Reserve System. The twelve-month London InterBank Offered Rate (LIBOR) is collected from ICE Benchmark Administration Limited. Two conventional, fixed mortgage rates are also included, the fifteen-year conventional mortgage interest rate and the thirty-year conventional mortgage interest rate. These two mortgage interest rates come from Freddie Mac. Finally, the Aaa corporate bond yield is collected from Moody's. The number of observations for each series is provided in the "Sample" column of Table 1.

Interest rate pass-through is measured through the cointegration of the policy interest rate and other retail interest rates. Prior to testing for cointegration unit roots must be established for each of the variables involved. Data ranges vary for the series and are provided in Table 1: Unit Root Tests. We use the Dickey-Fuller GLS (Elliott, Rothenberg, and Stock, 1996) and Ng and Perron (2001) unit root tests. The number of lags is determined using the modified Akaike Information Criteria (Ng and Perron, 2001), and both tests include an intercept term. Table 1 shows the results of these tests. In all cases we fail to reject the null of a unit root with both tests. Subsequent testing rejects the null of a unit root for first-difference of all series. Thus, the results suggest that all series are integrated of order one.

Table 1: Unit Root Tests

			DF-GLS		Ng and Perron	
Variable	**Sample**	**Number of Observations**	**Number of Lags***	**Test Statistic**	**Number of Lags***	**Test Statistic (Z_t)**
Effective Federal Funds Rate	01:1982 - 9:2018	441	9	-0.13	2	-1.81
Three-Month Treasury	12:1982 - 9:2018	441	6	0.03	6	-1.62
One-Year Treasury	08:1982 - 9:2018	441	16	0.06	3	-1.47
Two-Year Treasury	08:1982 - 9:2018	441	3	0.15	3	-1.46
Five-Year Treasury	08:1982 - 9:2018	441	3	0.29	5	-1.38
Seven-Year Treasury	08:1982 - 9:2018	441	3	0.37	5	-1.40
Ten-Year Treasury	08:1982 - 9:2018	441	3	0.45	3	-1.53
Sixty-Day Commercial Paper	11:1997 - 8:2018	259	3	-1.00	3	-2.48
Twelve-Month LIBOR	04:1987 - 8:2018	392	10	-0.87	2	-2.21
Fifteen-Year Mortgage	06:1992 - 8:2018	325	5	0.20	2	-2.71
Thirty-Year Mortgage	08:1982 - 8:2018	440	3	0.74	5	-0.86
Aaa Bond	10:1997 - 8:2018	261	6	-0.49	6	-2.66

* Number of lags in DF-GLS determined using the modified Akaike Information Criteria and all tests include an intercept.
Five-percent critical value in the DF-GLS test is -1.94 and the five-percent critical value in the Ng and Perron test is -2.91.

Cointegration Tests

We use the method of Johansen (1995) to test for cointegration. This method allows tests of long-run pass-through and weak exogeneity. Equation (1) provides the vector autoregression model (VAR) with k lags used to test for cointegration:

$$y_t = \mu + \rho D_t + \sum_{i=1}^{k} \pi_i y_{t-i} + e_t \quad (1)$$

Here, y_t is a $(2\ x\ 1)$ vector that contains the interest rate of interest and the effective federal funds rate, π_i ρ_i are a $(1\ x\ 2)$ matrices of parameters, and D_t is a matrix of deterministic variables. The number of lags, k, is determined using the Akaike Information Criteria (AIC). To identify the appropriate deterministic components we employ the Pantula principle (Johansen, 1992). Using the Pantula principle, four specifications are considered: no deterministic components, a constant restricted to the cointegration vector, an unrestricted constant, and an unrestricted constant and a time trend restricted to the cointegration vector. We start with the most restrictive (no deterministic components) and test for cointegration, and select the first specification that rejects the null hypothesis of no cointegration.

Table 2 provides the results of the cointegration tests. The sample period and number of k lags included in the test are provided in columns two and three. All but one test includes a restricted constant in the test and all tests exclude a restricted time trend. The cointegration tests between the three-month Treasury rate and the effective federal funds rate excludes all deterministic components. The cointegration test results, columns five through eight, indicate that the null of no cointegration is rejected between each interest rate and the effective federal funds rate in all cases. All tests fail to reject the null hypothesis of one stationary combination of the respective interest rate and the effective federal funds rate.

Table 2: Cointegration Tests - Effective Federal Funds Rate and Interest Rate

				Cointegration Test Results			
Variable	**Sample**	**Number of Lags**	**Deterministic Components**	**Trace Test Statistic, r = 0**	**Critical Value, r = 0**	**Trace Test Statistic, r <=1**	**Critical Value, r <= 1**
Three-Month Treasury	12:1982 - 9:2018	11	None	15.68	12.32	2.58	4.13
One-Year Treasury	8:1982 - 9:2018	6	Rest. Constant	33.92	20.26	7.23	9.16
Two-Year Treasury	8:1982 - 9:2018	6	Rest. Constant	28.26	20.26	7.36	9.16
Five-Year Treasury	8:1982 - 9:2018	6	Rest. Constant	22.51	20.26	7.52	9.16
Seven-Year Treasury	8:1982 - 9:2018	6	Rest. Constant	21.69	20.26	7.39	9.16
Ten-Year Treasury	8:1982 - 9:2018	6	Rest. Constant	20.95	20.26	7.18	9.16
Sixty-Day Commercial Paper	11:1997 - 8:2018	9	Rest. Constant	39.42	20.26	5.07	9.16
Twelve-Month LIBOR	4:1987 - 8:2018	14	Rest. Constant	28.89	20.26	3.84	9.16
Fifteen-Year Mortgage	6:1992 - 8:2018	9	Rest. Constant	39.28	20.26	3.19	9.16
Thirty-Year Mortgage	8:1982 - 8:2018	6	Rest. Constant	27.47	20.26	7.85	9.16
Aaa Bond	10:1997 - 8:2018	9	Rest. Constant	32.89	20.26	4.28	9.16

Normalizing on the various interest rates, column two of Table 3 provides the long-run coefficients on the effective federal funds rate. In many cases the long-run coefficient is close to one which implies that the given interest rate and the effective federal funds rate move in a one-to-one fashion in the long run. The cointegration vectors for assets with shorter durations, the three-month through five-year Treasuries, the sixty-day commercial paper, and the twelve-month LIBOR rates, have long-run coefficients on the effective federal funds rate that range from 0.949 to 1.040. With long-run coefficients essentially equal to one, this implies that for rates in which a constant was included in the model, that constant is the long-run spread between that interest rate and the effective federal funds rate. The 3-month Treasury rate did not include a constant and so it is essentially equal to the federal funds rate in the long run. Assets with longer durations have long-run coefficients on the effective federal funds rate that range from 0.740 to 0.899.

To test for complete pass-through, the multivariate autoregressive model in Equation (1) is formed into Equation (2) below. The pass-through test restricts the vector $\boldsymbol{\beta}'$ in Equation (2) to $[1 \quad -1 \quad *]$, where "$*$" denotes an unrestricted constant in the cointegration vector, if included. Columns four and five of Table 3 provide the results of the pass-through tests.

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_{k-1}\boldsymbol{\Delta y}_{t-(k+1)} + e_t \quad (2)$$

For seven of the interest rates included in Table 3, the test fails to reject complete pass-through between the effective federal funds rate and the rate of interest. In other words, the test fails to reject that the given interest rate and the effective federal funds rate move in a one-to-one relationship in the long run. Interestingly, we reject complete pass-through between the three-month Treasury rate and the effective federal funds rate. Finally, we also reject complete pass-through between the effective federal funds rate and the three long-term interest rates: the two conventional mortgage interest rates and the Aaa rate. This is consistent with Payne (2006) who cites switching costs and informational asymmetries as potential causes for a lack of complete pass-through to mortgage rates. In addition, Sellon (2002) states that pass-through should be more complete between rates of similar maturities. Therefore it is to be expected that long-term rates would have less complete pass-through from the effective federal funds rate.

Also of interest is the pattern of adjustment, given by the adjustment parameters from the cointegration tests presented in columns 6 through 9 of Table 3. The six Treasury rates, sixty-day commercial paper rate, and the twelve-month LIBOR rate are weakly exogenous. Therefore, the tests fail to reject the non-response of these interest rates to shocks in the rate of interest or the effective federal funds rate in the long run. Conversely, the effective federal funds rate adjustment parameters are statistically different from zero. Thus, the results suggest that the effective federal funds rate is the only interest rate in these relationships that responds to disequilibrium between these interest rates in the long run. Further, this process is faster when adjusting to short-term interest rate shocks.

Table 3: Cointegration Test Results: Long-Run Coefficients and Adjustment Parameters, Effective Federal Funds Rate and Interest Rate

			Pass-Through Test		**Adjustment Parameters**			
Variable	**Long-Run Coefficient**	**Standard Error**	**Chi-Square Test Statistic**	**p-value**	**Given Interest Rate**	**Standard Error**	**Effective Federal Funds Rate**	**Standard Error**
Three-Month Treasury	0.949	0.009	6.61	0.01	-0.031	0.049	0.114	0.046
One-Year Treasury	1.010	0.021	0.08	0.79	0.029	0.031	0.131	0.027
Two-Year Treasury	1.040	0.038	0.64	0.42	-0.004	0.022	0.070	0.018
Five-Year Treasury	0.953	0.070	0.24	0.62	-0.013	0.014	0.038	0.016
Seven-Year Treasury	0.899	0.080	0.80	0.37	-0.013	0.012	0.031	0.010
Ten-Year Treasury	0.832	0.089	1.63	0.20	-0.012	0.010	0.026	0.009
Sixty-Day Commercial Paper	1.010	0.006	1.90	0.17	-0.044	0.109	0.384	0.097
Twelve-Month LIBOR	0.983	0.020	0.56	0.46	-0.047	0.043	0.093	0.026
Fifteen-Year Mortgage	0.866	0.062	2.74	0.10	-0.022	0.010	0.037	0.007
Thirty-Year Mortgage	0.740	0.081	5.87	0.02	-0.034	0.009	0.008	0.009
Aaa Bond	0.810	0.060	4.11	0.04	-0.079	0.023	0.045	0.011

The three long-term interest rates (the fifteen and thirty-year mortgage interest rates and the Aaa rate) are endogenous in the long run, responding to shocks in both rates. From the results in Table 3, however, the effective federal funds rate also responds to shocks in the fifteen-year mortgage interest rate and the Aaa rate. In all, the effective federal funds rate responds in the long run to shocks in ten out of the eleven given interest rates. This result is consistent with Sarno and Thornton (2003) who find that with respect to cointegration between the effective federal funds rate and the 3- month Treasury bill rate, that the effective federal funds rate appears to do more of the adjusting towards long-run equilibrium. They propose that this phenomenon may be due to the Treasury bill acting preemptively to expected changes in the effective federal funds rate, making the effective federal funds rate appear to be reacting to changes in the Treasury bill rate.

The potential for structural breaks exists in our sample period. If a break exists, then the break may impact the adjustment parameters estimated in Table 3. To address this concern, the Quandt-Andrews test is utilized (Zivot and Andrews, 1992). Specifically, the Quandt-Andrews test is applied to the error-correction model for each given interest rate. Table 4 provides the results of the Quandt-Andrews tests. The results fail to reject the null of parameter constancy for six of the eleven error-correction models. The tests reject the null on long-run adjustment after the indicated break in the error-correction models where the dependent variable is the change in the one-year treasury rate, the change in the sixty-day commercial paper rate, the change in the LIBOR rate, the change in the fifteen-year mortgage rate, and the change in the Aaa rate.

Taking the results as a whole, however, they tend to support the findings in Table 3. More than half of the error-correction models for the given interest rate exhibit parameter constancy. For the error-correction models with adjustment parameter instability, only two of the interest rates were found to be endogenous in the long run in Table 3 and two are found to be weakly exogenous in the long run after the indicated break date. These results match those found in the full-sample results in Table 3. In only one case do the adjustment results after the break contradict the results in Table 3. The results suggest that the one-year treasury rate is endogenous in the long run after the indicated break date. Although the adjustment parameter in the one-year treasury error-correction model is marginally insignificant, the parameter itself is economically significant. The value of the parameter more than doubles after the indicated break date, and the result suggests that this rate adjusts to shocks in either rate in the long run.

Table 4: Stability Test Results

	Quandt-Andrews Test Results			Post-Break Adjustment Parameters for Given Interest Rate		
Variable	**Break Point**	**Max F-Statistic**	**p-value**	**Adjustment Parameter**	**Standard Error**	**p-value**
Three-Month Treasury	Jan., 1988	2.02	0.201	—	—	—
One-Year Treasury	Feb., 1989	2.85	0.010	0.071	0.039	0.068
Two-Year Treasury	Mar., 1990	2.20	0.117	—	—	—
Five-Year Treasury	Oct., 1996	1.98	0.231	—	—	—
Seven-Year Treasury	Oct., 1996	1.91	0.280	—	—	—
Ten-Year Treasury	May, 1996	1.79	0.378	—	—	—
Sixty-Day Commercial Paper	Mar., 2003	4.60	0.000	0.020	0.100	0.843
Twelve-Month LIBOR	Aug., 1992	2.61	0.000	-0.035	0.051	0.490
Fifteen-Year Mortgage	Jun., 2007	2.53	0.008	-0.034	0.017	0.038
Thirty-Year Mortgage	Jan., 1988	2.03	0.201	—	—	—
Aaa Bond	Nov., 2008	6.30	0.000	-0.127	0.037	0.001

Conclusion

In order to more fully understand the transmission of monetary policy to the macro-economy through interest rate targeting, we examine the magnitude and timing of interest rate pass-through from the effective federal funds rate to various primarily U.S. short-term and long-term interest rates. While we find that each of the eleven rates studied is cointegrated with the effective federal funds rate, only seven of the eleven exhibit complete pass-through in the long run. These mixed results are in line with the findings of other broad studies of interest rates, most of which examine non-U.S. rates. We also examine the pattern of adjustment to this long-run equilibrium relationship and find that in most cases, the effective federal funds rate is doing the adjusting. In eight out of the eleven cases, or seven out of eleven cases when including a potential structural break, the adjustment parameters show that the effective federal funds rate is the only rate responding to a disequilibrium shock to restore long-run equilibrium. Overall, in ten of the eleven cases, the effective federal funds rate is doing at least some of the adjustment. This runs counter-intuitively to the idea that policy changes are being passed through to retail interest rates which implies that the retail rates should be adjusting to a policy rate change. This warrants further study to explain this timing phenomenon.

References

Becker, Ralf, Denise R. Osborn and Dilem Yildirim. 2012. "A Threshold Cointegration Analysis of Interest Rate Pass-Through to UK Mortgage Rates." *Economic Modelling* 29: 2504-2513.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock. 1996. "Efficient Tests for an Autoregressive Unit Root." *Econometrica* 64: 813-836.

Harbo Hansen, Niels-Jakob. and Peter Welz. 2011. "Interest Rate Pass-through During the Global Financial Crisis: The Case of Sweden." *OECD Economics Department Working Papers* No. 855.

Hegwood, Natalie D. and M. H. Tuttle. 2017. "Conventional Mortgage Interest Rate and the Effective Federal Funds Rate Pass-Through." *Journal of Business Strategies* 34: 57-74.

Hofmann, Boris and Paul Mizen. 2004. "Interest Rate Pass-Through and Monetary Transmission: Evidence from Individual Financial Institutions' Retail Rates." *Economica* 71: 99-123.

Johansen, Soren. 1992. "Determination of Cointegration Rank in the Presence of a Linear Trend." *Oxford Bulletin of Economics and Statistics* 54: 383-397.

Johansen, Soren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. OUP Catalogue, Oxford University Press.

Ng, Serena and Pierre Perron. 2001. "Lag Length Selection and the Construction of the Unit Root Tests with Good Size and Power." *Econometrica* 69: 1519-1554.

Payne, James. 2006. "The Response of the Conventional Mortgage Interest Rate to the Federal Funds Rate: Symmetric or Asymmetric Adjustment?" *Applied Financial Economics Letters* 2: 279-284.

Payne, James E. and George A. Waters. 2008. "Interest Rate Pass Through and Asymmetric Adjustment: Evidence from the Federal Funds Rate Operating Target Period." *Applied Economics* 40: 1355–1362.

Sander, Harald and Stephanie Kleimeier, 2004. "Convergence in Euro-Zone Retail Banking? What Interest Rate Pass-Through Tells us about Monetary Policy Transmission, Competition and Integration." *Journal of International Money and Finance* 23: 461-492.

Sander, Harald and Stephanie Kleimeier. 2006. "Convergence of Interest Rate Pass-Through in a Wider Euro Zone?" *Economic Systems* 30: pp. 405-423.

Sarno, Lucio and Daniel L. Thornton. 2003. "The Dynamic Relationship between the Federal Funds Rate and the Treasury Bill Rate: An Empirical Investigation." *Journal of Banking and Finance* 27: 1079-1110.

Sellon, Gordon H. 2002. "The Changing U.S. Financial System: Some Implications for the Monetary Transmission Mechanism." Federal Reserve Bank of Kansas City, *Economic Review*, 1st Quarter: 5-35.

Stiglitz, Joseph E. and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71: 393-410.

Zivot, Eric and Donald Andrews. 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.

Capitalization of Public Perception in MLB Player Salaries

Gayla Nicole Herbert and Jason Beck, Georgia Southern University
Casey Brasher, Pfeiffer University

Abstract

Typical human capital models for professional athlete salaries usually focus on on-field performance measures. We explore the possibility that fan interest in a player could also be capitalized by a player through a higher salary. We use Google search intensity as a measure of fan interest in MLB free agents and find evidence that higher search intensity is correlated with higher salaries, controlling for measurable performance factors.

Introduction

Every year scores of professional baseball players enter free-agency as a direct result of the 1976 collectively bargained agreement between the Major League Baseball Player's Association (Player's Association, henceforth) and team owners. This agreement awarded *free-agency* to all players with six or more years of service in Major League Baseball (MLB) after the 1976 season.[1] As a result of competition between teams, the salaries of these MLB players have converged much closer to the players' marginal revenue products (MRP).

Scully's 1974 seminal article was one of the first forays into the realm of sports and economics and examined the compensation of MLB players in the context of the monopsonistic nature of the MLB labor market of the time. Over the past 50 years, Scully's influential paper has prompted a large literature examining the effects of on-field performance on player pay. Work in this area, however, often overlooks the fact that ultimately professional athletes are entertainers and thus their value to a profit maximizing team is derivative of their ability to generate revenue. While better on-field performance will translate into more team wins and thus will increase revenue, there may be factors beyond what would strictly be considered on-field performance measures that fans (i.e. customers) value. For instance, it is possible that two players with similar measureable performance could equally impact a team's likelihood of winning games, but nonetheless attract different levels of interest from fans. In this way, they could have different impacts on team marginal revenue by generating different levels of viewership, ticket sales, etc. If higher levels of fan interest translate into a larger marginal impact on revenue, it is possible that players may be able to capitalize on this through higher salaries.

This paper follows in Scully's (and others') footsteps by using the wealth of readily available micro-level data on MLB players to explore the influential factors on player pay. We augment this with the inclusion of a measure of fan interest in the player. Major League Baseball is particularly well-suited for examining such compensation issues because it is the only major American sports league that does not institute a set salary cap. Thus, MLB teams are free from the exogenous player pay restrictions that other leagues are often subject to.[2]

This paper proceeds as follows: First, we examine the literature as it pertains to player compensation. Next, we discuss the data and outline an empirical model of player compensation in MLB. Next, we discuss the results of *nonperformance measures* on the compensation of free-agent MLB players. Finally, we conclude with closing remarks regarding this and future studies.

Literature

The public availability of data has made Major League Baseball (MLB) a topic of research in Sports Economics for decades. In particular, Scully's 1974 article, with his formal modeling and estimation of MLB players' MRP, marked the beginnings of a considerable amount of research on the MLB labor market. Since Scully, researchers have continued exploring the MLB labor market by examining such issues as pay discrimination, imperfect competition (monopsony), and the impact of player performance statistics on player compensation.

A number of studies have explored compensation in MLB by comparing player salaries to estimates of players' MRP, with the idea being that in a competitive labor market player salaries would approximate players' MRP. Differences between the MRP of players and player salaries has sometimes been defined in the literature as salary exploitation—arising due to the three-tiered structure of the MLB labor market which distorts a free labor market and reduces competition for player services. Estimating a recursive system of equations, Scully (1974) found that MLB players were paid salaries a mere 10-20 percent of their MRP, providing the first statistical evidence of team owners' ability to suppress the salaries of MLB players and extract the lion's share of rents accruing from a monopsony. Arguing that the Scully model is likely simultaneous and using a two-stage least squares approach, Medoff (1976) likewise finds evidence of the salary exploitation of players, though his estimates suggest MLB players receive salaries in the range of 40-50 percent of their MRP.

It should be noted that both Scully and Medoff use data prior to the 1976 season, when the reverse clause affected player contracts and free-agency was contractually prohibited. Studies after the 1976 season—the season in which free-agency was court-awarded to all MLB players and, subsequently, agreed to by players and team owners for players with six or more years of service in MLB through collective bargaining—typically find that free-agency reduced monopsony rents accruing to team owners an led to salaries more in-line with players' MRP (Sommers and Quinton, 1982; Raimondo, 1983). However, studies on salary exploitation after the 1976 season continued to find that team owners were still able to exert monopsonistic power over reverse clause players, with non-free agent players receiving salaries in the range of the prior estimates of MRP by Scully and Medoff.

Krautmann (1999) takes a free market return approach to estimating players' MRPs by modeling MRP as a function of labor inputs employed by profit-maximizing firms. Using a sample of free agent players only, Krautmann first estimates a salary equation that includes key player performance measures. The estimated regression coefficients in the salary equation are assumed to yield the incremental competitive market reward for each measures of player performance included in the salary equation. Assuming these performance measures are paid a wage rate equal to the measure's MRP, and the relevant performance measure are controlled for in the salary equation, the predicted salaries of the sample of free agent players should approximate free agent players' MRP. Krautmann thus finds no salary exploitation of free-agent players, and much smaller estimates of the exploitation of reverse clause player by a factor of 4 or 5 compared to estimates generated by the Scully model.

Other studies examine factors omitted from previous estimates of the Scully model and likely to cause Krautmann's competitive market return model to overstate or understate MRP. In particular, Mullin and Dunn (2002) add to the literature by including "star quality"—defined as some tangible or intangible characteristic relating to a player that causes team owners and fans to place a unique value on a specific player—in the estimation of MRP. To capture the effect of star quality, Mullin and Dunn use the current market value of each individual player's baseball card as a proxy for star quality. Estimating a model in the framework proposed by Scully, the authors find star players in MLB to be significantly underpaid relative to estimates of star players' MRP. Humphreys, Simmons, and Soebbing (2014) find a positive correlation between a National Football League player's salary and being arrested. In another study, but this time examining market size, Burger and Walters (2003) find that larger market teams tend to derive 5 to 6 times more marginal revenue per win compared to teams in the smallest markets, with the potential for revenues per win to be particularly pronounced for teams in postseason contention. Therefore, creating the incentive for the largest market teams to pay more for players whose performance generates more wins. An important implication of these findings is that free agent player salaries depend not only on the number of teams bidding for the player's services, but also on the market size of the teams bidding for the free agents services; thereby, creating a situation in which the large market teams need not pay salaries equal to the free agent players' MRP if rival bidders are small market teams.

There has been considerable attention paid to the performance-related determinants of MLB player salaries. Inspired by the Moneyball hypothesis outlined by Lewis (2003), Hakes and Sauer (2006) estimate a series of wage equations in an effort to test the efficiency of the MLB labor market during the period 2000-2004. In particular, they use on-base percentage and slugging percentage as explanatory variables in both a team success equation and a wage equation, arguing that if the MLB labor market is efficiently operating, the measure contributing most to team success should also be the measure incrementally rewarded the most. Although both performance measures are positive and statistically significant in each regression, on-base percentage is found to contribute more to teams' win percentage by a factor of approximately 2.5 per one-point increase, yet on-base percentage is incrementally paid far less relative to slugging percentage. This provides evidence of teams mispricing player skill and an inefficient labor market during the period studied. Other studies similar in nature to Hakes and Sauer have found MLB player salaries to be significantly-related to fielding percentage (Kahn, 1993), position (Wallace, 1988), consistency and durability (Stone and Pantuosco, 2008), contract length (Krautmann and Oppenheimer, 2002), and race (Johnson, 1992).

Terry, McGee, and Kass (2018) take a look at MLB player salaries by exploring the relationship between contract premiums and factors not related to a players on-field performance. The authors measure contract premiums as the difference between the salary received by free agent players and free agent players' MRP in the season going into their contract year. Employing a random effects model that regresses contract premiums on select player performance and non-performance factors, the authors find several factors not related to players' on field performance play a statistically significant role in the salary determination of free agent players in MLB. In particular, their findings suggest that players playing in a larger MLB market during their contract year receive a contract premium, as well as players playing on more successful teams. Additionally, teams that are relatively more financially successful, teams in playoff contention the previous season, and multiple teams bidding for the services of free agent players lead to contract premiums for free agent players in MLB.

Data and Model

MLB free agents were identified via the ESPN Free Agent Tracker website for years 2015, 2016, and 2017.[3] Since salary

information is required for the analysis, the data were restricted to players that were signed by a major league team by the first day of the following season. Since pitchers and position players cannot be directly compared in a hedonic salary model (i.e. statistics like "earned run average" would not apply to a right fielder), the observations were restricted to only include position players (infielders, outfields, and catchers).

The basic model includes player characteristics such as age (and age squared), home runs per 100 at bats, average number of stolen bases per year, on-base percentage, and controls for infield, outfield, catcher, and designated hitter. Note that some available variables were omitted due to collinearly. For instance, players' battering averages are often highly correlated with on-base percentage and thus excluded. Career on-field statistics, as well as contract information were collected from the ESPN website.

To measure the level of player popularity, and index was constructed using Google Trends. Unfortunately, Google Trends does not provide raw, level data, such as the number of searches on a specific term (in our case, a player) over a specified time frame. Instead, it allows a user to compare intensity of search on up to five specific search terms relative to each other, over a user-specified time period. These results are normalized, 0-100, with 100 being the highest single day level of search intensity for any of the five search terms in the query. Each search term's score is an average over the time frame relative to the single day peak of any of the included search terms. As such, one five-player set of scores would not be directly comparable to a completely different five-player set of scores since they would be normalized to a different single-day peak. However, by using one player as a reference, and leaving that player in all searches, a relative measure can be created. Each player's score is then a relative index, not directly interpretable, but hopefully useful as a relative measure of Google search intensity. Since each set of scores for each of the three years is relative to a reference player from that year, this creates difficulty in cross-year comparisons across the three MLB seasons (2015-2017) as the scale is different in each year. Thus, three separate models are initially estimated, one for each year. A pooled model is also estimated, though interpretation on the search intensity variable is difficult.

One potential problem with the above methodology is that ambiguity of search terms may introduce noise to the measure. For instance, one player in our data is named Carlos Santana. As this player shares a name with an internationally known musician and recording artist, it might be assumed that the Carlos Santana in our data set might have an inflated popularity score as the search term "Carlos Santana" may pick up searches for both individuals. However, when constructing the query, Google Trends is able to create a distinction between the two Santanas, labeling one as "Carlos Santana- musician" and the other as "Carlos Santana- baseball player". While it is difficult to say how effective this sorting process is, there appears to be at least some evidence that Google is able to distinguish between the two in that the baseball playing Carlos Santana has a much lower search index score than the musician Carlos Santana and this score is roughly in line with the scores of other MLB players of similar performance and salary. Note that every player name entered into the query was recognized by Google Trends as an MLB player and thus, ostensibly at least, the scores used are not generated simply on the player's name, generally defined, as a search term.

Table 1: Summary Statistics

	2015		2016		2017	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
salary	7751163	6627045	7213121	5975237	6449755	6449755
age	33.232	2.524	32.574	2.902	31.970	2.276
age squared	1110.628	170.580	1069.340	186.374	1027.147	150.961
HR per 100AB	2.976	1.346	3.489	1.354	3.228	1.345
SB per year	7.955	6.936	5.267	7.349	8.418	7.931
OBP	326.651	22.874	321	28.980	326.058	20.231
outfielder	0.395	0.494	0.340	0.478	0.352	0.485
catcher	0.186	0.393	0.234	0.427	0.176	0.386
adjusted_score	4.161	10.038	1.702	2.441	3.105	2.645
n=	43		47		34	

The following model is estimated via OLS with robust standard errors four times: once for each of the year 2015, 2016, and 2017 free agents, and then once as a pooled sample.

$Ln(salary_i)= f(C_i, P_i, G_i)$

Where:
$Ln(salary_i)$ is the natural log of the player's one-year salary
C_i= a vector of characteristics for player i
P_i= a vector of performance characteristics for player i
G_i= fan interest for player i as measured by Google search intensity

Results

Table 2: OLS Results with Robust Standard Errors

	Dep.Var:	ln(salary)						
	2015		2016		2017		Pooled	
Variables	Coef	P-val	Coef	P-val	Coef	P-val	Coef	P-val
constant	25.279	0.002	3.945	0.433	-15.896	3.623	6.642	0.240
age	-0.719	0.093	0.594	0.104	1.745	0.087	0.340	0.349
age squared	0.009	0.125	-0.104	0.084	-0.029	0.056	-0.007	0.238
HR per 100AB	0.237	0.010	0.176	0.082	0.115	0.318	0.236	0.00
SB per year	0.211	0.159	0.039	0.010	-0.023	0.261	0.021	0.068
OBP	0.071	0.032	0.069	0.097	0.133	0.530	11.823	0.00
outfielder	0.722	0.001	-0.022	0.934	0.150	0.523	0.3037	0.046
catcher	0.219	0.486	-0.007	0.980	0.007	0.984	0.117	0.517
year_2016							-0.303	0.049
year_2017							-0.052	0.704
search_intensity	0.255	0.00	0.14	0.00	0.084	0.113	0.025	0.007
n=	43		47		34		124	
F-stat	18.233		22.65		9.47		14.59	
Prob>F	0.00		0.00		0.00		0.00	
R squared	0.652		0.524		0.592		0.470	
Adjusted R^2	0.571		0.434		0.481		0.424	

Coefficients on control variables largely match expectations and are in line with other research. Age of the player appears to increase salary at a decreasing rate, and higher offensive characteristics, such as homeruns per 100 at bats, and on base percentage, are associated with higher salaries. T-stats are somewhat lower than hedonic wage models in other work, but this could be due to the relatively small sample size for each of the three yearly samples. The forth model is a single regression using all three years of data. Note that the variable of interest, *search_intensity*, is scaled differently in each year and thus difficult to interpret.

The variable of interest, *search_intensity*, was positive in each of the four models and statistically significant in the 2015, 2016, and pooled specifications. Since the unit of measurement of the *search_intensity* variable is a relative measure of an index, it is not directly interpretable but its positive sign and statistical significance in three of the four models provides at least some justification for the inclusion of non-performance measures in these types of human capital models. These results suggest that, controlling for performance, higher search intensity is associated with higher salary and may have explanatory power when estimating wages. Note that the inclusion of the *search_intensity* variable into the pooled model increase the adjusted R-squared from 0.396 to 0.424 further suggesting that this variable has explanatory power.

Conclusion

Many previous studies have explored the relationship between salaries of professional athletes and their measurable performance statistics. We hypothesize these may be incomplete because they typically omit nonperformance attributes that

may influence revenues of their teams. If these non-performance measures indeed impact player salaries, then estimates from models where they are omitted should be interpreted with caution given these estimates may suffer from omitted variable bias.

To try and account for this, we develop a human capital model that includes a non-performance measure in a model of MLB free agent players and explore the possibility that players are able to capitalize on their off-field recognition. We use a measure of Google search intensity as an independent variable and find that while controlling for traditional performance measures, Google search intensity was positively associated with higher player salary and had predictive power on earnings in two of the three estimated yearly models, as well as the pooled model. This suggests that there is at least some evidence that off-field fan interest may influence player salary and that this could be an issue for further research.

One limitation of this study lies in the fact that non-performance fan interest measures are difficult to measure. Here, relative Google search intensity is used but future research may be able to utilize a better measure. Additionally, other professional sports leagues could be examined in a similar fashion to see if this effect is observable in other sports.

Notes

1. Prior to 1976, all player contracts specified that a player belonged, in perpetuity, to the team, unless that player was traded or sold to another team. After 1976, only players with fewer than 6 years of service were subject to the reverse clause.
2. Despite no set salary cap, teams can be charged a luxury tax for excessive spending.
3. http://www.espn.com/mlb/freeagents

References

Burger, John D., and Stephen JK Walters. 2003. "Market size, pay, and performance: A general model and application to Major League Baseball." *Journal of Sports Economics,* Vol. 4, No. 2, pp. 108-125.

Fizel, John. *Handbook of sports economics research*. Routledge, 2017.

Hakes, Jahn K., and Raymond D. Sauer. 2006. "An economic evaluation of the Moneyball hypothesis." *Journal of Economic Perspectives,* Vol. 20, No. 3, pp. 173-186.

Humphreys, Brad R., Rob Simmons, and Brian P. Soebbing. 2014. "Does crime pay? The effect of off-field behavior on player salaries in the National Football League." *Fox School of Business Research Paper,* Vol. 16, No. 5

Johnson, Bruce. 1992. "Team racial composition and players' salaries." *Diamonds are forever: Business of baseball*: 189-202.

Kahn, Lawrence M. 1993. "Free agency, long-term contracts and compensation in Major League Baseball: Estimates from panel data." *The Review of Economics and Statistics,* Vol. 75, No.1, pp. 157-164.

Krautmann, Anthony C. and Margaret Oppenheimer. 2002. "Contract length and the return to performance in Major League Baseball." *Journal of Sports Economics,* Vol.3, No. 1, pp. 6-17.

Krautmann, A. C., & Oppenheimer, M. 2002. Contract length and the return to performance in Major League Baseball. *Journal of Sports Economics*, Vol. 3, No. 1, pp. 6–17.

Lewis, Michael. *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004.

Medoff, Marshall H. 1976. "Monopsonistic exploitation in professional baseball." *Quarterly Review of Economics and Business,* Vol. 16, No. 2, pp. 113-121.

Mullin, Charles J., and Lucia F. Dunn. 2002. "Using baseball card prices to measure star quality and monopsony." *Economic Inquiry,* Vol. 40, No. 4, pp. 620-632.

Treme, Julianne, and Samuel K. Allen. 2011. "Press pass: Payoffs to media exposure among National Football League (NFL) wide receivers." *Journal of Sports Economics,* Vol. 12, No. 3, pp. 370-390.

Raimondo, Henry J. 1983. "Free agents' impact on the labor market for baseball players." *Journal of Labor Research,* Vol. 4, No. 2. pp. 183-193.

Scully, Gerald W. 1974. "Pay and performance in major league baseball." *The American Economic Review,* 64: 6, pp. 915-930.

Sommers, Paul M., and Noel Quinton. 1982. "Pay and performance in major league baseball: The case of the first family of free agents." *The Journal of Human Resources,* Vol. 17, No. 3, pp. 426-436.

Stone, Gary, and Louis J. Pantuosco. 2008. "Estimating baseball salary equations from 1961-2005: A look at changes in major league compensation." *International Journal of Sport Finance,* Vol. 3, No. 4, pp. 228-238.

Terry, Ryan P., Jeffrey E. McGee, and Malcolm J. Kass. 2018. "The not-so-free agent: Non-performance factors that contribute to free agent compensation premiums." *Sport Management Review,* Vol. 21, No. 2, pp. 189-201.

Wallace, Michael. 1988. "Labor market structure and salary determination among professional basketball players." *Work and Occupations* Vol. 15, No. 3, pp. 294-312.

Zimbalist, Andrew. 1992. "Salaries and Performance: Beyond the Scully Model." *Diamonds are forever: The business of baseball*: 109-133.

An Analysis of the High-performing Property Asset Manager in the Multi-family Real Estate Industry, According to Executives

Christine A. Jubelt, Rollins College

Abstract

Responding to investor demand for high-performing real estate assets and the critical hardship for highly qualified Property Asset Managers in the residential multi-family real estate industry, research was conducted to identify and rank order of importance unique competencies surrounding high-level performance Property Asset Managers in high-performing work settings. A three-step research methodology was used to collect data. The analyzed data identified executive's views of 23 competencies and definitions of competencies of the high-level performance of the exemplary Property Asset Manager. In addition, the relative importance of the 23 recognized competencies of high-level performance Property Asset Managers was identified. Limited research is available related to competencies of the Property Asset Manager. This research study provides a comprehensive listing of additional competencies (honesty and integrity, clear communication, efficiency, passion, proactive, delegates effectively, confident, financial acumen, recruiter, marketing knowledgeable, creativity, adaptability, education focused, develops employees and employee development, being entrepreneurial, emotional intelligence, team focused, loyalty, inspiring, a visionary, and involved), thereby complementing recent literature and adding to existing knowledge in the area of managerial competencies.

Introduction

This exploratory research deliberates the perception of executives regarding the competencies required of Property Asset Managers in High-performing Work Settings (HPWS) within the residential multi-family real estate industry in the U.S. The purposes of this study include: identifying unique competencies held by high-level performance Property Asset Managers in the view of executives in the industry, and exploring the relative importance of the identified competencies of high-level performance Property Asset Managers in views of the executives in the industry in order to thereby extend knowledge about competency modeling and the identified competencies in the multi-family real estate management industry. In addition, there is very little research in this area. This research related to competencies contributes to the limited research available.

This research provides not only benefits to the multi-family industry operations, but provides a clear path to stronger financial performance. According to Lawler (1994) the main purpose for this approach is "there is only one compelling reason for adopting a skill or competency-based approach: to create a competitive advantage, because it leads to an organization performing better" (p. 12). That is, "ultimately the best competitive advantage in global markets is the performance capability of the organization" (Lawler, 1994, p. 12).

The Property Asset Manager is the central operative "consistently associated with the day-to-day operations and maintenance of the property(s) and its people – site personnel and tenants/residents" (Fields, 2015, p. 45). Regarding the industry as a whole, over the past few years there has been a change in the role of the Property Asset Manager, "real estate management has probably seen the greatest shift in the core responsibilities and expectations of Property Managers in recent years" (Fields, 2015, p. 45). This indicates "the industry requires a more sophisticated representative to mitigate risks, embrace technology, build successful relationships, and ultimately understand financial implications to affect change at the site-level" (Fields, 2015, p. 45). The Property Asset Manager operates "in a business environment that favors flexibility and quick reactions to market conditions, a more financially adept Property Manager provides a competitive advantage" (Fields, 2015, p. 46).

There is a need for exemplary performance of Property Asset Managers to operate residential communities in the multi-family housing industry, which:

> serves a vital role in the real estate market place as one of four households in the United States live in multi-family homes (http://www.nahb.org). Many desirable features of multi-family housing as well as changing demographics have exacerbated the popularity of multi-family housing as a housing choice. (Zietz, 2003, p. 185)

The changing role of Property Asset Managers is also related to changes in the reasons individuals seek multi-family housing versus home ownership. More recently, Americans enjoy a busy lifestyle; they desire simpler more convenient means to live, which reduce the typical homeowner responsibilities, such as maintenance, lawn care, home care, property taxes, and other

homeowner expenses. In addition, the workforce is more mobile today than in the past, providing more incentives for individuals transferring to new locations to seek multi-family housing. (Zietz, 2003, p. 185).

The basic organizational structure of real estate multi-family management organizations is represented in Figure below. The organizational structure encompasses four levels. These are the strategy (C-level), functional, mid-level manager, and supervisor levels. The Property Asset Manager in this research study was operating at the supervisor level. The identified proposition and research questions, related to this study, were as follows: What were the unique competencies held by high-level performance Property Asset Managers in the view of executives in the industry? What was the relative importance of the identified competencies for high-level performance Property Asset Managers in the views of executives in the industry? Overall, these research questions explore the competencies, of high-performing Property Asset Managers from the viewpoint of executives in the industry.

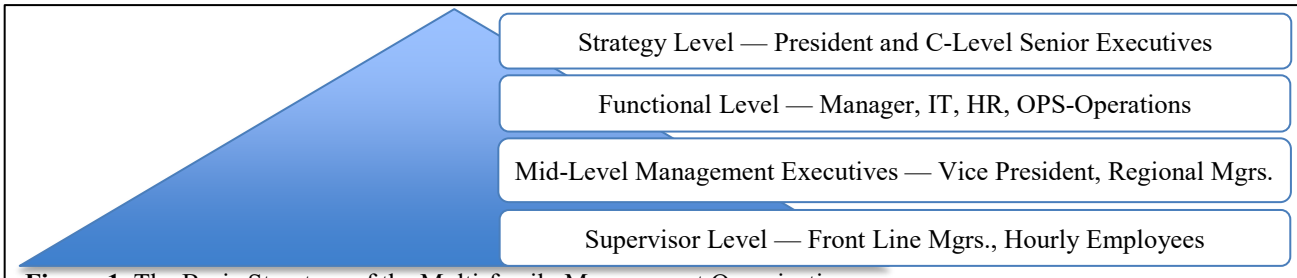


Figure 1. The Basic Structure of the Multi-family Management Organization.

One intended benefit from this study was to identify competencies, as these "competencies communicate to workers what tools they need in order to demonstrate effective performance. If we study exemplary performers in the organization, we can replicate what they do" (Rothwell, Graber, & Dubois, 2015a, pp. 1-63). Competencies are a valuable constituent of stratagem intended for profitable real estate multi-family asset performance. By conducting research identifying these competencies, the intent was to provide valuable and relevant information to the multi-family industry conducive to hiring, developing, training, and setting performance standards for Property Asset Managers who provide high-level performance results. In addition, research that identifies competencies may drive the design of industry certification program mechanisms currently in place in the United States.

The exemplary performer drives results in the multi-family real estate industry, a service industry. In the HPWS, a service-quality-oriented HPWS influences individual service performance, in part, through the mediation of employee human capital, psychological empowerment, and perceived organizational support, and highlights the importance of incorporating employee perspectives into the examination of HPWS (Liao, Toya, Lepak, & Hong, 2009, p. 388).

The exemplary performing manager in the HPWS then provides a higher quality level of service leading to higher customer satisfaction. In addition, "competencies contribute to individual exemplary performance that creates reasonable impact on business outcomes" (Society for Human Resource Management, 2012, p. 5). Another benefit from this research is providing relevant information to the multi-family real estate industry. In particular, the results would provide desirable competencies of high-performing managers. Residential firms could use this information to train, recruit, and develop managers who provide HPWS, as well as successful and profitable performance.

The conceptual basis surrounding this research referenced and recognized:

> high-performance work practices (HPWP; e.g., valid selection procedures, job-focused training, performance-based incentives and compensation, etc.) may serve a critical mediating role between HPWP and organizational outcomes, such as productivity, turnover, sales growth, and profitability" (as cited in Sikora, Ferris, & Van Iddekinge, 2015, p. 1908).

In addition, "two theoretical perspectives" (as cited in Sikora et al., 2015, p. 1909), suggest a "social context model highlights the links between organizational climate, HPWP and their implementation, employee attitudes and behaviors, and ultimately, organizational effectiveness" (Sikora et al., 2015, p. 1909). The authors go on to say: "by illustrating these sequential relationships, social context theory provides an overall context for understanding the important process of how line manager HPWP implementation efforts shape employee outcomes" (Sikora et al., 2015, p. 1909).

An additional perspective included Human Resource Management (HRM) that is, "within Ferris et al.'s overall context, Bowen and Ostroff (2004) provided a more specific understanding of how the 'strength of an HRM system' impacts the HPWP – firm performance relationship" (as cited in Sikora et al., 2015, p. 1909). The authors state "Bowen and Ostroff argued that contexts or climates that are 'strong' are those where individuals possess a common perception and interpretation of events and

shared understandings regarding expected behaviors and outcomes" (as cited in Sikora et al., 2015, p. 1909). In addition, the authors go on to incorporate the value of HR practices into the HPWP and:

> further suggested that employees' attitudinal and behavioral responses to HR practices are a function of only the practices and systems they perceive to exist. Thus, for HR practices to gain employee attention and to demonstrate intended effects on employee attitudes and behaviors, an organization's line managers need to make HR practices salient to employees. (Sikora et al., 2015, p. 1909)

The diagram below in Figure 2 is provided to facilitate the understanding of the research.

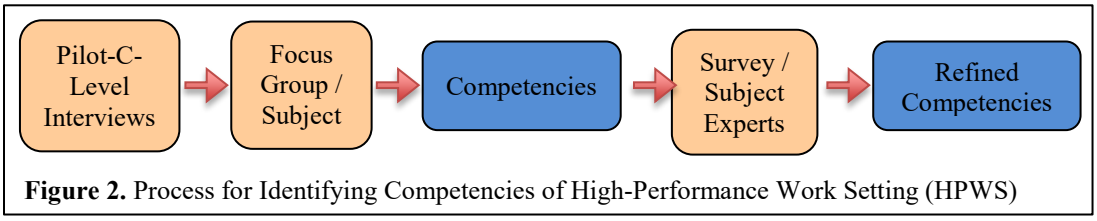


Figure 2. Process for Identifying Competencies of High-Performance Work Setting (HPWS)

To identify the refined competencies, the process included pilot (interviews), and focus groups with subject experts who determined and selected competencies. The researcher identified, refined, and ranked the most relevant competencies through a survey. Figure 3 below illustrates the process and research we are endeavored to investigate and study. The model explains the flow of the research to identify the refined competencies (the level of importance of the competencies and the Property Asset Manager) in the HPWS.

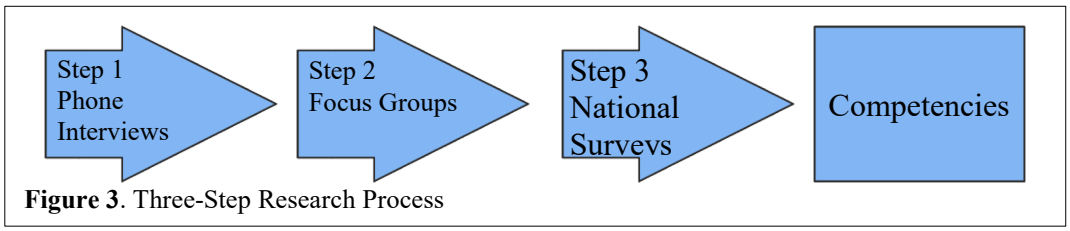


Figure 3. Three-Step Research Process

The derivative surrounding this research included the social context theory and the Resource Based View (RBV). The reason for using the RBV was that it supported the competency-based approach and competitive advantage. The competency-based approach offered a method to provide a competitive advantage for the firm. According to proponents of the RBV special types of competencies, usually called "core" competencies are recognized as, "sources of sustained competitive advantage, they are firm resources that are valuable, rare, imperfectly imitable, and nonsubstitutable" (Prahalad & Hamel, 1990, p. 7; Barney, 1991, p. 116).

The RBV was used to illustrate the competencies of the exemplary, high-performing Property Asset Manager in this study. The competencies, in essence, are unique to the exemplary Property Asset Manager, an asset difficult to imitate. The exemplary Property Asset Manager was considered an organizational asset and firm resource. This asset and firm resource was a rare, and difficult to imitate, resource providing an organization sustained competitive advantage. In addition, Barney (1991) states, "a resource based model of sustained competitive advantage anticipates a more intimate integration of the organizational and the economic as a way to study sustained competitive advantage" (p. 116). Notwithstanding, Barney (1991) further indicates, "indeed it may be the case that a manager or a managerial team is a firm resource that has the potential for generating sustained competitive advantages" (p. 117).

Research Process and Methodology

The process utilized in this research included both qualitative and quantitative methods. Quantitative data were analyzed using descriptive statistics. Qualitative data was attained through three different methodological research steps. Both qualitative and quantitative data facilitates triangulation to this research study providing data, findings, and conclusions that are rigorous and convincing. Triangulation of data from interviews, focus groups, and surveys, and two different research methods in this study provided "a fuller picture of what is happening...a qualitative study with quantitative data from a survey" (Meyers, 2013, p. 5). The qualitative research methodology was implemented as a preliminary approach to the quantitative study. The qualitative methodology in this research study was accomplished through gathering data and information and talking to

executives and subject experts about competencies. The overall intention of the research was to study exemplary, high-level performing Property Asset Managers.

The methodology in this research performed to identify competencies also consisted of information ascertained from literature, in addition to prior Certified Apartment Manager (CAM) research performed by this researcher in 2016. Information from literature identified and applied to the research methodology in this research study were Rothwell, Graber, and Dubois (2015a) and Marshall, Goebel, and Moncrief (2002). In addition, the literature review informed this research study by providing information for the executive interviews in Step 1.

Each competency identification presented variations in methodology and process due to the unique circumstances surrounding firms, and specific organizational issues and concerns (Rothwell et al., 2015a). Other researchers found out that "we are unable to provide you with a single, one size-fits all competency identification process of strategy (even by using competency lists we have provided) that you can apply to every competency identification project" (Rothwell et al., 2015a, pp. 2-18).

The overall methodology selected for this study facilitated an exploratory research study of competencies through both analysis and perceptions of executives. That is, this researcher first identified the competencies in Step 1 (from the perceptions of executives) through open-ended interviews. Then, further identification of competencies was accomplished in Step 2, the focus groups. Step 3 was implemented to evaluate these identified competencies through a survey of national-based subject experts. Following this, the researcher reviewed the qualitative and descriptive data compiled from all three steps and the resultant survey, and determined whether there is agreement with the initial Step 1 survey, and determined the level of importance of each competency. The 3-step research process was utilized as seen in Figure 3 below.

Step 1 Research- Interviews

Step 1, the interviews, was accomplished by an open-ended telephone interview performed on 10 C-level industry executives. The interviews were shaped from knowledge assimilated from the literature review and the researcher's industry experience. The purpose and anticipated outcome of these initial interviews was to learn from the executives their perceptions of the competencies of the exemplary manager and provide answers to recognized research questions.

The anticipated outcome of the interview included identifying the executive's perceptions of competencies of the Property Asset Manager and informing the C-level executives of the research being performed. This was accomplished through interviews with the executives from multi-family real estate management firms selected from a national listing of industry executives provided by the National Apartment Association. The executives function at the strategic level with an operations backgrounds, the executive titles included Chief Operating Officer and Chief Executive Officer, although the title varied based on the organizational structure of each firm. The National Apartment Association Education Institute (NAAEI) provided the list of C-Level executives. Executives were selected from this list; each has over 10 years executive-level experience in managing multi-family real estate investments. The selected interviewees were rich in experience to expedite the process; they were recruited from the industry.

Executives were asked identical open-ended questions queried in the same order during the telephone interview. The open-ended question format at the initial stage of this research was intended to provide consistency, yet aimed to attain results with richness and intensity to deliver valid data and results relevant to the industry. The intent was to explore the executive's perceptions through qualitative research, that is, "qualitative researchers argue that if you want to understand people's motivations, their reason, their actions, and the context for their beliefs and actions in an in-depth way" (Meyers, 2013, p. 5). That is, at minimum, the interview will lend face validity to the research study. This means the interviews "show that you did indeed talk to some of the key people" (Meyers, 2013, p. 242) within the industry, and ascertained executives' perceptions. This not only occurred in the executive interview in Step 1, but also applied to the focus group activity in Step 2, discussed in depth below.

Step 2 Research – Focus Groups

This national study held a focus group in three locations in the U.S. The participants of each focus group were industry experts (supervisors of exemplary, high performers). The industry or subject experts were selected utilizing a national listing of Regional Vice Presidents and Regional Managers provided by the NAAEI. Six to ten subject experts were selected to participate in each of the focus groups. One focus group was held in Atlanta, GA, another in Dallas, TX, and the third in Seattle, WA. By holding focus groups in three different regions of the U.S., we provided a selected sample of subject experts providing results demonstrating a broad sample population the U.S. in an attempt to assimilate and better represent a national population. This approach was adapted from a similar national research study "of sales executives were held in three geographically diverse locations throughout the U.S." (Marshall, Goebel, & Moncrief, 2002, p. 247).

The focus groups in Step 2 each consisted of subject experts who worked collaboratively with each other and the facilitator to identify competencies. The facilitator utilized the data collected from the focus groups and executive interviews to identify competencies of the exemplary Property Asset Manager. These identified competencies were evaluated in a survey at Step 3. To reinforce methodology in this research study, especially utilization of the survey and focus group, Rothwell et al. (2015a) says that:

> since circumstances influence your options for competency identification, we believe that you should begin by using a survey approach—possibly coupled with first using a competency card-sort activity, or a disciplined and well-planned focus group activity to identify a high-quality, critical mass of personal functioning competencies. (pp. 2-37)

In the focus groups, this researcher utilized the Rapid Results Assessment Method (RRAM) to determine the competencies and tasks performed by the Property Asset Manager, as well as the strength of attributes in relation to the position (high-performing Property Asset Manager). The Rapid Results Assessment Method "has proven to be a very effective method of quickly determining, at relatively low cost, the tasks that must be performed by persons employed in a given job or occupational area" (Rothwell et al., 2015a, pp. 2-91).

From the task, the competency was identified. The task underwent analysis to determine the specific skills, knowledge, and attitudes the worker, in this case the Property Asset Manager, requires (Rothwell et al., 2015b, pp. 2-91). The industry experts formed a committee, or focus group, and worked under the guidance of a facilitator. In this study, the specific purpose was to identify competencies specific to the exemplary Property Asset Manager in a high-performance setting. The RRAM in this study was performed by the focus group through the leader or facilitator.

The focus groups' process in this study also utilized information and input from the executives' interviews in Step 1. That is, the Step 1 results and content were intended to provide richer material intended for utilization in the specific questions presented to the focus group. The focus groups' course of action included utilizing a validated skills listing provided by the NAAEI. The firm Castle Worldwide and Dr. Jim Henderson validated this listing in 2015. The listing was explored with the focus group participants in Step 2, identifying the general areas of responsibility of the Property Asset Manager. The facilitator reviewed the listing or categories of key activities with the focus group to provide all focus group participants with identical information from the beginning to complete their common understanding of the Property Asset Manager job. The entire listing was provided to each focus group participant as a handout. Knowledge and skills are the best-known competencies used by employees to achieve outputs or results expected of them (Rothwell et al., 2015a, pp. 1-17). The skill listing contains skills required of a Property Asset Manager. This research study identified competencies of the exemplary, high-performing Property Asset Manager. The expected outcome of Step 2 was met; to provide a listing of competencies from subject experts of the high-performance setting of the Property Asset Manager. The results of Step 2 research produced a listing of competencies that was evaluated through a survey of subject experts implemented in Step 3 of the research.

Step 3: Research- Survey

Step 3 was comprised of a survey taken by industry experts, supervisors, and regional managers who are supervisors and generally hire the Property Asset Manager nationally. They were considered mid-level management executives in the property management industry. The intent was to validate the competencies identified in the focus groups and Step 2 of this research. Additionally, the survey was also used to establish the relative importance of the different competencies from the viewpoint of the respondents. The survey was applied to 1,500 subject experts utilizing a random sample of data from three sources provided by the NAAEI. The sample was randomized as we were attempting to infer what the so-called "universe" or population of subject experts or supervisors think. These three sources are the ALN Apartment Data Inc. database, containing information of 9,000 multi-family property management companies nationwide; the Certified Apartment Property Asset Managers (CAPS) database in the United States; and, the Membership listing of the National Apartment Association (NAA).

An advantage to this method of selection was that the listing of experts was large. A disadvantage is that the listing of NAA membership does not always list the title of the member. Some members have not listed their titles or titles have changed. The implications for this research are that the random sample will be derived from a database that contains less than 9,000 property management companies. To assure we identify the industry expert some were eliminated due to the lack of titles so that the face validity of the survey was increased.

The online survey was prepared and distributed utilizing Qualtrics and analyzed utilizing SPSS. The survey was constructed utilizing a 7-point Likert scale. The 7-point scale is comprised of the number 1 signifying "of no importance" through the number 7 signifying "of the utmost importance." In this research study, the Likert scale was applied to a situation actually rating the importance of a diverse set of competencies. In addition to rating the importance of each competency, respondents ranked the competencies from 1 to 23 based on their relative importance.

The intent of the research was to identify and provide a listing of competencies of Property Asset Managers in HPWS identified by executives or industry experts. The survey was intended to be constructed utilizing collected data attained from the focus groups in Step 2 research. The survey was developed based on coded qualitative data from the focus groups. The

NAAEI agreed to provide an incentive to any participants who completed the study and did so within two weeks. The NAAEI also requested demographic and educational data to be collected at the end of the survey. This includes age, gender, education level, and certifications.

Results

The results include the generation of competencies, response rate, demographic data, and results of statistical analyses. The findings include analyses to address the research proposition and questions explored in this research study. These are as follows: What are the unique competencies held by high-performance managers in the view of executives in the industry, and what is the relative importance of the identified competencies for high-performance Property Asset Managers in the view of executives in the industry? These research questions explore the desirable competencies of Property Asset Managers from the viewpoints of executives in the industry.

The identification of competencies of the exemplary Property Asset Manager took place through comprehensive, rigorous research methodology including two initial steps, followed by a survey questionnaire, a third step. Within the survey there were 23 items (competencies) measured on a 7-point Likert-type response, additionally, participants ranked the competencies from 1 to 23, based on their relative importance. The scale was analyzed using SPSS software and Clemson University Likert-type scale response anchors, Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation, and Tourism Management. Prior to launching the questionnaire, validation took place through pretesting of the instrument by subject experts and executives of the multi-family real estate industry.

A total of 150 usable survey questionnaires were returned, a response rate of 13%. That is, there were 1,168 email surveys opened, 250 started, and 150 were usable. The respondents were geographically diverse throughout the United States and represented approximately half female and half male respondents. The survey questionnaire was issued to 4,900 executives randomly selected from a list of 9,000 executives provided by the NAAEI. The 4,900 industry executives were dispersed across the United States. The usable surveys represented the geographic locations and general population derived from the NAAEI national listing of membership and conference attendance provided. The survey was distributed online via Qualtrics software, an identical letter of instruction was provided to each email. The ability to opt out of the survey was provided to each participant along with assurance that confidentiality would be preserved regarding his or her identity and response.

The demographic data of the survey respondents indicated there were 52.6% male respondents and 47.4% female respondents. The majority of respondents, 89.5%, were over 35 years old; 58.7% were between the ages of 35 and 54, while 31.3% were between the ages of 55 and 64, and 8% were over 65. The respondents were rich in experience, 95.4% had over 7 years' industry experience and 92.2% had over 10 years' industry experience

The executives that responded were dispersed throughout the U.S. Respondents located in the United States from the Northeast totaled 19.6%; the Southeast, 35.3%; Central, 22.9%; the Northwest, 3.3%; and the Southwest comprised 19%. The educational background of respondents varied from no college attendance to those respondents attaining higher professional degrees; 22.7% do not hold a college degree, 8.4% hold a 2-year college degree, 42.2% hold a 4-year college degree, 15.6% hold a master's degree, and 5.2% hold professional degree (JD or MD). Regarding industry certification, 69.5% of respondents hold industry certifications.

The results of the first two steps in this research study, executive phone interviews and focus groups, indicated the following 23 competencies and definitions were identified for the exemplary Property Asset Manager.

1. Honesty and integrity refer to the Property Asset Manager who people will confide in and trust, and who has a servant mindset.
2. Clear communication is defined as the manager who articulates, verbally communicates effectively, and has good writing skills.
3. Efficiency refers to the manager who is organized, manages time effectively, and resolves issues effectively. They are a researcher who checks pricing and sourcing of materials prior to purchase.
4. Passion refers to the manager who has fun in their work, is self-motivated, driven to excel, and passionate about their job, career progression, the property managed, and its people. They own their own growth.
5. Proactive refers to the manager who is decisive and takes risks as part of the execution process. She presents solutions when going to supervisors with problems, having thought them through. She is intuitive, periodically and independently revisits systems in place, checking if they make sense and updating as necessary. She walks and inspects her property independently and acts on improvements.
6. Delegates effectively refer to the Property Asset Manager who delegates without micromanaging, she trusts but verifies.
7. Confident describes the manager who is driven and decisive. She has confidence in skills and abilities, and knows when to question a directive or not, and when to pick battles.
8. Financial acumen refers to the manager who has analytical skills, analyzes financial operations of the property,

understands and is able to execute the investment strategy. She finds creative ways to maximize revenue and minimize expenses.

9. Trainer refers to the manager who trains teams effectively to adapt to a changing environment and develops team members to increase bench strength.
10. Recruiter is defined as hiring high-caliber employees, recruiting appropriately qualified applicants.
11. Marketing knowledgeable is defined as the Property Asset Manager who has knowledge of the surrounding market conditions, and can develop a comprehensive marketing plan.
12. Creativity refers to the manager who is curious, inquisitive, resourceful, creative, and inspirational in problem solving, someone who creates value with new initiatives, an "out of the box" thinker.
13. Adaptability is defined, as the manager who sees everything as possible, is flexible, resilient, and the champion of company direction, someone who welcomes and understands change, and supports initiatives. She has the ability to adapt leadership style to different generations.
14. Education focused refers to the manager who owns their own growth, participates in further education through designations and industry classes, a forever student of the business.
15. Develops employees and employee development refers to the manager who holds herself and all associates accountable, gives positive and directive feedback to employees, coaches employees proactively, evaluates employee performance systematically, and coaches employees timely.
16. Being entrepreneurial refers to the manager who has an ownership mentality. She understands owner objectives and strategizes accordingly, she is a problem solver able to identify problems and offer solutions.
17. Approachability refers to the manager who fosters an environment where it is acceptable to ask questions. She facilitates a cohesive, collaborative environment. She is the sounding block/wall sometimes just listening and not sharing.
18. Emotional intelligence is defined as the manager with interpersonal skills such as calm under pressure, patient, likeable, able to connect with others, professionally mature, and able to effectively and respectfully communicate with people at all levels.
19. Team focused describes the selfless manager who shares in recognition, is not selfish, and puts the team forward. She holds team meetings, keeps team members engaged, and leads by example. She is willing to work with others and complete work outside her own job description.
20. Loyalty is defined as the manager who trusts in firm leadership and company initiatives.
21. Inspiring describes the manager who is a motivator of others.
22. A visionary is defined as the manager who develops and implements the use of a strategic vision of the organization in the future.
23. Involved describes the manager who is involved in the industry and community.

The survey questionnaire referred to and defines the exemplary manager as "the most productive performers of any group are called exemplars or exemplary performers" (Rothwell et al., 2015a, pp. 1-27). The exemplary manager is defined as "best in class or most productive worker" as opposed to the manager who is the "fully successful performer, an experienced worker who is not best in class" (Rothwell & Lindholm, 1999, p. 91).

Survey respondents were fairly balanced with respect to gender (male 52.6%, female 47.4%). The majority of respondents were over 35 years old (58.7%), and had over 10 years of industry experience (92.2%). The majority of respondents had a 4-year degree, or higher (63.6%), and most held industry certifications (69.6%). This is beneficial to organizations and the industry because this profile describes executives of this industry.

Other valuable information and data provided the geographical location managed by the executives (subject experts) was dispersed throughout the US. The dispersed locations provided data representing national respondents and a national perspective and views of executives in the multi-family real estate industry. From an industry perspective, national views of executives are helpful to industry training initiatives especially the NAAEI, and the ability to implement the identified competencies and executives' national views throughout training and programs such as certification programs across the United States.

Within the multi-family real estate industry there are various sized firms ranging from less than 1,000 to over 15,000 units. Of the respondents in the current survey, 20.4% managed less than 1,000 units; 23.7% managed between 1,001-2,000 units; 21.1% managed between 2,001-5,000 units; 8.6% managed between 5,001-7,500 units; 13.2% manage between 7,501 and 15,000 units; and 13.2% manage over 15,000 units. See Table 4 below.

Table 1. Number of Units Managed by Real Estate Multi-Family Firms

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Less than 1,000 units	31	20.0	20.4	20.4
	between 1,001- 2,000 units	36	23.2	23.7	44.1
	between 2001-5000 units	32	20.6	21.1	65.1
	between 5001-7,500 units	13	8.4	8.6	73.7
	between 7501- 15,000 units	20	12.9	13.2	86.8
	Over 15,000 units	20	12.9	13.2	100.0
	Subtotal	152	98.1	100.0	
Missing		3	1.9		
Total		155	100.0		

In order to investigate the relative importance of the 23 competencies identified through interviews and focus groups, executives both rated and ranked the importance of each. Results indicated that respondents did not differentiate their ratings of the 23 competencies. Specifically, the average correlation between the 23 importance ratings was .901. The mean level of importance for the 23 competencies as rated by respondents was 5.4-6.8, which indicates that they were all deemed to be important. Next, the pattern of respondents' rankings was examined. Specifically, the level of agreement among respondents as to the Top 5 most important competencies. The information in Table 2 below demonstrates six different firm sizes and the percentage of respondents who ranked each competency as being within their Top 5.

Table 2. Executives Ranking of Competencies by Firm Size

Competency	Rank	% Ranked Top 5	% Ranked Top 5 Firm Size 1000 Units & Under (22)	% Ranked Top 5 Firm Size 1001-2000 Units (30)	% Ranked Top 5 Firm Size 2001-5000 Units (25)	% Ranked Top 5 Firm Size 5001-7500 Units (25)	% Ranked Top 5 Firm Size 7501-15000 Units (15)	% Ranked Top 5 Firm Size Over 15000 Units (15)
Honesty & Integrity	1	90.3	91.3	90.3	88	100	87.5	86.67
Clear Communication	2	59.5	51.6	41.94	52	45.45	75	80
Passion	3	40.2	29	31.03	44	63.64	40	35.71
Emotional Intelligence	4	32.5	12.9	36.67	24	18.18	40	62.5
Efficiency	5	28.1	40.9	22.58	24	18.18	31.25	26.67
Confident	6	26.3	19.4	30	28	9.09	26.67	21.43
Financial Acumen	7	26.2	12.9	29.03	20	25	20	40
Proactive	8	26.1	1.7	40	24	33.33	26.67	14.29
Adaptability	9	25.8	19.4	29	20	45.45	20	20
Team Focused	10	23.2	16.1	29.03	28	27.27	40	40
Develops Employees	11	22.9	9.7	36.67	20	9	13.33	35.71
Approachability	12	19.5	12.9	16.67	12	36.36	26.67	21.43
Loyalty	13	19.0	12.9	16.67	16	25	20	13.33
Marketing Knowledgeable	14	15.0	12.9	13.33	20	25	6.25	7.14
Inspiring	15	14.2	3.2	16.67	20	9	13.33	13.33
Involved	16	12.8	0	6.9	8	9.09	13.33	0
Trainer	17	8.5	1.8	16.67	20	0	0	0
Delegates	18	5.9	3.2	6.67	8	0	6.67	7.14
Visionary	19	5.9	0	3.33	16	0	6.67	7.14
Entrepreneurial	20	5.1	8.7	3.45	4	18.18	0	0
Recruiter	21	3.4	0	3.33	4	0	5	7.14
Education	22	2.6	0	17.87	4	9	0	0
Creativity	23	1.7	4.5	0	0	0	0	7.14

A number of noteworthy trends were revealed by these data. First, the same three competencies were agreed upon as being highly important by the largest percentage of respondents regardless of firm size (i.e., honesty and integrity, clear communication, and passion). Second, for all but the smallest firms the same ten competencies are important. These were honesty and integrity, clear communication, passion, emotional intelligence, efficiency, confident, financial acumen, proactive, adaptability, and team focused. Third, a greater percentage of respondents from the largest firms ranked emotional intelligence and financial acumen as being highly important, whereas a far lesser percentage of respondents from smaller firms listed these competencies as being among the most important. Fourth, team focus was agreed upon as a Top 5 competency for firms in the two largest categories to a greater degree than for smaller firms. Fifth, respondents from the smallest firms showed the least agreement overall with respect to their most highly ranked competencies. Moreover, these respondents differed from those from larger firms with respect to the importance placed on efficiency. 40% of respondents from the smallest firms listed efficiency as a Top 5 competency compared to only 18-31% of respondents from the larger firms. Finally, the same six competencies were least likely to be ranked as highly important regardless of firm size. These are: delegates, visionary, entrepreneurial, recruiter, education, and creativity. In sum, the pattern of results provides evidence to support the proposition that the competencies perceived as most important do vary as a function of firm size.

Conclusions

Limited research is available related to competencies of the Property Asset Manager. This research study provides a comprehensive listing of additional competencies (honesty and integrity, clear communication, efficiency, passion, proactive, delegates effectively, confident, financial acumen, recruiter, marketing knowledgeable, creativity, adaptability, education focused, develops employees and employee development, being entrepreneurial, emotional intelligence, team focused, loyalty, inspiring, a visionary, and involved), thereby complementing recent literature and adding to existing knowledge in the area of managerial competencies.

Examination of the rankings respondents assigned to the 23 competencies provided some support for the research proposition. As expected, the relative importance of the competencies differed as a function of firm size. The overall top ten ranked competencies are similar for all but the smaller sized firms. This indicates executives view similar competencies in the five larger sized firms (honesty & integrity, clear communication, passion, emotional intelligence, efficiency, confident, financial acumen, proactive, adaptability, and team focused) as important. In the top three largest firms (5,001 to over 15,000 units), executives viewed approachability highly in addition to the top ten ranked competencies previously mentioned. This indicates the manager who fosters an environment where it is acceptable to ask questions, a cohesive, collaborative environment is an important competency for Property Asset Managers in the three largest firms. Emotional intelligence and team focused were competencies ranked by executives in the two largest firms as critical. This suggests the Property Asset Manager of the two largest firms interact with a higher number of employees therefore it is critical to helm interpersonal skills such as calm under pressure, patient, likeable, able to connect with others, professionally mature, and able to effectively, and respectfully communicate.

Respondents from smallest sized firms showed less agreement with respect to their rankings than did those from the larger firms. Only four competencies were agreed upon as being most important by 20% or greater of these respondents. These competencies were honesty and integrity, clear communication, passion, and efficiency. Although three of these were also agreed upon as highly important by the remainder of the firms, efficiency was not. This suggests that in smaller firms (in executives' views) efficient use of resources and time is more important in determining a Property Asset Manager's effectiveness than it is for Property Asset Managers at larger firms. Larger firms have firm or corporate support and departmental structures and functions such as marketing departments, revenue management support, purchasing departments, and advertising support.

The expectation that competencies vary as a function of firm size suggests that the Property Asset Managers role would vary with firm size due to variables such as number of employees and resources. Larger firms are more hierarchical and have more sophisticated support systems to assist the manager such as marketing, purchasing, and training departments. Smaller firms have a less sophisticated support system, requiring the Property Asset Manager to operate and wear "many hats" to function effectively. The importance of this proposition and rationale that competencies vary as a function of firm size enables firms to tailor identified competencies to the specific training, evaluation, and recruiting needs unique to each firm size.

Guided by the proposition and research questions, this study has identified the unique competencies that subject experts in the multi-family real estate industry expect of their Property Asset Managers, with the belief that such competencies can produce exemplary, high-level property management and lead to strong, performance-based results for their organizations. In addition, this paper has explored the relative importance of the various identified competencies. By identifying, classifying, and discussing the key competencies for Property Asset Managers, according to their supervisors and executives in the multi-family real estate industry, this research provides invaluable information that is highly relevant to both research and practice.

References

Barney, J. (1991a). Firm Resources and sustained Competitive Advantage. *Journal of Management*, *17*(1), 97.

Barney, J. (1991b). Special Theory Forum: The Resource-Based Model of the Firm: Origins, Implications, and Prospects. *Journal of Management*, *17*(1), 97.

Barney, J., Wright, M., & Ketchen Jr., D. J. (2001). The resource-based view of the firm: Ten years after 1991. *Journal of Management*, *27*(6), 625.

Cadogan, and Fahey (JBR 2005). *Journal of Business Research*. pp. 148-151. doi:10.1016/j.jbusres.2005.03.001.

Ferris, G. R., Arthur, M. M., Berkson, H. M., Kaplan, D. M., Harrell-Cook, G., & Frink, D. D. (1998). Toward a social context theory of the human resource management-organization effectiveness relationship. *Human Resource Management Review*, *8*(3), 235.

Fields, T. J. (2015). Blurred Lines, Redefined. *Journal of Property Management*, *80*(1), 42-46.

Harrell-Cook, G., & Mahoney, J. T. (2001). Manufacturing Advantage: Why High-Performance Work Systems Pay Off. *Academy of Management Review*, *26*(3), 459-462.

Hui, L., Keiko, T., Lepak, D. P., & Ying, H. (2009). Do They See Eye to Eye? Management and Employee Perspectives of High-Performance Work Systems and Influence Processes on Service Quality. *Journal of Applied Psychology*, *94*(2), 371-391.

Juttner, U., & Wehrli, H. P. (1994). Competitive advantage: Merging marketing and the competence. *The Journal of Business & Industrial Marketing, 9*(4), 42-53.

Ku, E. (2014), Putting forth marketing competencies strength with collaborating partners in the hotel industry. *Service Business, 8,* 679-697.

Kyle, R., Spodek, M., & Baird, F. (2010). *Property Management*. 8th ed. Chicago: Dearborn Real Estate Education.

Lawler, E. E. (1994). From job-based to competency-based organizations. *Journal of Organizational Behavior*, Vol. 15, 3-15.

Lung-Far, H. (2014). A Study of the Relationships Among Self-Directed Learning, Managerial Competency and Manager's Job Performance. *Proceedings for The Northeast Region Decision Sciences Institute (NEDSI)*, 989-994.

Multifamily Housing, Multifamily Book Project (MBP) LLC, 2015 ISBN: 978-0-9964180-0-3

Real Estate Management Professional Leadership Competency Model, The. (2013). *IREM White Paper.* Chicago: Institute of Real Estate Management, 2013.

Rothwell, W. J., & Lindholm, J. E. (1999). Competency identification, modeling and assessment in the USA. *International Journal of Training & Development*, *3*(2), 90.

Rothwell, W., Graber, J., & Dubois, D., (2015a). *The Complete Competency Toolkit*, Vol 1, Second Edition, HRD Press.

Rothwell, W., Graber, J., & Dubois, D., (2015b). *The Complete Competency Toolkit*, Vol 2, Second Edition, HRD Press.

Semeijn, J. H., Van Der Heijden, B. M., & Van Der Lee, A. (2014). Multisource Ratings Of Managerial Competencies And Their Predictive Value For Managerial And Organizational Effectiveness. *Human Resource Management*, *53*(5), 773-794. doi:10.1002/hrm.21592

Sikora, D. M., Ferris, G. R., & Van Iddekinge, C. H. (2015). Line Manager Implementation Perceptions as a Mediator of Relations Between High-Performance Work Practices and Employee Outcomes. *Journal of Applied Psychology* Vol. 100, No. 6, 1908–1918.

Society for Human Resource Management. (2012). SHRM Competency Model. Retrieved from: http://www.shrm.org/hrcompetencies/pages/model

Spencer, L., & Spencer, S. (1993). *Competence at Work: Models for Superior Performance*. New York: John Wiley and Sons.

Zietz, Emily. (2003). Multifamily Housing: A Review of Theory and Evidence. *Journal of Real Estate Research,* Vol. 25 (2), 185-244

The ETF Liquidity Illusion

Richard A. Lewin & Marc J. Sardy, Rollins College

Abstract

We consider the prospects for a liquidity mismatch between ETFs and their underlying investments in the context of previous market stress events. As the global ETF industry eclipsed $5 trillion in 2019, the impact of ETFs on underlying stocks has become extremely acute. Our paper considers the non-zero probability of a risk mismatch in liquidity between an ETF and the underlying securities it holds, in the event of a (not-low-probability) panic sell-off in the market. Liquidity for an ETF asset may, or may not be, constant through time and thus may become a potential flashpoint for a future financial crisis.

Introduction

Exchange Traded Funds (ETFs) have been available on US public markets since 1993 when the Securities and Exchange Commission (SEC) approved the first ETF - a broad-based domestic equity fund tracking the S&P 500 index (Shilling, 1993). According to the Investment Company Factbook 2018 (ICI, 2018), assets under management in exchange-traded funds have ballooned from $151 billion in 2003, after a slow initial decade, to more than $4.7 trillion through the final quarter of 2018. ETFs offer the systematic advantages of low expense ratios and reduced transaction costs, simplicity, a high degree of diversification, coupled with tax efficiency and transparency. Moreover, given the burgeoning size of the market, now approaching 18% of all US equities, the impact of open-ended ETFs on their underlying stocks may also have become quite significant.

Although much of the secondary trading in ETFs does not directly influence the underlying securities held therein, the net withdrawals and receipts must still be reconciled at the market close, which enforces consistency between the stock of assets according to the net flow of funds on a daily basis. Notable areas of concern previously addressed include market efficiency arguments, where there may be much less pressure for accountability to shareholders (Israeli, Lee, and Sridharan, 2016). The market efficiency issue being that as more of the free-float equity becomes locked-up within passive funds, which typically do not exercise their voting rights, there may be less ability for residual active investors to exercise market discipline against miscreant managers. Whilst many advocates of ETFs still see significant advantages to their structure, prior to his recent passing John Bogel, the founder of Vanguard and pioneer of low costs index funds, made a compelling case (Bogel, 2018) for concern and promulgated a need to mandate voting by index funds. Clearly once a significant minority of free-floating shares become locked-up within these passive vehicles, there risks an abrogation of corporate governance responsibilities should voting rights remain perpetually unexercised.

Previous academic research has generally focused on determining whether ETFs may increase volatility and actually reduce liquidity of the underlying securities, by restricting free-float as well as increasing the co-movement and co-integration in returns and liquidity of the underlying component securities (Bhattacharya and O'Hara, 2016). The concurrent trading of ETFs and the vast array of securities they now hold, represent a growing challenge to upholding the law of one price through limits on potential arbitrage. ETF prices ought to be kept continuously in line with the intrinsic values of their underlying security constituents through a process of arbitrage, in which market makers, authorized participants (APs), as well as hedge funds and other institutional investors, participate. As passive investment vehicles absorb an ever greater degree of available equity such arbitrage mechanisms may no longer be able to function appropriately.

This paper thus examines the potential for massively magnified risks should such funds run into difficulties in a future liquidity crisis. We only have to consider the dire consequences of the famous money market funds which astonishingly 'broke the buck' in the great recession, freezing redemptions globally, to understand the magnitude of the inherent problem. During ordinary market conditions, counterparty risks typically fade like noise into the background, but as many investors contemplate more robustly potential liquidity events, as part of their 'Black Swan' preparation strategy, scrutiny of the processes underlying ETF activities become paramount. Investors may naively conceive ETFs as some kind of liquidity panacea and a great way to diversify whilst getting around active management fees. Whilst the passive versus active management argument has moved roundly in favor of index funds, in no small part due to reversion to the mean of active managers through time, the darker machinations behind ETF functioning remains obfuscated behind the Baumian curtain.

Investors may be suffering from a liquidity illusion if they are endeared to their diversified ETF comfort blanket, as investors who owned VIX ETFs rudely discovered in February 2018 when these declined 90% in a few days over the ensuring market volatility, and never came back, not all ETFs are created alike. Credit Suisse Group AG moved to liquidate one ETF product focused on 'short on volatility', and trading in more than a dozen similar products were also halted after their values precipitously sunk toward zero. VelocityShares Daily Inverse VIX, a short-term exchange traded note (ETN) which was known

by its trading symbol XIV, had been worth almost $2bn just the previous month (Foerster and Evans, 2018). Investors in all those 'diversified' index funds (be it ETFs or just plain-vanilla funds) swiftly discovered that when you own even a diversified basket of uniformly overvalued assets, that diversification and implied liquidity may be merely illusory.

Liquidity Issues

Implicitly leveraged ETFs are designed specifically to amplify returns (risk being the counterparty in this double-edged sword arrangement, which means that they inevitably amplify any potential losses), while inverse ETFs are designed to profit when the underlying benchmark declines in value. By corollary we note that these instruments create unlimited risk, should the opposite event happen. Most notably in raising our concerns of inherent liquidity risk were a couple of the 'flash crashes', and particularly the tumultuous events occurring in May 2010 and August 2015, when on both occasions the market fell precipitously and unexpectedly, and ETF prices along with their underlying assets were no longer trading in lockstep as they are mandated to do. Another exacerbating factor is the disproportionate share of the index that a handful of the major stocks now represent. Indeed the S&P 500 Index for example is a market capitalization weighted index, which means larger companies will have a higher weighting within the index. Investors in these type of indexes may not be quite as diversified as they seem to think they are, due to increasing concentration risk. At one point in 2018, the top 10 stocks which includes the celebrated FAANG stocks (Facebook, Apple, Amazon, Netflix & Google/Alphabet) represented almost 30% of the entire S&P500 index (with the technology sector alone representing 26%), which potentially imperils investors due to much higher levels of concentration risk than a seemingly diversified index otherwise belies. During a dysfunctional market, where investors try to sell ETFs at several times their average daily volume, these ETFs in turn were forced to go out and attempt to sell the underlying assets they held by way of basis reconciliation. Supply then temporarily exceeded demand, and thus the prices of these assets should have dropped precipitously.

Intriguingly, what transpired was instant reaction, especially so from the less liquid stocks within the S&P500 index, followed by a virtual suspension of trading activity in these lesser stocks during the ensuing carnage. Larger more liquid stocks continued to be driven-down along with the index however, in some cases becoming tantamount to penny stock valuations suggesting that the more liquid components were almost certainly being sold off relative to the less liquid ones, just to maintain parity of the overall ETF valuation versus the underlying constituent basket. Our conclusion in the aftermath of which is that extra risk was being borne by market makers by dropping a greater component of the larger more liquid holdings, simply to counteract the illiquidity of their own long positions within smaller constituents of the S&P500 market index. Even assuming they took option derivative positions to restore such temporary imbalances, cascade risk remained had indexes continue their descent over subsequent trading sessions. This scenario would cause the price of ETFs to decline further, creating losses for exiting ETF holders and permanent losses for any remaining ETF holders, in addition to a raft of potential capital gains exposure, or potentially in a worst case scenario both. Indeed something similar already happened during the 2008 financial crisis to action-rate securities, which were extremely liquid, right up until the very moment they suddenly were not. Furthermore, in that instance it took investors several years to get their money back. Liquidity can thus remain a trigger for the next crisis even when trust in diversified trading instrument is brought into question by market participants, especially with so much leverage already implicit in the system. The prospect of a liquidity mismatch between an ETF and its underlying investments also raises questions about credit lines being extended by syndicates of lenders to certain ETF sponsors. Should assets held by an ETF be viewed as too illiquid to sell, particularly in times of severe market stress, managers would instead be required to tap into available credit lines to cover any interim redemptions, in the potentially misguided expectation that volatility should subside and underlying assets could then be sold at more reasonable and reliable prices in the future.

We note that the same salutary lesson was imbued in the construct of Long Term Capital Management (LTCM) in 1998 where they were led to increasingly magnify their hedged positions just prior to the Russian default, by increasing borrowings on the erroneous supposition that provided they remained hedged, markets would eventually revert to 'normal' behavior. The flaw in this logic remains that it relies on volatility subsiding, and hence prices rising, or at least stabilizing. Otherwise, existing investors would simply be left footing the bill for these tapped out credit lines in the same way that the syndicate of banks were on the hook for the Fed sponsored bailout of LTCM. This is especially risky with the more levered ETF products, due to the exacerbation of implied leverage as asset values implode, coupled with magnification of liquidity squeezes at precisely the same moment. Essentially, whenever a lack of liquidity event manifests itself, it is rather like attempting to squeeze a camel through the eye of the proverbial needle. It is however, the *raison d'être* for the much-maligned cash reserve requirement, which at low nominal rates of interest tends to act as an anchor dragging on performance in the good times, but is an absolutely essential buffer to avoid the necessity of simply hoping for the best by reliance on lines of credit at such a critical liquidity hiatus. More generally, there remains a remarkable potential cascade effect for equity markets, and in any case this is a risky way to obfuscate underlying market volatility to investors, as eventually the fund must reimburse any such loan positions. That is not to say that a well-managed credit line would not help to smooth out short periods of market illiquidity.

However, by extension, if it proved a more protracted and prolonged reaction given there are none who can readily divine markets, it is essentially a bet on liquidity and more stable markets resuming. A report by the Financial Stability Board (2011) noted that while credit facilities could reduce financial stability risks short term, they could also act to raise leverage on already distressed funds, and thereby compound the threat of contagion across the wider financial system. Just because an ETF contains a plethora of positions it does not follow that it is unaffected by overall market volatility shifts. The fallacy here, often misattributed to Keynes, is that markets can remain irrational far longer than you can remain solvent (Shilling, 1993).

Another issue is that legacy holders of an ETF ultimately shoulder the burden of these risks, given that those who have redeemed early on are already safe. With an ETF structure there is no redemption penalty associated with removing your money during any specific period, as is typically faced with other mutual funds as you still have secondary and primary trading that offer more avenues for liquidity. Naturally the potential for large swings in underlying values will mainly depend on the scope of the fund. An ETF that tracks a broad market index, such as the S&P 500 (however we note that there is notable variation in the performance of the 27 ETFs that mimic the S&P500), is likely to be less volatile than another ETF that tracks a specific industry or sector, such as oil services. Therefore, it is vital for investors to be aware of the fund's focus and precisely what types of underlying investments it includes in its portfolio.

In the case of international or global ETFs, the fundamentals of the country that the ETF is following are also important, as is the creditworthiness of the particular currency of that country. Political risk, economic and social instability will also contribute an outsize role in determining the outcomes of any ETF that invests in a specific country or region. Together these factors must be borne in mind whenever making decisions regarding the viability of a particular ETF instrument from an investment perspective. Investors must be especially vigilant in respect of any subsector that inadvertently overpromises liquidity, or where investors themselves have just simply assumed much greater liquidity exists in the ETF instrument than the underlying asset class can actually support. To suggest, for example, that a derivative instrument based on the bond market can be more liquid than its underlying constituents is borne out of either delusion or hubris. An ETF can only be as minimally liquid as the least-liquid security encompassed within its portfolio, and that is the correct worst-case proxy for investors to use in order to factor in the correct liquidity risk perspective under conditions of market stress. An ETF's liquidity is correlated with its underlying holdings' liquidity, as opposed to its average daily trading volume or total assets under management, which are more often misspecified by investors as their working proxies for an ETF's implied liquidity.

A classic example being emerging markets and high-yield bonds, which could easily face a liquidity crunch, should the tide of liquidity which has driven more esoteric offerings, finally turn. For regulators the billion-dollar question remains whether such market stress events result in contagion or not. The imperative here is to know what the ETF is actually tracking, and understand the underlying risks associated with its individual components. ETFs are also not costless, they still have an implicit bid-ask spread, which can vary from one penny to several dollars, and will widen dramatically during enduring market stress events. From tracking errors with their evolving indices and trading close to net asset value, disparities may grow exponentially during flash crashes. Moreover, in the extreme scenario of an ETF closing down, the fund will be liquidated and shareholders actually paid out in cash. However, the ETF will most likely realize significant net capital gains during this liquidation process, which it will be obliged to pass on to shareholders of record. Moreover, there will also be transaction costs and uneven tracking errors incurred during this process and various other more minor grievances including shareholders bearing the *pro rata* legal costs of technically winding up the fund.

Challenges to liquidity are especially peaked during market stress events or extreme financial crises, when the complex structures of ETF products and underlying counterparty risks implied, need the most careful scrutiny. This coincides with the SEC preparing a wide-ranging review of the booming ETF industry, and reflects concerns that recent huge capital flows into this area may well be exacerbating prospective market volatility as the free-float underlying available in the general market diminishes. For example, Virtu Financial, a high frequency broker, urged the US Securities and Exchange Commission to examine potential operational risks associated with exchange-traded funds (Bullock, 2016). In their letter to the regulator, they highlighted this as an "area that clearly deserves a closer look". In particular, Virtu's concerns surrounded risks that it was exposed to if trading ETFs very near the close of business. In response, the regulator's working group is anticipated to review every aspect of the ETF industry and market consequences of its enormous growth.

It is important to note that market makers in ETFs are required to submit orders to their respective fund sponsors to create or redeem ETF share units, depending on market demand. Inherently market makers bare this risk in order to earn the spread differential between bid/offer prices on the ETFs themselves, and any difference between the pricing of the underlying components and the resale value of the ETF unit. ETF market makers are obligated to maintain a two-way market in the ETFs, right up and until the close, while simultaneously buying and selling the underlying constituents. Inevitably no market maker can control the precise number of executions needed approaching the close, nor can they robustly anticipate individual closing positions with any great degree of accuracy in advance of the closing bell. Risk will thus be compounded if an ETF specialist is unable to submit, create and redeem units right up until the market close. Implicit risk factors are therefore passed onto investors via wider spreads in the embers of the trading day.

Anecdotally at least, "electronification" of ETFs in comparison to their explosive growth trajectory has been quite ponderous. Accelerating accurate underlying processing should facilitate trading by market makers much nearer to the close of business. However, the multiplicity of interactions required for redemption and creation of ETF units are still being conducted through dedicated email, or oftentimes directly over the phone. This exposes the entire market to far greater risks of human input errors or 'fat finger' events and omissions in processing transactions, undermining fund providers reporting accuracy in respect of the component basket versus net asset values, which are utterly crucial to market making. End users would find such procedural improvements especially important since the reduced risk would translate into lower costs and narrower spreads.

Many ETFs market makers opine that derivatives have had no impact on their market making, or their ability to hedge for leveraged funds, but specific scenarios that could arise include significant counterparty risk implicit in these instruments which simply cannot be ignored. We note that the SEC's intention (SEC, 2015) to limit aggregate derivatives' exposure to 150% of the funds' total assets, or 300% if the derivative is intended to reduce losses arising from market moves, portends that these risks remain a real danger. Whilst this might protect investors against a measure of amplified losses, our concern here remains the one size fits all threshold-based approach being adopted, as the real focus ought to be on counterparty risk and liquidity of these derivatives instruments themselves, coupled with their impact on the funds that use them. ETFs should reflect the underlying values of their securities, plus or minus a liquidity premium. The SEC, in a postmortem of the 1987 collapse, advocated for a single security that could represent the broad market, on the theory that such an issue could increase overall market liquidity. ETF trading can have an impact on overall market moves, which has been hotly debated for years, particularly within the context of the active versus passive debate. The concern arises because investors essentially buy every stock in an index when they buy a passive fund: if a crash were to result in heavy selling of index funds, the entire basket would be sold indiscriminately regardless of the fundamentals of the individual underlying securities. And because ETFs are a way to make immediate decisions on the overall market (in contrast to mutual funds, which only price and trade at the end of day), selling them could mean the rout becomes even more indiscriminate in respect of what securities are actually being sold.

The hyperbolic growth in ETFs has created a situation of extreme illiquidity for many of the underlying stocks. Calamia *et al.* (2016) found evidence to suggest that the relationship between ETF and asset class liquidity does not provide the full picture. As an example Johnson & Johnson is a stock where according to Bloomberg, 6.36% of shares outstanding are held by ETFs. This translates to about 170.7 million shares, compared with a 30-day average trading volume of 5.6 million. Those 170.7 million shares, in other words, represent more than 30 days of average daily trading. The real liquidity issue may thus be capacity, as the real liquidity issue lies in the asset classes that ETFs cover, and just how big and liquid the underlying markets remain to accommodate large and growing ETFs. The less liquid the asset class, and the bigger the fund, the more likely an ETF sponsor is going to run into these issues. ETF providers can increase their liquidity through their choice of replication strategy. For example, for less liquid markets such as emerging markets, sponsors like Lyxor use a synthetic ETF structure (FSB, 2011b). In this example the assets owned by the ETF can be restricted to quality European assets, whereas the performance of the emerging market can be precisely replicated using a performance swap. This means the ETF holdings are more liquid than the underlying asset provided the counterparty risk is notional. On 24 August 2015 some of the major US ETFs diverged from their net asset value (NAV) by up to 50%, creating significant losses for investors. In contrast, over the same period, Lyxor observed no more than 3% of divergence between the price of its S&P 500 ETF and the intraday net asset value of the fund. The key disparity of behavior can be attributed to differences in the markets.

In the US, ETFs are considered independent vehicles with their own source of liquidity, separate to that of the underlying assets, whereas in Europe ETFs are considered an extension of their underlying assets. As such, European exchanges tend to enforce far stricter rules to keep ETF prices in line with moves in their underlying assets. When a large selloff happens in financial markets we would expect to see that S&P components should move proportionally as their role in the S&P 500 is identical and scaled by weight. Thus we should see proportional moves in components across the range of constituents. However, in the event of a run on the ETF or 'flash crash' we might see distorted movements in the components that would make it harder for the ETF to execute a proportional selloff. Thus, ETF performance impact on the S&P 500 may lead to distortions in the performance of the S&P components themselves, as funds become desperate to unload given major market moves by investors unloading funds in a market crash.

We call S&P components "S" and the S&P is a weighted average of its components based on market cap, where *W* is the weighting given each component *n* and *R* represents the return on the asset *n*.

$$R_{S\&P500} = \sum_{n=1}^{500} WnRn \quad (1)$$

However, in the event of a market crash, liquidity will become more problematic for the smaller components, if Size is represented by 'L' then:

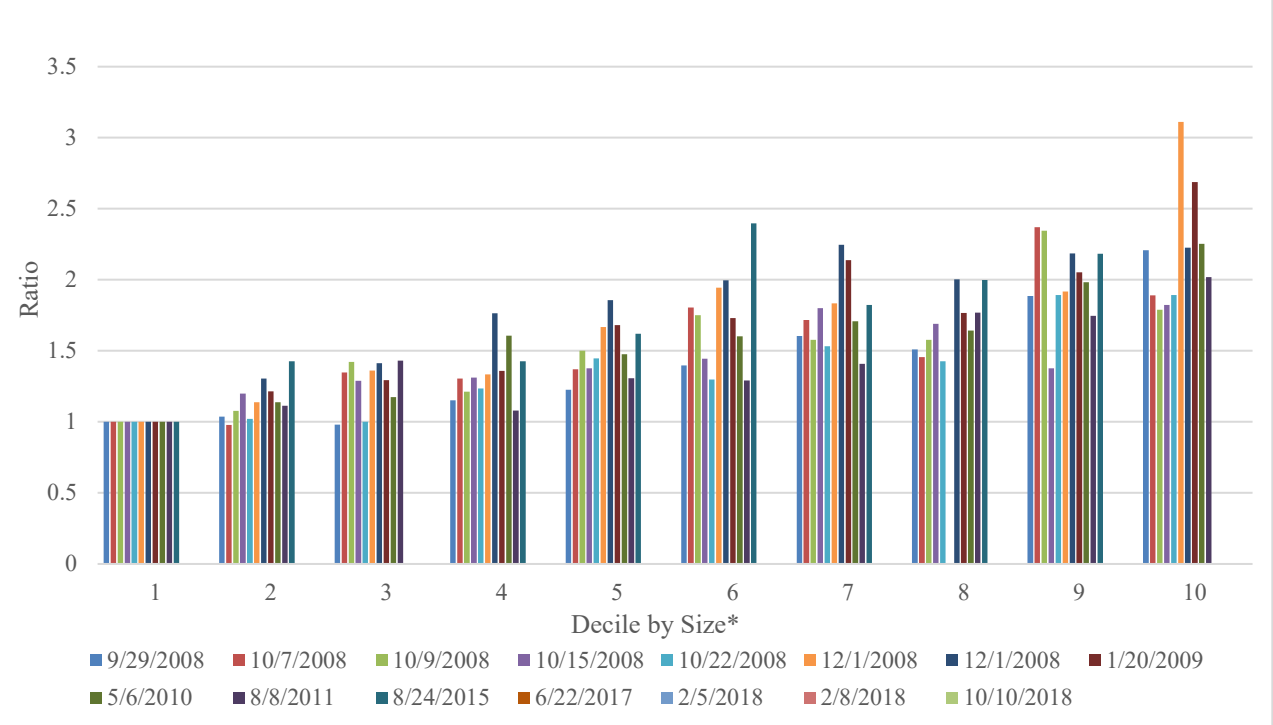
$$L_1X_1 > L_2X_2 > \ldots\ldots\ldots L_nX_n \quad (2)$$

So ETFs may try to modify the order of the firms they sell off, to help them manage their shortage of funds compounded by an investor's ETF selloff of many funds simultaneously. More weight would inevitably be put on the larger more liquid firms to allow the fund to match the aggregate S&P performance.

$$\text{Thus if; } L_1X_1 = \sum_{n=450}^{500} W_n X_n \quad (3)$$

Then for each share of X_1 sold, there could be performance offset for say the bottom 50 firms within the S&P 500. Thus, the effect would be to see more shares of the top firms sold off heightening risk associated with these firms. However, in practice, our results are uneven across the components of the S&P that we examine and appear to behave in the reverse order we initially hypothesized. We examined S&P 500 components from different one-day event drops across the S&P 500 using data from Bloomberg. We looked at their component weights and several other key measures. The main metric we used was to look at the trading volume on a large down market day, divided by the firm's shares outstanding. These were broken out into deciles, based on the market capitalization of firm, as the following figure illustrates:

Figure 1: Ratio of Volume to Shares Outstanding for largest Market Moves

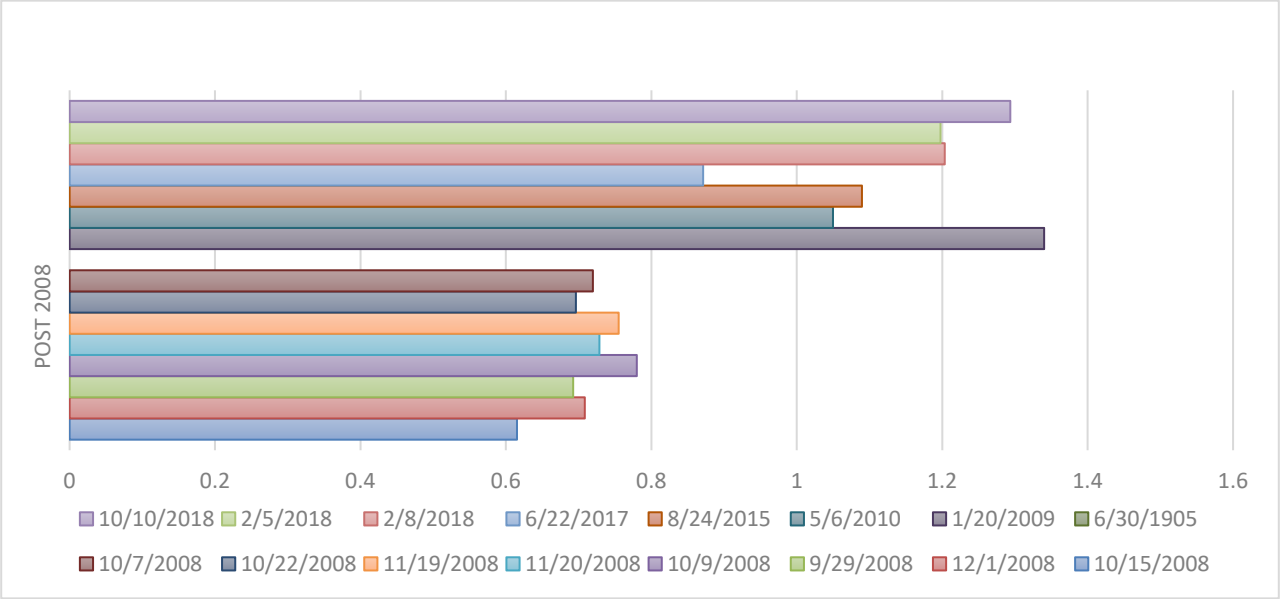


Note that in the above figure 1, three deciles were removed in order to better display the data. Those not shown included decile 3 on 6/22/17 which had a value of 6.05, decile 8 on 12/01/2008 which had a value of 3.36 and Decile 10 on 6/22/2017 which had a value of 3.70.

These results appear somewhat counterintuitive at first. The largest firms have the lowest ratio of trading volume to shares outstanding. We used this as a proxy for the percent of the firm that turned over on any specific date. The smallest firms have the largest ratio of volume to shares outstanding. We hypothesize that the level of institutional traders compelled to hold the firms within the S&P 500 ETF might lead to a distortion in the movement of the underlying that puts more pressure on smaller firms, whereas the institutional holdings might represent a larger percentage of total shares outstanding within the largest components. As might be anticipated, the financial crisis that began in 2008 is one of the most volatile years in the last two decades when many of the top market drops over the last 20 years occurred. Thus, the top 10 firms during those dates were a smaller percentage of trading volume versus total shares outstanding. This suggests that the bulk of the average volume versus shares outstanding is in the remaining 40 firms. In the 2008 period, the top 10 firms represented approximately 70% of the average volume versus shares outstanding. However, it seems that after 2008, the ratio of the top 10 firms as ranked by market capitalization seem to have attained a much higher ratio, averaging 15% above the ratio of the top decile.

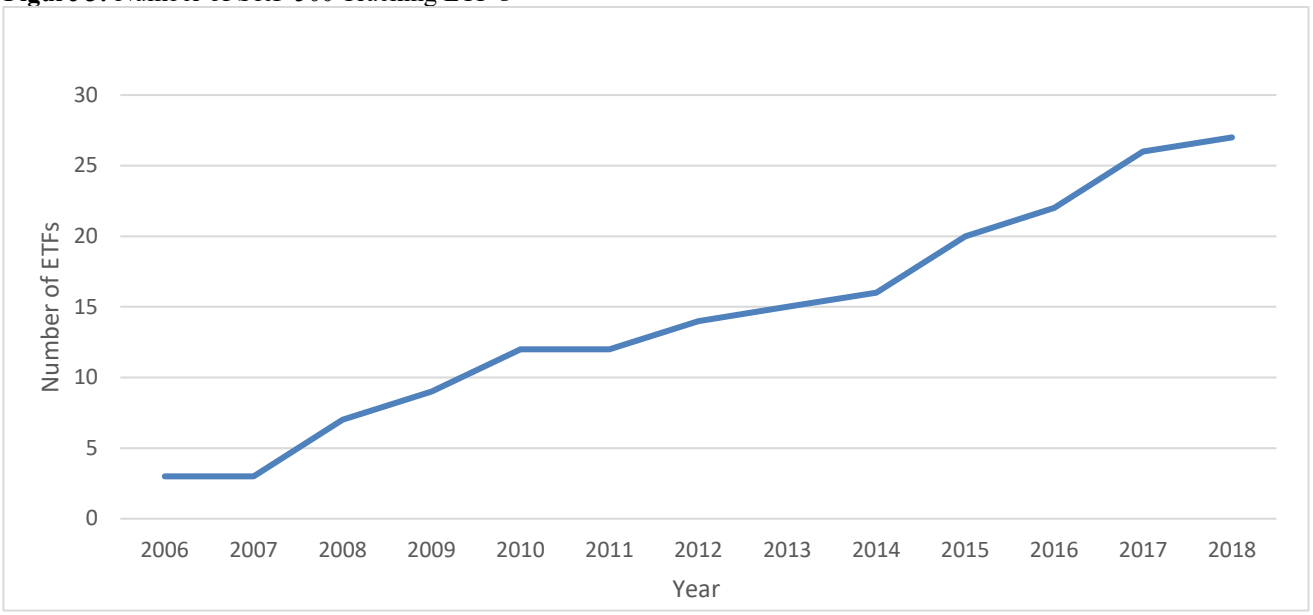
In figure 2 for the period subsequent to 2008, the ratios appear much larger than the rest of the top decile which supports our hypothesis that firms are disproportionately liquidated by size, with the largest being sold off in a large bear market more rapidly than they are being acquired in a bull market.

Figure 2: Ratio of Trading Volume to Shares Outstanding Top 10 vs 1st Decile



This is an important distinction and one which allows ETF managers to mask the illiquidity issues involved. During the same period market priorities of investors may have shifted as well. More investors became interested in ETFs as a safe vehicle for diversification compared to other investment opportunities. As a result of th increasing move from active to passive management there has been a proliferation of ETFs that follow the S&P 500, from about 4 officially tracking the S&P 500 in 2008 to about 27 by end of 2018 as illustrated in figure 3.

Figure 3: Number of S&P 500 Tracking ETF's



This represents a 20% annual growth rate in the funds that have to buy and sell all of the underlying constituents of the S&P 500 index. This does not include all of the firms that track the S&P components in some form or another however, and even

within these there are notable tracking errors. Liquidity of a given stock is a function of the free float available and the proportion of institutionally held shares. As more funds become housed within ETFs, as a proxy for mimicking aggregate market returns, it will be increasingly rare for such funds to sell the underlying shares or reduce positions, other than very slowly over time unless an actual emergency or panic occurs. Therefore in a 'flash crash' scenario, it may be anticipated that institutions may just sit out the issue if one ETF is down due to investor flight, as this is now on the shoulders of the managers of the ETF to liquidate their underlying positions. Interestingly because they have to unload large positions across several areas, their models may suggest that should a large volume of stock be sold in MSFT for example, then the fund should divest too, which could ensnare algorithmic traders to simultaneously attempt to exit, causing prices to freefall as occurred in 2010 and 2014 Here a small number of traders had taken the market down so quickly, that most institutional investors were unaware, unable or unwilling to react (IOSCO, 2011, 2013).

On the one hand had ETFs placed a brake on some movements of the companies they trade-in, in theory there would be more stability, but once settled in an environment of less free-float there should be more marginal impact of actual underlying trading. The bulk of shares tied up in these forms of vehicle are thus unwittingly accentuating smaller movements due to reduced free-float in the market. This plays into the hands of high frequency traders who already dominate trading volumes and may more easily influence and / or manipulate market prices with a flood of orders. Blackrock, Fidelity, Vanguard and State Street among others, have the largest exposure in terms of their liquidity effect on the remaining free-float in the market. Looking at a return shift in respect of a percentage change in volume over shares outstanding, our theory contends that a smaller return should shift smaller volume relative to shares outstanding at the lower end of the range.

Conclusions

Junk bonds were all the rage in 1980s, and then they suddenly fell from grace (Hurduzeua and Popescu, 2015). Technology and dot-com stocks observed almost the same maniacal treatment in the late 1990s, and investors are well versed in how that story unfolded. ETFs are most likely heading down that same well-trodden path, as the inherent realities of passive index domination begin to surface, or as Lucius Seneca might have more eloquently put it, 'Time will discover the truth.' As we have illustrated, ETFs have inadvertently absorbed much of the liquidity of the underlying S&P 500 components due to restricting the remaining free-float available, particularly among the more dominant firms. We contend that this has created a potentially perverse situation where the prudent advisors who recommend ETFs as a safe diversified investment product, may be inadvertently fueling an edifice for their clients that becomes more unstable, rather than less, the larger it gets. This seemingly innocuous practice appears to provide fertile ground for a future bubble to develop within the global stock market and an area of compelling future research.

References

Agarwal, Vikas, E., Paul E.Hanouna, Rabih Moussawi, and Christof W. Stahel, 2018. "Do ETFs Increase the Commonality in Liquidity of Underlying Stocks?", *28th Annual Conference on Financial Economics and Accounting; Fifth Annual Conference on Financial Market Regulation 2018.* Available at SSRN: http://dx.doi.org/10.2139/ssrn.3001524.

Bhattacharya, A., and M. O'Hara, 2016. Can ETFs increase market fragility? Effect of information linkages in ETF markets. *SSRN Electronic Journal*. http://ssrn.com/abstract=2740699.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi. 2014. Do ETFs Increase Volatility? *NBER Working Paper No. w20071*, pp.1-59. Available at SSRN: https://ssrn.com/abstract=2430066.

Bogle, John C. 2018. *Stay the Course: The Story of Vanguard and the Index Revolution*, Wiley.

Bullock, Nicole. 2016. Virtu calls on US regulator to examine risks of ETFs. *Financial Times* (Nov 22nd, 2016) https://www.ft.com/content/1d00d414-b0cf-11e6-9c37-5787335499a0.

Calamia, Anna, Laurent Deville, and Fabrice Riva. 2016. The Provision of Liquidity in ETFs: Theory and Evidence from European Markets. *SSRN Electronic Journal.* https://ssrn.com/abstract=2835907.

Dannhauser, Caitlin. 2017. The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs). *Journal of Financial Economics,* Vol 125, Issue 3, pp.537-560.

Financial Stability Board. 2011a. "Potential financial stability issues arising from recent trends in Exchange-Traded Funds (ETFs)". FSB Note on ETFs. https://www.fsb.org/2011/04/potential-financial-stability-issues-arising-from-recent-trends-in-exchange-traded-funds-etfs/.

Financial Stability Board. 2011b. "Re: Feedback by Lyxor Asset Management on the note issued by FSB". https://www.fsb.org/wp-content/uploads/c_110910h.pdf.

Foerster, Jan-Henrik and Rachel Evans. 2018. Credit Suisse Fund Liquidated, ETFs Halted as Short-Vol Bets Die. *Bloomberg*. https://www.bloomberg.com/news/articles/2018-02-06/credit-suisse-is-said-to-consider-redemption-of-volatility-note.

Hill, Joanne, Nadig Dave, and Matthew Hougan. 2015. A Comprehensive Guide to Exchange-Traded Funds (ETFs). *Research Foundation Books,* 3, pp.1–181.

Hurduzeua, Gheorghe and Maria-Floriana Popescu. 2015. "The History of Junk Bonds and Leveraged Buyouts" *Procedia Economics and Finance,* Emerging Markets Queries in Finance and Business, Vol.32, pp.1268-1275.

ICI. 2018. "A Review of Trends and Activities in the Investment Company Industry", Investment Company Fact Book, 58th edition, Investment Company Institute, www.icifactbook.org.

IOSCO. 2011. "Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency", *Report of the Technical Committee of International Organization of Securities Commissions* CR02/1, pp.1-60. https://www.iosco.org/library/pubdocs/pdf/IOSCOPD354.pdf.

IOSCO. 2013. "Principles for the Regulation of Exchange Traded Funds" *Final Report Board of the International Organization of Securities Commissions*, FR06/13. https://www.iosco.org/library/pubdocs/pdf/IOSCOPD414.pdf.

Israeli, Doron, Charles Lee, and Suhas Sridharan. 2017. "Is There a Dark Side to Exchange Traded Funds? An Information Perspective", *Review of Accounting Studies*, Vol. 22, pp.1048-1083. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2625975.

SEC. 2015. "SEC Proposes New Derivatives Rules for Registered Funds and Business Development Companies" https://www.sec.gov/news/pressrelease/2015-276.html.

Shilling, A. Gary, (1993), Scoreboard, *Forbes*, Volume 151, issue 4, pp.236. https://www.fsb.org/wp-content/uploads/c_110910h.pdf.

Smith, Lisa. 2018. An Inside Look at ETF Construction, *Investopedia.* https://www.investopedia.com/articles/mutualfund/05/062705.asp#ixzz5TLib4r4j

Bootstrapping a Zero-Coupon Yield Curve: An Exercise for Finance Students

William A. Reese, Jr. and Russell P. Robins, Tulane University

Abstract

Since Treasury bills, notes and bonds of different maturities have different coupon rates, a traditional yield curve cannot be used to accurately price other bonds. A zero-coupon yield curve must be constructed from information found in current Treasury debt instruments through a process known as "Bootstrapping". We have developed an Excel spreadsheet with an imbedded macro that will obtain current Treasury quotes from the Wall Street Journal's website, download them to the spreadsheet, and use them to bootstrap a zero-coupon yield curve. Using this spreadsheet in the classroom allows students to better understand how bonds are priced through experiential learning.

Introduction

Introductory finance courses at both the undergraduate and graduate level typically teach students that the value of any financial asset is the present value of its (expected) future cash flows. This involves two steps. First, the cash flows must be forecast. Second, an appropriate discount rate (or appropriate discount rates) should be selected. It is crucial that the discount rate(s) reflect the riskiness of the cash flows.

When pricing Treasury notes or bonds, the cash flows are easy to determine. The semiannual coupon payments are set from the bond's coupon rate and the face value is paid at maturity. For simplicity, we often speak of a bond's price, face value, and coupon payments as percentages of the face value. So a bond with a face value of $10,000, a coupon rate of 6.50%, and a price of $10,100 will usually be referred to as having a face value of 100, semiannual coupon payments of 3.25, and a price of 101.

The interest rate that equates the observed price of the bond with the present value of its future cash flows is the bond's yield. With semiannual coupon payments, a semiannual yield is easily calculated. By convention, the yield-to-maturity of the bond is its semiannual yield multiplied by 2.0 (this ignores compounding within each year, but everyone who deals in bonds realizes this). An investor will earn this yield if he holds the bond until maturity, if all the cash flows are paid in full and on time, and if the investor reinvests each coupon payment at that yield until the bond matures.

To price a bond, introductory finance courses teach students to discount the bond's face value and coupon payments at its semiannual yield. Thus, the bond's yield must be known in order to find its price. When students ask where we find this yield, we typically tell them that similar bonds will have similar yields, so you look at the yield for another bond that has the same credit quality (for Treasuries, that would mean another Treasury), and the same number of years till maturity. What this neglects however, is that the bond we are seeking to price probably has a different coupon rate from the (similar) bond that we are using to obtain a yield. While this subtlety may not be necessary to explore in an introductory finance course, it is important enough to include in any advanced course dealing with fixed-income analysis (FIA) and in-fact is critically important to bond traders who want exact prices of bonds with face values of $1 million or more.

In advanced courses that teach FIA, it is taught that a Treasury bond or note should be viewed as a series of independent cash flows and that each cash flow should be discounted to its present value at the discount rate that investors require for a single zero-coupon cash flow guaranteed by the U.S. Treasury. Currently, the treasury issues 6-month and 1-year Treasury bills – each of which offer a single cash flow at maturity. These bills are auctioned weekly. Thus, it is simple to look at the semiannual yield that can be earned on the most recently issued 6-month and 1-year bills to find the discount rates that investors require for any treasury-guaranteed cash flows occurring 6-months and 1-year from now.

The challenge starts when we try to find a single treasury-guaranteed cash flow that will be paid 18 months (1.5 years) from now. The yield on a Treasury note or bond that matures in 18 months reflects three cash flows – a coupon payment that comes in six months, another coupon payment that comes in one year, and a third coupon payment (plus the face value) that is paid in 18 months. Roughly speaking, the yield is an average of the discount rates that investors require on these three separate cash flows – not the discount rate for a single cash flow to be received in 18 months. However, if we have already found the semiannual yield that investors require for cash flows that the treasury promises in six months and one year from actively traded Treasury bills, and we observe the price investors are willing to pay for a coupon-paying Treasury note or bond that matures in 18 months, we have an equation with only one unknown – the discount rate that investors currently require for a treasury-guaranteed cash flow that is scheduled to be paid 18 months from now. Finding that rate in this manner is known as bootstrapping.

The Bootstrapping Process

The second edition of the fixed income text that is used in the CFA Institute Investment Series says, "The theoretical spot rates for Treasury securities represent the appropriate set of interest rates that should be used to value default-free cash flows. A default-free theoretical spot rate curve can be constructed from the observed Treasury yield curve" (Fabozzi, 2007, p. 135). What is referenced by Fabozzi as a spot rate curve, can also be called a zero-coupon yield curve (Petitt, Pinto, and Pirie, 2015). Teaching students how to construct such a curve and how it can be used to price both risk-free and risky bonds is the object of this exercise.

Once we have used the bootstrapping technique to find the rate that investors require for a single cash flow to be received in 18 months, we can find the rate they require for a single cash flow to be received in two years by observing the price and coupon payments for a Treasury bond or note that matures in two years. We use the 6-month, 1-year and 18-month discount rates that we have already found, and solve for the 2-year discount rate that equates the present value of these four separate cash flows with the observed price of the bond. This process can be carried on for up to 30 years since the U.S. treasury issues 30-year bonds. In our exercise, we carry out this bootstrapping process for ten years.

As an example, suppose we observe the following Treasuries, which are actively traded:

Maturity (years)	Coupon Rate (%)	YTM(%)	Price ($)
0.5	NA	8.0	96.15
1.0	NA	8.3	92.19
1.5	8.5	8.9	99.45
2.0	9.0	9.2	99.64
2.5	11.0	9.4	103.49
3.0	9.5	9.7	99.49
3.5	10.0	10.0	100.00
4.0	10.0	10.4	98.72
4.5	11.5	10.6	103.16
5.0	8.75	10.8	92.24
5.5	10.5	10.9	98.38
6.0	11.0	11.2	99.14
6.5	8.5	11.4	86.94
7.0	8.25	11.6	84.24
7.5	11.0	11.8	96.09
8.0	6.5	11.9	72.62
8.5	8.75	12.0	82.97
9.0	13.0	12.0	104.30
9.5	11.5	12.4	95.06
10.0	12.5	12.5	100.00

As mentioned earlier, the 0.5 and 1.0-year treasuries are Treasury bills and are sold at a discount with no coupon payments. Thus the six-month zero-coupon rate is 8.0%/2 = 4.0% and the one-year zero-coupon rate is 8.3%/2 = 4.15%. In this exercise, we call each of these spot zero-coupon rates "Zs" where Z_1 is the six-month zero-coupon rate, Z_2 is the one-year zero coupon rate, Z_3 is the 1.5-year (18-month) zero-coupon rate, etc. While Z_1 and Z_2 are easily calculated, $Z_3 – Z_{20}$ need to be bootstrapped from the data. To find Z_3, we look at the 1.5-year treasury with an 8.5% coupon rate and a price of 99.45. An 8.5% coupon rate means that each semiannual coupon payment will be 4.25% of the bond's face value. Since we are setting the bond's face value at 100, the coupon payments will 4.25. This price of 99.45 is the present value of the bond's three future cash flows with each cash flow being discounted at its market-determined zero-coupon rate:

$99.45 = 4.25/(1+Z_1) + 4.25/(1+Z_2)^2 + 104.25/(1+Z_3)^3$.

$99.45 = 4.25/(1.04) + 4.25/(1.0415)^2 + 104.25/(1+Z_3)^3$.

Therefore, $Z_3 = 4.465\%$ semiannually with a bond-equivalent-yield of 8.93%. We can extend this out to find Z_4. Here, the coupon rate is 9%, so the semiannual coupon payment will be 9/2 = 4.5.

$99.64 = 4.50/(1+Z_1) + 4.50/(1+Z_2)^2 + 4.50/(1+Z_3)^3 + 104.50//(1+Z_4)^4$.

$99.64 = 4.50/(1.04) + 4.50/(1.0415)^2 + 4.50/(1.04465)^3 + 104.50//(1+Z_4)^4$.

Therefore, $Z_4 = 4.624\%$ semiannually with a bond-equivalent-yield of 9.247%. If this is continued for all ten years of data, we find that $Z_{20} = 6.812\%$

While extending this process out all the way to Z_{20} can certainly be done, I'm sure the reader can see that it will take students quite a bit of time. Additionally, all the values used in this example are made-up. Obviously, these yields are much higher than we observe today. However, once we have all the Zs, we can price any Treasury with a maturity up to 10 years by discounting its cash flows as separate zero-coupon cash flows, discounted at the market-determined Zs that we have bootstrapped. Once we have a price for our bond, its yield-to-maturity can be found in the usual way by finding a single discount rate that equates that price with the present value of the coupon payments and face value. In the example above, a 10-year bond with a 10% coupon rate will be discounted at the Zs to a price of 85.35477. Its yield will be 12.618%.

Corporate bonds are typically priced off of a treasury that has the same maturity and coupon rate. Since corporate bonds are subject to the risk of default, their Zs (zero-coupon discount rates) are set at some value above the Zs of the treasury they are being priced off of in order to reflect this credit risk. So if a company is looking to issue a corporate bond with a 10% coupon rate (like the example above) and it is determined that investors will require an additional yield of 1.00% as compensation for its credit risk, that 1.00% needs to be added to $Z_1 – Z_{20}$ to come up with the appropriate discount rates for the corporate bond's cash flows. Of course, this will result in the corporate bond having a lower price than the otherwise similar treasury.

The Spreadsheet

Our spreadsheet allows students to price a 10-year corporate bond with whatever coupon rate and credit spread (over a comparable treasury) that the student (or instructor) chooses. It uses the bootstrapping procedure explained above with current treasury data. It does this through the use of a macro which is embedded in the spreadsheet. The macro goes to the Wall Street Journal's website to obtain the Journal's current Treasury bill, note and bond quotes. The WSJ website with the data is found at https://www.wsj.com/market-data/bonds/treasuries

To access our spreadsheet, students are instructed to go to https://breesefine7110.tulane.edu/bootstrapping-2/ From here, they can access the instructions which will tell them to download and run the WSJ executable file from Box. The Box site can be found directly at https://tulane.box.com/s/so124zgbuwkrkaigpai9esxiytnn4zvi

Once this executable file has been run, students are instructed to open the excel file titled "Bootstrapping-Spreadsheet" on the website and click on the "to get started, click here" button, the spreadsheet will immediately fill with the necessary current data obtained from the WSJ website. This allows students to learn how to bootstrap and construct a zero-coupon yield curve using current data without having to spend hours looking for the data and bootstrapping each "Z" by hand. Complete directions can be found in Appendix 1 – Instructions for Students.

If the executable file has been run, our spreadsheet can access the WSJ Treasury Quote data. The spreadsheet determines today's date through the "Today" function and from it, calculates dates which are six months, one year, 18 months, two years, etc. from today – all the way up to 10 years from today. When a student clicks on the blue "to get started, click here" button, the macro downloads the current Treasury data from the WSJ website and looks for the Treasuries that mature closest to those dates. If two different Treasury issues mature on the same date (or the same number of days from the six-month date), our algorithm selects the Treasury that is priced closest to par value. The 6-month and one-year selections will always be zero-coupon Treasury bills and the remaining 18 selections will be coupon-paying Treasury notes or bonds. For the Treasury bills, our spreadsheet reads the yield-to-maturity and calculates Z_1 and Z_2 from them. For the Treasury notes and bills, our spreadsheet reads their coupon rate and price – the data that is needed to bootstrap $Z_3 – Z_{20}$.

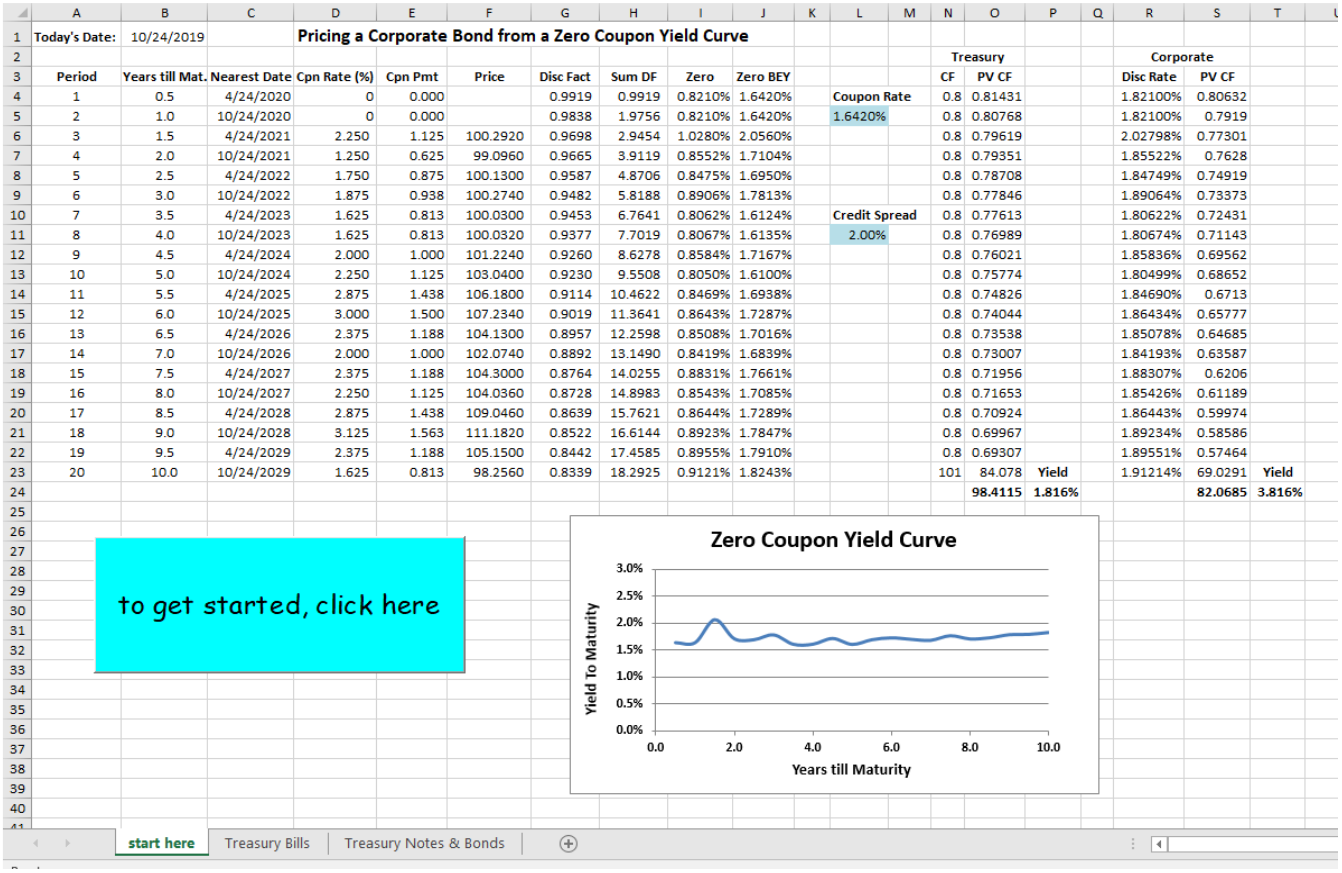
With all 20 Zs having been calculated, our spreadsheet can price any Treasury bill or note with a maturity of up to 10 years. By default, the spreadsheet prices a 10-year Treasury note, but students can easily modify it to price a treasury with a shorter maturity. Since students have been taught that the price (value) of a Treasury is the sum of the present values of its cash flows – with each cash flow being discounted at its market-determined zero-coupon rate – students only need to select a coupon rate to have the spreadsheet calculate the price of their desired Treasury. In addition to calculating this bond's price, our spreadsheet also calculates its yield-to-maturity.

To price a comparable corporate bond, students need to select a credit spread. The spreadsheet adds this credit spread to each of the Zs ($Z_1 – Z_{20}$) and discounts the cash flows of a bond with the selected coupon rate at these higher discount rates. The sum of those discounted cash flows gives the student the bond's price and its yield is calculated from that price.

One additional feature of the spreadsheet is that it graphs a Zero-Coupon Yield Curve from the Zs that it calculates. This allows instructors to discuss the term structure of interest rates with students while referring to a completely up-to-date yield curve.

We have found that this spreadsheet is best used after students have been taught how to bootstrap and have had a chance to bootstrap (by hand) one or two treasuries with fairly short maturities (perhaps two years). As with many time-saving math tools, the bootstrapping process is best understood before students are told that there is a spreadsheet that will do most of the work for them.

When complete, the spreadsheet will appear as on the following page.



Today's Date:	10/24/2019		Pricing a Corporate Bond from a Zero Coupon Yield Curve											Treasury			Corporate	
Period	Years till Mat.	Nearest Date	Cpn Rate (%)	Cpn Pmt	Price	Disc Fact	Sum DF	Zero	Zero BEY					CF	PV CF		Disc Rate	PV CF
1	0.5	4/24/2020	0	0.000		0.9919	0.9919	0.8210%	1.6420%		Coupon Rate			0.8	0.81431		1.82100%	0.80632
2	1.0	10/24/2020	0	0.000		0.9838	1.9756	0.8210%	1.6420%		1.6420%			0.8	0.80768		1.82100%	0.7919
3	1.5	4/24/2021	2.250	1.125	100.2920	0.9698	2.9454	1.0280%	2.0560%					0.8	0.79619		2.02798%	0.77301
4	2.0	10/24/2021	1.250	0.625	99.0960	0.9665	3.9119	0.8552%	1.7104%					0.8	0.79351		1.85522%	0.7628
5	2.5	4/24/2022	1.750	0.875	100.1300	0.9587	4.8706	0.8475%	1.6950%					0.8	0.78708		1.84749%	0.74919
6	3.0	10/24/2022	1.875	0.938	100.2740	0.9482	5.8188	0.8906%	1.7813%					0.8	0.77846		1.89064%	0.73373
7	3.5	4/24/2023	1.625	0.813	100.0300	0.9453	6.7641	0.8062%	1.6124%		Credit Spread			0.8	0.77613		1.80622%	0.72431
8	4.0	10/24/2023	1.625	0.813	100.0320	0.9377	7.7019	0.8067%	1.6135%		2.00%			0.8	0.76989		1.80674%	0.71143
9	4.5	4/24/2024	2.000	1.000	101.2240	0.9260	8.6278	0.8584%	1.7167%					0.8	0.76021		1.85836%	0.69562
10	5.0	10/24/2024	2.250	1.125	103.0400	0.9230	9.5508	0.8050%	1.6100%					0.8	0.75774		1.80499%	0.68652
11	5.5	4/24/2025	2.875	1.438	106.1800	0.9114	10.4622	0.8469%	1.6938%					0.8	0.74826		1.84690%	0.6713
12	6.0	10/24/2025	3.000	1.500	107.2340	0.9019	11.3641	0.8643%	1.7287%					0.8	0.74044		1.86434%	0.65777
13	6.5	4/24/2026	2.375	1.188	104.1300	0.8957	12.2598	0.8508%	1.7016%					0.8	0.73538		1.85078%	0.64685
14	7.0	10/24/2026	2.000	1.000	102.0740	0.8892	13.1490	0.8419%	1.6839%					0.8	0.73007		1.84193%	0.63587
15	7.5	4/24/2027	2.375	1.188	104.3000	0.8764	14.0255	0.8831%	1.7661%					0.8	0.71956		1.88307%	0.6206
16	8.0	10/24/2027	2.250	1.125	104.0360	0.8728	14.8983	0.8543%	1.7085%					0.8	0.71653		1.85426%	0.61189
17	8.5	4/24/2028	2.875	1.438	109.0460	0.8639	15.7621	0.8644%	1.7289%					0.8	0.70924		1.86443%	0.59974
18	9.0	10/24/2028	3.125	1.563	111.1820	0.8522	16.6144	0.8923%	1.7847%					0.8	0.69967		1.89234%	0.58586
19	9.5	4/24/2029	2.375	1.188	105.1500	0.8442	17.4585	0.8955%	1.7910%					0.8	0.69307		1.89551%	0.57464
20	10.0	10/24/2029	1.625	0.813	98.2560	0.8339	18.2925	0.9121%	1.8243%					101	84.078	Yield	1.91214%	69.0291	Yield
															98.4115	1.816%		82.0685	3.816%

We have also found it helpful to show students the algebra behind the spreadsheet calculations of the Zs prior to giving them access to the spreadsheet, but after working through a couple of bootstrapping examples by hand. Here is how we explain it to them:

The term "discount factor" means $\frac{1}{1+z_1}$ or $\frac{1}{(1+z_2)^2}$ or $\frac{1}{(1+z_3)^3}$ etc.

So "sum of the discount factors" $= \frac{1}{1+z_1} + \frac{1}{(1+z_2)^2} + \frac{1}{(1+z_3)^3}$ etc.

If we have a bond maturing in 1.5 years (3 semiannual periods from now):

$$P = \frac{C}{1+z_1} + \frac{C}{(1+z_2)^2} + \frac{C+100}{(1+z_3)^3}$$

Which can be rearranged as:

$$(1+z_3)^3 = \left(\frac{C+100}{P - \left(\frac{C}{1+z_1} + \frac{C}{(1+z_2)^2} \right)} \right)$$

and

$$z_3 = \left(\frac{C+100}{P - \left(\frac{C}{1+z_1} + \frac{C}{(1+z_2)^2} \right)} \right)^{1/3} - 1$$

If we pull out the Cs in the denominator, we get:

$$z_3 = \left(\frac{C+100}{P - C\left(\frac{1}{1+z_1} + \frac{1}{(1+z_2)^2} \right)} \right)^{1/3} - 1$$

Notice the sum of the discount factors in the denominator.

At this point, instructors who want to give students a more complex excel-task can assign them to design a spreadsheet using this format that will bootstrap a bond with as many semiannual periods as desired – by using only one row for each semiannual period. Of course, this is exactly what our spreadsheet does, so this task should only be assigned if the students have not yet seen the spreadsheet.

Once students understand the terms "discount factor" and "sum of the discount factors" as used above, they can more easily see how our spreadsheet calculates the Zs in column I by first calculating each discount factor in column G and the sum of the discount factors in column H.

Appendix 1 contains the instructions we give students – explaining where to go to find the spreadsheet and exactly what steps need to be taken to access it and get the macro to download the current WSJ Treasury quotes. It is important to note that this spreadsheet was designed to be used with PCs. Additionally, the computer using this spreadsheet must be set up to allow macros and be connected to the internet.

Appendix 2 lists the questions that we typically assign students to answer for this exercise. While questions 1-3 merely require the students to place the correct values in the cells with the blue highlights, question 4 requires students to go a bit further. Since any bond which is selling at par value (100) has a coupon rate equal to its yield-to-maturity (and thus its semiannual coupon rate equals its semiannual yield), solving for the YTM is essentially solving for the coupon rate. Since we know the Zs and the price of the bond (100), the coupon payments are the only unknown in our equation. Doing this for a series of different maturities will allow students to construct a par bond yield curve. Sundaresan (2002) describes how a par bond yield curve is frequently used in industry.

We use questions 5 and 6 in appendix 2 to see if students can apply their understanding of different theories of the term structure to today's yield curve. Of course, these questions should only be included if these concepts have already been covered in class or in assigned readings.

Conclusion

Advanced courses in Fixed Income Analytics teach students that bonds should be priced as a series of zero-coupon cash flows, with each cash flow being discounted at its own separate zero-coupon rate. Bootstrapping is the process through which we find those zero-coupon rates. While it is important to show students how the Bootstrapping process works, they can understand it much more completely through experiential learning where they get an opportunity to bootstrap a zero-coupon yield curve using current data without spending several hours doing it by hand.

Our spreadsheet allows students to access current Treasury bill, note and bond quotes from the Wall Street Journal website and use that data to bootstrap current and accurate zero-coupon rates, which are used to price a 10-year Treasury note with whatever coupon rate the student choses. Students can also price a 10-year corporate bond with a defined credit spread to the comparable treasury. Finally, the current zero-coupon yield curve is graphed, allowing for further discussion on the term structure of interest rates. We have found that using this spreadsheet in class and in assignments has contributed to our students' understanding of the Bootstrapping process in the study of fixed income analytics.

References

Fabozzi, Frank J. 2007. *Fixed Income Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.
Petitt, Barbara S., Jerald E. Pinto, and Wendy L. Pirie. 2015. *Fixed Income Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.
Sundaresan, Suresh. 2002. *Fixed Income Markets and Their Derivatives*. Cincinnati, OH: South-Western

Appendix 1 – Instructions for Students

1. Go to https://tulane.box.com/s/so124zgbuwkrkaigpai9esxiytnn4zvi or just click on the "Link to Bootstrapping folder in Box" on the website.
2. Click the Download button
3. Select "Run"
4. If it says that this download could harm your computer, select "Actions"
5. If given the choice, select "Run Anyway". If that choice isn't available, select "More Options", and then select "Run Anyway".
6. A black window might appear and then disappear a bit later – that is normal.
7. Double click on the "Bootstrapping Spreadsheet" from the class website (just below "Questions to Answer").
8. Click "Save As" and save it to your desktop (or wherever you want), and then open the spreadsheet
9. Enable Editing and Enable Content if requested by the spreadsheet.
10. Verify that today's date appears in cell B1
11. Click on the large blue button that says, "to get started, click here".
12. As soon as you click on the blue button, a macro in the spreadsheet will go to the Wall Street Journal's website and extract the current data for Treasury Bill, Treasury Note and Treasury Bond quotes. This data will show up in the worksheets titled "Treasury Bills", and "Treasury Notes & Bonds".
13. The worksheet titled "start here" will read data in the other two worksheets and insert the proper data in cells J4-J5, D6-D23, and F6-F23.
14. With this current data inserted into the spreadsheet, Z_1 – Z_{20} will be calculated, a zero-coupon yield curve will be bootstrapped and graphed, and the price of a theoretical 1.488% 10-year Treasury Note will appear in cell O24. Its Yield-to-Maturity will be in cell P34.
15. If you want to find the price and yield of a theoretical 10-year Treasury Note with a different coupon rate, just change the value in cell L5 (blue highlight).
16. To price a Corporate bond, select a credit spread and insert it into cell L11 (blue highlight). The price of this bond will appear in cell S24. Its yield will be in cell T24.
17. Take a close look at the Excel formulas that were used to obtain all these values. You should be able to follow and understand each of them.

Appendix 2 – Suggested Questions for Students

1. What is the price of a bootstrapped 2.4% (coupon rate) 10-year Treasury note?
2. What is the YTM of the Treasury note in question 1?
3. Price a 10-year 2.4% corporate bond so that it has a 50- basis point credit spread over the Treasury note you bootstrapped.
4. Using the zero-coupon rates (the semiannual z values) that that were found in the bootstrapping spreadsheet, find what the YTM would be for a previously-issued Treasury bond that matures in exactly two years and is currently selling at par value.
5. Look at the Zero Coupon Yield Curve that the bootstrapping spreadsheet made. How would you describe the shape of this yield curve?
6. Based on the Unbiased Expectations Theory of the term structure, what does this yield curve tell us about investors' expectations over the next ten years?

Canadian Hockey Leagues Game-to-Game Performance

Nick Riccardi, Syracuse University

Abstract

This study examines game-to-game performance of players across the three Canadian Hockey Leagues (Western Hockey League, Ontario Hockey League, and Quebec Major Junior Hockey League) for the 2017-2018 season. It tests the importance of factors such as rest, travel, weather conditions, and more. Data for this study were collected from each of the three CHL websites and from www.weatherunderground.com. The null hypotheses of different factors affecting performance were tested through regression models using Ordinary Least Squares. The dependent variables, used across different specifications, were on-ice performance variables such as points, goals, and penalty minutes on a per-game basis.

Introduction

The Canadian Hockey League (CHL) consists of three major junior hockey leagues for 16 to 20 year old players. The Western Hockey League (WHL), Ontario Hockey League (OHL), and Quebec Major Junior Hockey League (QMJHL) serve as platforms for players hoping to reach the major leagues. As is the case for all professional-level sports leagues, the players in these leagues have a number of outside factors that may impact their performance, yet they have no control over. For players who are trying to prove themselves to National Hockey League Executives, some factors can significantly affect their performance on the ice.

Some of these outside factors that can have an impact on performance include home ice advantage, the amount of days between games, the distance they must travel for each game, and weather effects that can affect ice conditions. These factors can have both positive and negative effects on performance, and can also affect different players in different ways. Considering the differences between the WHL, OHL and QMJHL, these factors can also have different effects on player performance within each league compared to each other. For example, the Western Hockey League covers far more land than the other two leagues, so factors like days of rest and travel distance will most likely affect players differently. In fact, George Johnson describes life as a WHL player with the childhood song "The Wheels on the Bus" and emphasizes that sometimes teams will be on the bus for 15 hours at a time (Johnson, 2010).

The most pronounced outside factor in terms of its impact on performance is playing games at home. In all sports, to some degree there is an advantage of playing in one's home facility rather than playing on the road. The advantage is the greatest in hockey, as per 100 games played at home, the home team on average wins 53 games, loses 30, and ties 17. Excluding ties this would mean home teams win 64 percent of their games (Schwartz & Barsky, 1977). Furthermore, the National Basketball Association (NBA) offers a clear assessment of the advantage of playing at home, due to crowd effect, because the Los Angeles Lakers and Los Angeles Clippers share a stadium. When the two teams play each other, the "home" team has their win likelihood raised by an estimated 21-22.8 percentage points (Boudreaux, Sanders, & Walia, 2017).

In order to run a profitable business, the CHL schedules games to produce the highest attendance possible, which often results in three games in four days. Playing night after night does not necessarily bode well for players trying to perform at their highest level. Inversely, having long stretches of days without playing can also negatively affect performance, as players often go on hot streaks throughout the season or can get "rusty" through lack of actual game action. Prior research conducted on the NBA on rest and travel effects have shown that players perform worse when they only have one day between games, while having more than one day in between games improves performance. Additionally, peak performance comes with three days in between games, while performance begins to tail off with four or more days. Travel was also seen to have an effect on performance, as visiting teams were affected by days of rest more than home teams (Steenland & Deddens, 1997).

The goal of this research is to identify which outside factors significantly affect player performance for major junior ice hockey players and in which way. Furthermore, an attempt is made to consider these leagues in the scope of a labor market and how workers are affected by the scheduling of their work. Through a method of ordinary least squares regression, an analysis of these factors' effects on performance is conducted for not only the CHL as a whole, but also each league individually.

Empirical Model

I hypothesize that playing at home, the amount of days between games, travel, and ice conditions caused by weather effects have an impact on player performance. Therefore, points, goals, assists, and even penalty minutes can be affected by these outside factors. In order to test the significance and the effect that these factors have on performance, I ran four regressions with these factors as independent variables, and each of the four performance variables as the dependent variables. The points

and penalty minutes models will be shown in the body of the paper as Table I and Table II, while the goals and assists models did not have as conclusive results. The data consists of the game logs from the top three hundred scorers from each of the three CHL leagues. Weather data from www.weatherunderground.com was matched with these game logs.

In the models, "playerhome" represents whether the player was playing at home or on the road, where the value "0" is for a road game, and "1" is for home. The "daysbetween" variable is for how many days between the current game and the last game they played. Given the offseason, holiday breaks, and players coming off of injuries, the maximum for this variable was restricted to 10 days. The distance measured in miles from a player's home city to the road city in which they are playing in is represented by the "travel" variable. Lastly, "temp," "hum," "precip," and "pres" are used as weather variables, and are temperature, humidity, precipitation, and pressure, respectively. These models also included squared variables for both "daysbetween" and "travel" to test for a nonlinear relationship between performance and the aforementioned variables. All models are fixed for both player and opponent effects.

The main focus will be on the model which uses points as a dependent variable, as this is the best all-encompassing variable for player performance. The results of this model are in the table below. Coefficients on the independent variable and the p-value associated with that coefficient is presented for each league and for the pooled sample (including all three leagues).

Table 1: Regression Results CHL Player Performance

	OHL	***OHL***	***QMJHL***	***QMJHL***	***WHL***	***WHL***	***CHL***	***CHL***
Variable	***Coefficient***	***Prob.***	***Coefficient***	***Prob.***	***Coefficient***	***Prob.***	***Coefficient***	***Prob.***
Intercept	2.836	*0.013*	-0.031	*0.957*	-0.919	*0.242*	0.216	*0.584*
Player Home	0.090	*0.032*	0.128	*0.000*	0.103	*0.000*	0.114	*0.000*
Days Between	0.036	*0.002*	0.033	*0.000*	-0.006	*0.492*	0.022	*0.000*
Days Between^2	-0.004	*0.001*	-0.003	*0.000*	0.001	*0.301*	-0.003	*0.000*
Travel	-0.000	*0.383*	0.000	*0.060*	0.000	*0.699*	0.000	*0.046*
Travel^2	0.000	*0.228*	0.000	*0.158*	0.000	*0.885*	0.000	*0.125*
Temperature	0.000	*0.410*	-0.000	*0.246*	0.001	*0.070*	0.000	*0.473*
Humidity	-0.001	*0.069*	-0.000	*0.638*	0.001	*0.063*	-0.000	*0.576*
Precip.	-0.064	*0.323*	0.058	*0.016*	0.051	*0.343*	0.036	*0.035*
Baro. Pressure	-0.039	*0.294*	0.017	*0.351*	0.046	*0.071*	0.011	*0.403*
QMJHL							-0.128	*0.000*
WHL							0.050	*0.000*
Player Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opponent Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

The table is constructed to show how each of the variables affect each league and the CHL as a whole. Taking a broad look at all three leagues together, there are several variables that are significant. First, and most obvious, is that playing at home is significant at the 1% level with a positive coefficient of .114. Furthermore, both "daysbetween" variables are significant at the 1% level, which signifies that the relationship between points and the amount of days between games is nonlinear. Understanding the true coefficient for "daysbetween" will be discussed later. The travel variable was also significant, at the 5% level, but with such a small coefficient, nothing can be said about its effect on performance without looking at it visually. Lastly, precipitation was significant at the 5% level as well, with a small but positive coefficient. For the CHL, the OHL was used as a dummy variable to look for a difference between leagues, which was found for both of the other leagues at the 1% level. Players in the WHL scored more points than the OHL, while players in the QMJHL scored less.

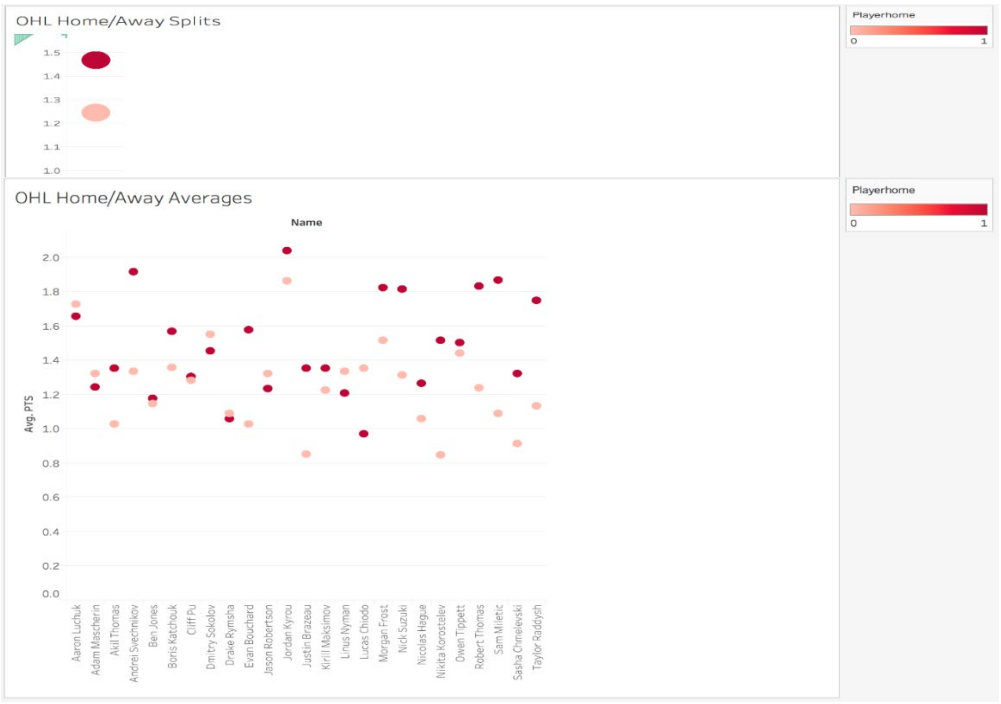
Before moving on, a quick look at the OHL, QMJHL, and WHL models show differences between some variables. First, unlike the other two, the WHL didn't have significance for either of the "daysbetween" variables, while the QMJHL was the only league with significance for travel. As far as the weather variables go, the OHL found significance at the 10% level for humidity, with a negative effect, while the QMJHL saw significance at the 5% level for precipitation with a positive effect. The WHL had significance at the 10% level and positive coefficients for all four weather variables besides precipitation.

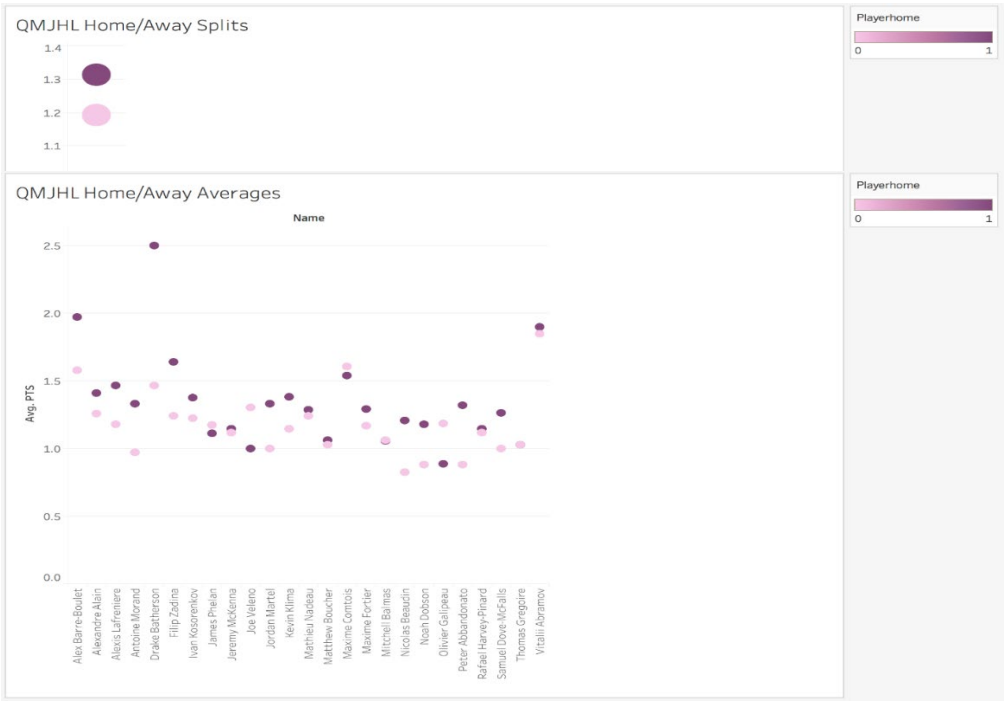
Impact of Playing at Home

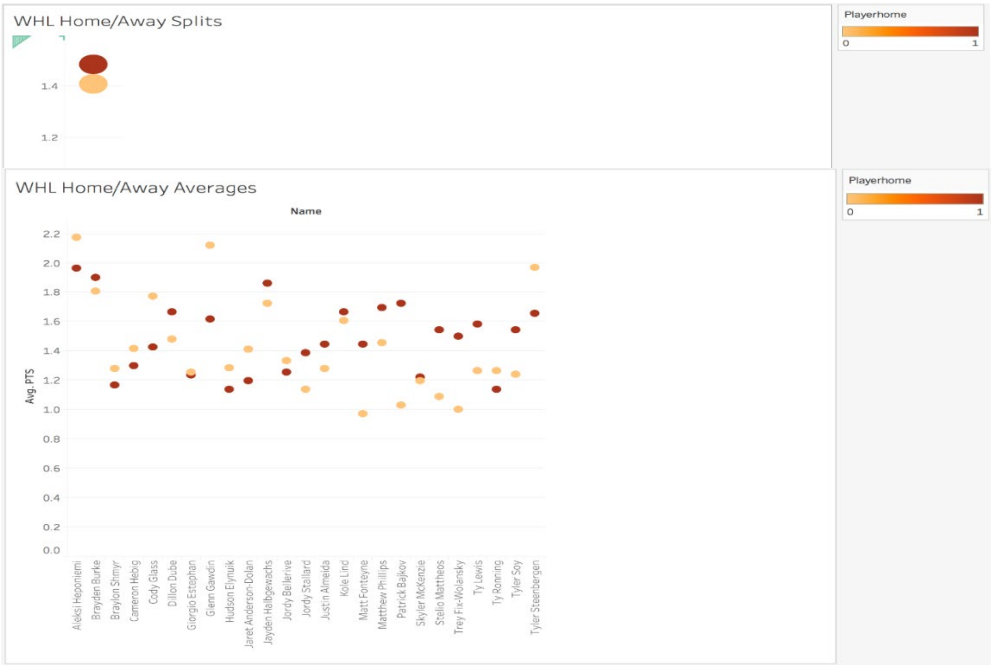
It comes as no surprise that the biggest effect out of the variables tested came from a home ice advantage. All three leagues saw significance with a positive effect. A look at how home ice advantage contributes to scoring points can be seen in the graphs below that include the top 25 scorers from each of the three leagues. The choice to use the top 25 scorers rather than the entire sample size was decided by the fact that the data has too many players to look at as a whole, and the top scorers in theory should perform at a high level regardless of where they are playing.

The first graph for each league shows the combined home and away point averages for the top 25 scorers. The darker circle in each graph represents the home average while the lighter circle is for on the road. In all three leagues, averages at home were

higher than on the road. For the second graph for each league, the top 25 scorers are labeled on the x-axis with average points on the y-axis. The two circles for each player represent their averages at home and on the road, with the darker circles representing being at home and the lighter on the road. The OHL and QMJHL show that out of their top 25 scorers, 18 and 19 of them, respectively, score better at home. The difference between playing at home and on the road is less noticeable for the WHL, as only 14 of the top 25 scorers produce more at home. Overall, there is a clear advantage to playing at home rather than playing on the road when it comes to performance.







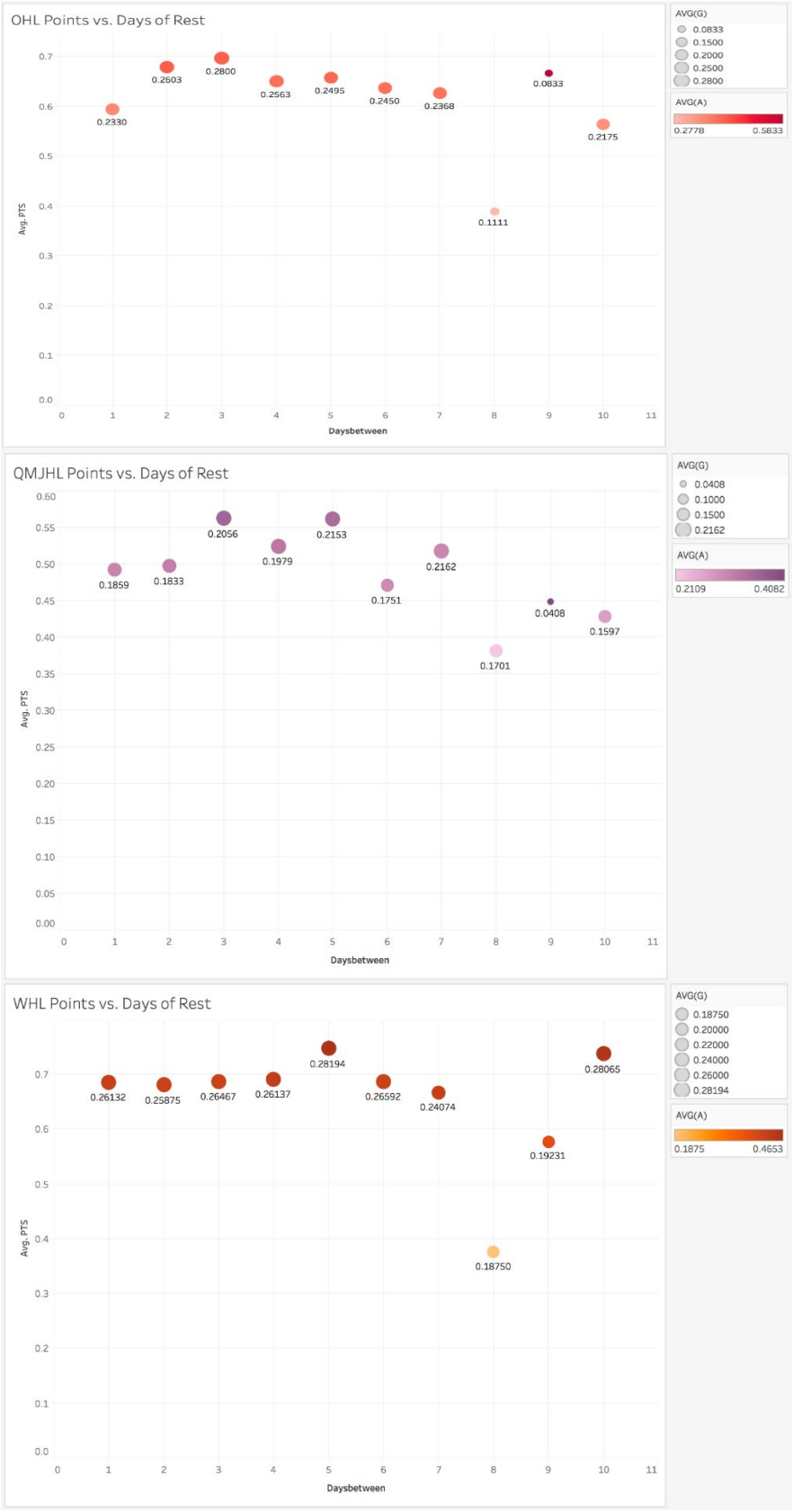
Days in Between Games

In order to assess the impact that the amount of days between games has on performance, based on the regression results, we must first adjust the overall coefficient. We do this by first multiplying the coefficient for "daysbetween" by the actual amount of days between and then adding it to the product of the "daysbetween" squared coefficient and the actual amount of days between. After doing this, we get the true coefficient for each amount of days between, which is represented in the graph below which shows the coefficient for the CHL as a whole.

The downwards concavity of this graph shows us that peak performance comes around four to five days between games. If we did not include the "daysbetween" squared coefficient and then find the true coefficient, we would have been left believing that as the amount of days in between games is increased, performance increases. This is clearly not the case, as the coefficient decreases after five days between games.

It is important to point out that the OHL and QMJHL coefficients produced graphs similar to the one above, but the WHL, which, individually, did not have significance for the "daysbetween" variables, showed a different relationship. However, because the variables were not significant, the relationship is not necessarily a perfect picture of how these variables affect performance. The graph for the true coefficient of the WHL is show below.

A look at actual point averages for each amount of days between games supports the curves for the OHL and QMJHL, while the WHL somewhat reflects the same pattern. Individual looks at each league are shown below.

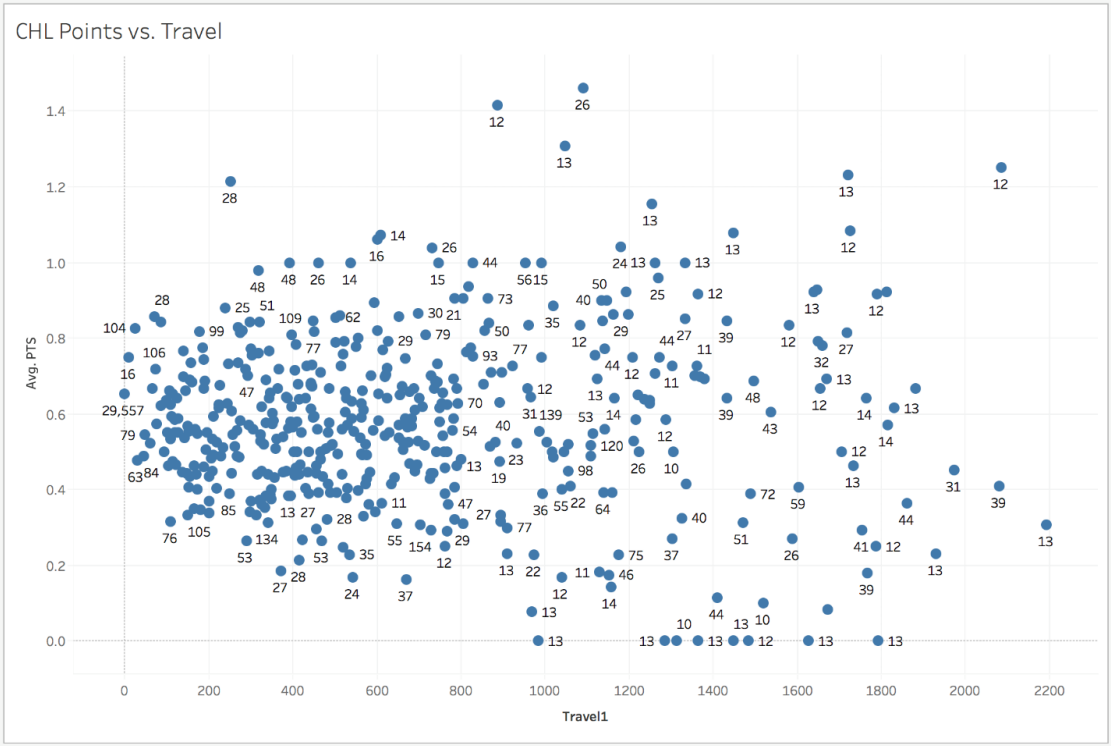

OHL Points vs. Days of Rest
Avg. PTS
Daysbetween
0.2330
0.2603
0.2800
0.2563
0.2495
0.2450
0.2368
0.0833
0.2175
0.1111
AVG(G)
0.0833
0.1500
0.2000
0.2500
0.2800
AVG(A)
0.2778
0.5833
QMJHL Points vs. Days of Rest
Avg. PTS
Daysbetween
0.1859
0.1833
0.2056
0.1979
0.2153
0.1751
0.2162
0.0408
0.1597
0.1701
AVG(G)
0.0408
0.1000
0.1500
0.2162
AVG(A)
0.2109
0.4082
WHL Points vs. Days of Rest
Avg. PTS
Daysbetween
0.26132
0.25875
0.26467
0.26137
0.28194
0.26592
0.24074
0.28065
0.19231
0.18750
AVG(G)
0.18750
0.20000
0.22000
0.24000
0.26000
0.28194
AVG(A)
0.1875
0.4653

Each of these graphs show the amount of days in between on the x-axis with average points on the y-axis. The size of each data point and the corresponding number under them represents the average number of goals scored, while the color range represents the average number of assists, with the darker the color representing a higher average. Each league shows similar trends but with different peaks. OHL players perform best with three days in between, having their highest average of both points and assists with this amount. QMJHL players have almost identical peak average points with three or five days in between, with slightly higher assists with three days and slightly higher goals with five days. Lastly, the WHL has peaks at both five and 10 days in between games. Points, goals, and assists are all slightly higher with five days than 10. Interestingly, both the OHL and QMJHL see their smallest, darkest circles with nine days in between games.

Inconsistencies from Traveling

In the model, travel was significant at the 5% level, but with such a small coefficient, the effects of travel were hard to interpret. Graphing miles traveled against average points tells the story of how travel affects performance. The relationship is show in the figure on the following page.

With distance (in miles) traveled on the x-axis, average points scored when traveling those distances on the y-axis, and the count of how many times that distance is traveled (the numbers under the data points), a clear relationship develops. Distance traveled doesn't necessarily provide a positive or negative effect on performance, but rather brings about inconsistency. As the data points move to the right in the graph above, they become more spread apart than they once were. Not only are there far worse performances the farther teams travel, but there are also far better ones. Although not a perfect analogy, this phenomenon might be explained by the Alchian-Allen Theorem which states that as transportation costs increase, consumption shifts toward higher-grade products (Potts, 2014.) In this case, as a possible explanation of improved performance, the further a player has to travel for a game, the more effort they may put into this game, possibly due to being exposed to different professional scouts in that region, compared to their own.



Penalty Minutes

Considering points can be used to evaluate performance and how it can be affected by outside factors, we will use penalty minutes to evaluate aggressiveness in the sport. Using the same independent variables and penalty minutes as the dependent variable, we get the following results for our model.

Table 2: Penalty Minutes Regression Results

Variable	***OHL Coefficient***	***OHL Prob.***	***QMJHL Coefficient***	***QMJHL Prob.***	***WHL Coefficient***	***WHL Prob.***	***CHL Coefficient***	***CHL Prob.***
Intercept	-0.859	*0.493*	2.177	*0.058*	-0.572	*0.751*	1.486	*0.042*
Player Home	-0.043	*0.320*	-0.006	*0.893*	-0.262	*0.000*	-0.081	*0.001*
Days Between	-0.020	*0.115*	0.012	*0.470*	-0.073	*0.001*	-0.025	*0.008*
Days Between^2	0.002	*0.206*	-0.000	*0.827*	0.009	*0.000*	0.003	*0.001*
Travel	-0.000	*0.635*	0.000	*0.737*	-0.001	*0.001*	-0.000	*0.012*
Travel^2	0.000	*0.725*	0.000	*0.519*	0.000	*0.001*	0.000	*0.010*
Temperature	0.002	*0.005*	-0.001	*0.144*	0.003	*0.001*	0.001	*0.000*
Humidity	-0.001	*0.472*	0.001	*0.195*	-0.002	*0.064*	-0.000	*0.575*
Precip.	0.043	*0.507*	-0.006	*0.852*	0.409	*0.001*	0.003	*0.930*
Baro. Pressure	0.036	*0.377*	-0.047	*0.212*	0.059	*0.319*	-0.028	*0.247*
QMJHL							0.153	*0.000*
WHL							0.244	*0.000*
Player Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opponent Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

These results show that playing at home is significant at the 1% level with a negative effect on penalty minutes. Furthermore, both "daysbetween" and travel variables are significant, with all but travel being significant at the 1%. Lastly, temperature positively affects penalty minutes at the 1% level. Perhaps higher temperatures lead to more aggressive behavior resulting in more penalties.

Considering that "daysbetween" squared is significant, the effects are nonlinear. After finding the true coefficient for the amount of days in between games, it shows the opposite effect than it does with performance. Having three to five days in between games shows lowest numbers in penalty minutes, while not having much rest increases penalty minutes, and having extended days off greatly increases aggressiveness. This is most likely due to having an "itch" to get back out and play.

Conclusions

Junior hockey players in the Canadian Hockey League are constantly trying to prove themselves to National Hockey League teams in hopes they can make it into the league. By not making it to the NHL, players are either tasked with playing in minor leagues or overseas, which both have significant pay drop-offs in comparison to the NHL, or giving up hockey all together. With this in consideration, it is vitally important for players to perform at their peak ability in order to show what they can bring to a team. However, there are many outside factors that are out of their control that can affect their performance.

For most players, playing at home is usually beneficial. Not having to travel means players most likely feel fresher, and playing in front of one's home crowd most likely motivates most players. Playing at home has a significant positive effect in the OHL, QMJHL and WHL for points, goals and assists. Although playing at home seems to have less of an impact in the WHL than the other two CHL leagues, it is still regarded as a positive.

In addition to having home ice advantage, another factor that affects player performance is the amount of days in between games. As a whole, peak performance throughout the CHL comes with somewhere between three and five days between games. Only having one or two days between games can lead to players not having enough rest for games. Inversely, having six or more days between games can lead to poorer performance due to rust and the streaky nature of the game of hockey. With not a lot of rest and too much rest causing a decrease in production, performance as it relates to days in between games has an umbrella-like relationship.

Travel also plays a role in how players perform. Although traveling far distances doesn't really have a positive or negative effect, there is still an impact on performance. As travel distance increases, player performance tends to become less consistent. Average point totals tend to be either very high or very low, and it might just be that some players benefit and others play worse when having to travel.

As far as weather effects go, my research did not find anything that had that great of an impact on performance. However, other work conducted on the matter has. Northern cities tend to have higher quality ice conditions due to the colder temperatures and less humidity. This results in "fast" game conditions. On the other hand, warmer, more humid cities produce "slow" ice. These differences can have positive and negatives effects on players with different playstyles and can also lead to teams taking different strategies depending on where they are playing (Conetta, Weinberg, & Paul, 2016). In terms of my research, I did find temperature to have a positive effect on penalty minutes, which means that players tend to be more aggressive when temperatures are higher.

Overall, players playing at the major junior level of hockey have a lot to prove to NHL teams. The opportunity cost of not making it to the majors is substantial. There are many outside factors that can contribute to their performance on the ice and therefore affect their chances of being able to compete against the best players in the world.

References

Boudreaux, C. J., Sanders, S. D., & Walia, B. (2017). *A Natural Experiment to Determine the Crowd Effect Upon Home Court Advantage*. Journal of Sports Economics, 18(7), 737-749.

Conetta, C., Weinberg, J., & Paul, R. (2016). *Impact of Ice Quality on Team and Player Performance in the NHL*. Lecture presented at MIT Sloan Sports Analytics Conference, Cambridge, MA.

Johnson, G. (2010, May 4). Back On The Bus: Life in the WHL. Retrieved January 19, 2019, from http://tigershockey.com/back-on-the-bus-life-in-the-whl/

Potts, J. (2014). The Alchian-Allen Theorem and the Economics of Internet Animals. Retrieved January 19, 2019, from http://journal.media-culture.org.au/index.php/mcjournal/article/view/779

Schwartz, B., & Barsky, S. F. (1977). *The Home Advantage.* Social forces, 55(3), 641-661.

Steenland, K., & Deddens, J. A. (1997). *Effect of Travel and Rest on Performance of Professional Basketball Players.* Sleep, 20(5), 366-369.

Female Employment and Public Policy Impact on International Fertility

Yassaman Saadatmand, Georgia Southern University
Lainey Sapp, Georgia Ports Authority

Abstract

Over the past 50 years, fertility rates in developed countries have significantly decreased. In many cases, fertility rates are now below population replacement level, leading to long term economic challenges such as an increased dependency ratio and a decreased support ratio for social programs. As a result, governments have turned to incentivizing child-bearing to increase fertility rates. Three common subsidies are maternity- and parental-leave benefits, childcare subsidization, and family allowances. The effect of these three government subsidies on fertility rates is explored via multivariable regression analysis in data from 21 OECD countries between the years of 1982–-2013. The results indicate subsidies increase fertility rates with varying levels of magnitude. Maternity- and parental- leave benefits were seen to have a stronger impact on fertility rates implying they are more powerful policy tools for increasing fertility rates.

Introduction

Over the past half a century, fertility rates in developed countries have been falling. The primary reason for this is improvements in medical technology that have decreased infant mortality rates, allowing more children to reach adulthood. As a result, females have fewer children. Another cause for the decrease is a large increase in female labor force participation. As females have entered the labor force in mass numbers, the opportunity cost of having children has significantly increased. More recent evidence, however, suggests that female employment now increases fertility rates because females attain higher levels of affluence and can more easily bear the cost of children. (McNown 2003; Goldstein and Örsal 2010).

The drop in fertility rates is a serious issue for many developed countries. Their rates have fallen below the rate required to maintain the population level, potentially causing significant negative impacts on these countries' economies. As a result many governments have taken actions to raise fertility rates through public policies that incentivize child-bearing, such as maternity- and parental-leave benefits, childcare subsidization, and family allowances.

This leads to a two part question of whether female employment increases fertility rates, and whether public policies can also significantly increase fertility rates. The current study's empirical results suggest that both female employment and public policies have a significant impact on fertility rates. These results have major implications for governments in developed countries with low fertility rates. It suggests that public policies may be able to increase fertility rates, and thereby increase long-run potential GDP in these countries.

Literature Review

In the 20th century the world's population increased greatly, from 2 billion in 1927 to 6 billion in 1999 (Fitzgerald, B 2011.) This has been the byproduct of a worldwide decrease in the mortality rate caused by advancements in medical technology. However, in the recent decades the fertility rates in most developed countries have been falling. Some researchers have discussed the importance of government support for maintaining an appropriate rate of fertility in developed countries. Jinno & Masaya (2016), for instance, argue that, because of active child care policies in France and Sweden, the fertility rates are appropriately high, but in Japan, because child care services are not sufficiently supported, the fertility rate remains low. On the other hand, a body of literature connects the declining fertility in developed countries to the existence of the welfare state. For instance, De Jager (2013) shows that extensive social welfare policies negatively affect fertility rates. Furthermore, De Jager (2013) states that the disaggregate effects of four social welfare policies—public education, unemployment benefits, "old-age" and sickness benefits—all negatively affect fertility rates. Similarly, Evan, Tomáš & Vozárová, Pavla (2018) discuss extensively the negative influence of the welfare state policies, in particular its pension system, on the declining fertility rates.

In developed countries, besides the existence of a welfare state, other factors influence fertility rates. .For instance, in many of these countries, females have higher education and are in the labor force. Based on the New Homes Economic Theory, one of two effects can take place, either a substitution effect or an income effect. The substitution effect claims that, as females become employed, their opportunity cost of having children will rise (Goldstein & Örsal 2010; McNown 2003). The costs could consist of forgone wages from staying home to care for children, slower growth in wages caused by forgone experience,

and the risk of a job loss (Adsera 2003). For women with higher education, the opportunity cost will be even higher since higher education levels tend to come with higher wages. With a higher opportunity cost associated with having children, many females will delay having children until later in life and will have fewer children as a result of this postponement. This postponement has been seen in Organization for Economic Co-Operation and Development (OECD) country data, with the mean female age at first birth being 23.8 in 1970 and 27.2 in 2000. With females having children later in life, postponements are not likely to recuperate. This means they are more likely to have fewer children than they otherwise would have or they may not have any at all. Childlessness in OECD countries increased by one third among thirty year olds from 1960 to 1970, suggesting that full postponement recuperation does not take place (d'Addio & d'Ercole 2005). The income effect, on the other hand, claims that, as females become employed, they will have a greater ability to meet the expenses of more children. This will cause female employment to increase fertility rates (Goldstein & Örsal 2010).

Evidence of both of these effects are found in the literature. Before the mid- to late-1980s, the majority of findings suggested that the substitution effect was dominant. Butz and Ward (1979) found a robust negative relationship between female employment and fertility rates in U.S data from 1948 to 1975. Similar findings were seen in UK data from 1950 to 1985 (Ermisch 1988). In research published since the mid to late 1980s, the relationship between female employment and fertility rates reversed with the income effect becoming dominant. McNown (2003) found a robust, positive relationship between female employment rates and fertility rates in U.S. data in the 1990s. Similar results were found by Goldstein and Örsal (2010) in panel data for 22 OECD countries from 1986 to 2008. The reversal in the impact female employment has had on fertility rates is likely due to less relevance in the traditional model of the male being the breadwinner and female being the primary care giver (d'Addio & d'Ercole 2005).

Even though the income effect appears to now overshadow the substitution effect, the income effect has not been strong enough to offset other factors from decreasing fertility rates. The significant decrease in developed countries' fertility rates that has taken place over the past 50 years has led many countries' average fertility rates to fall below the population replacement rate of 2.1 (Adsera 2003). As the average female labor force participation rate increased, from 48% in 1975 to 64% by the late 1990s, the average fertility rate has decreased, from 2.9 in 1960 to 1.6 in the late 1990s (Adsera 2003).

Some economists argue that having a smaller population will be beneficial for these countries, allowing for more well-off countries in the long-run with a cleaner environment and more natural resources per person. Though the vast majority of economists claim that declines in fertility will have negative short-run and potentially long-run impacts that will undermine any potential benefits, they state that having both low mortality rates and low fertility rates will decrease the country's support ratio with workers not being able to produce as much as society wants to consume. It will also lead to an increase in the countries' dependency ratio, placing a greater financial burden on the working-age population to support retirees (Bloom, Prettner, & Strulik 2012). This large retired population will eat up much of the public budget with health care costs and pension funding, leading to an unlikely decrease in aggregate demand (d'Addio & d'Ercole 2005). It will lead to the crowding out of other programs. This crowding out makes it unlikely that this smaller working-age population will have a higher quality of productivity. Since productivity is unlikely to be higher and demand is unlikely to be lower, it leaves nothing to offset the decrease in aggregate supply. This makes it likely that low fertility rates will lead to economic stagnation.

As fertility rates in OECD countries have dropped, government officials have started to show concern. A 1979 survey shows that officials in 20 of the OECD countries believed that fertility rate levels were satisfactory, while officials in only four of the OECD countries believed that the rates were too low. In 2003, when fertility rates were significantly lower in many of the OECD, the surveys showed much different results. Officials in only 12 countries believed rates were satisfactory, while officials in 16 countries believed rates were too low. Government officials' pessimistic views of potential negative impacts associated with low fertility rates have led many to craft public policies that incentivize higher fertility rates (d'Addio & d'Ercole 2005).

These policies are meant to reduce the financial burden of having children through reducing the opportunity cost of having children. Family allowance policies directly subsidize the cost of having children. These are typically in the form of government subsidies, with monthly payments to low income families to mitigate the financial burden of raising children. Other policies help females combine work with raising children. In these programs, the government uses subsidies to help mitigate specific financial burdens faced by employed females, subsidies like maternity- and parental-leave benefits and childcare subsidization. Kalwij (2010) tested the impact of these three public policies on fertility rates in 16 western European countries from 1980 to 2003. He combined micro-level data with macro-level data through pairing these countries' mean per person public policy dollar amounts with micro-level control data for individuals in each country. These variables were regressed against each country's aggregate fertility rate. His results suggest that both maternity- and parental-leave benefits and childcare subsidization increased fertility rates, while family allowances did not have a significant impact.

Theory and Model

These results lead to a two part question of whether the income effect is dominant, and whether public policies can also

significantly increase fertility rates. Mid- to post-1980s research has shown that, as female employment increases, fertility will also increase (Adsera 2003; Goldstein and Örsal 2010; McNown 2003). Previous findings have also suggested that public policies, like maternity- and parental-leave benefits and childcare subsidies that lessen the financial burden of employed females, will have the most robust impact on increasing fertility rates. Other public policies that directly subsidize the cost of having children may have no significant impact at all (Kalwij 2010).

In order to test these hypotheses, macro-level data were taken from 21 OECD countries for the years 1980 to 2013. The countries included are Australia, Austria, Belgium, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Poland, Portugal, Spain, Sweden, Switzerland, and the United Kingdom.

The variables in the models are all in first differences. We employed the first differences in order to control for unit roots found in each of the variables. The variable URB, shown below, is in second difference due to a unit root that was found in the first difference variable. The difference-in-differences model controls for unobserved variable effects, making fixed cross-section effects unnecessary. Random cross-section effects were employed instead. The regression also suffered heteroscedasticity, meaning the variance of the error term was not constant. The white (diagonal) coef covariance method was used to control for heteroscedasticity, since T > N. This method controls for heteroscedasticity across both cross-sections and over time. Equation 1 was used to test the impact of various governments' expenditures have had on fertility rates.

$$\Delta FERT_{it} = \beta_0 + \beta_1 \Delta LEAVE_{it} + \beta_2 \Delta CHILD_{it} + \beta_3 \Delta ALLOW_{it} + \beta_4 \Delta FEMP_{it} + \beta_5 \Delta UNEMP_{it} + \beta_6 \Delta MAR_{it} + \beta_7 \Delta(\Delta URB_{it}) + \beta_7 \Delta INC_{it} + U_{it} \quad (1)$$

The sources of data are presented in the Appendix. The dependent variable is the fertility rate. The primary variables of interest are the public expenditures: maternity- and paternal-leave benefits (LEAVE), childcare subsidization (CHILD), and family allowances (ALLOW). Maternity- and paternal-leave benefits are mean, per person subsidization per infant for employed females, for infants up to age one. Childcare subsidization is mean, per person subsidization per young child for employed females, for young children up to age five. Family allowance is mean, per person subsidization per child for employed females, for children up to age sixteen. Each of these subsidies is adjusted for purchasing power and held in constant U.S. dollars. Previous research has found that both maternity- and parental-leave benefits and childcare subsidization help females combine employment with having the family size they desire, thus, lowering the opportunity cost of having children and increasing the fertility rate (Kalwij 2010). Since these subsidies have been shown to have a positive impact on fertility, the regression results are expected to be positive.

The secondary variable of interest is the number of employed females (FEMP), which has substantially increased over the past century. Research after the 1980s suggests that employed females are more likely to have children because they can better afford to have them (Goldstein & Örsal 2010). Based on research, the resulting coefficient on FEMP is expected to be positive.

The first control variable is the total unemployment rate (UNEM). The higher the unemployment rate, the higher the employment uncertainty. Since past research suggests an income effect is now dominant with both males and females, it is likely that uncertainty in employment will cause females to put off having children until employment levels become stabilized. This suggests that the sign of UNEMP will be negative. Other factors may also be at play, like whether unemployment is seen as temporary or persistent. If unemployment is seen as temporary, females may see this time of unemployment as a less costly time to have children since the opportunity costs involved are low. Whether a secondary provider is in the household will also play a role in whether times of unemployment will increase or decrease fertility rates (Adsera 2003).

The second control variable is the marriage rate (MAR). Traditionally, procreation took place within the context of marriage. This social code is starting to dissolve in some developed countries. In Nordic countries, more than half of all births occur out of wedlock. In the 1960s, this number was closer to one in ten. Even though many no longer adhere to this social code, many still follow it. In Japan and countries in Southern Europe, the birth rates outside of marriage are much lower (d'Addio & d'Ercole 2005). Higher marriage rates will likely lead to higher fertility rates, so the expected sign of MAR will be positive.

The last two control variables are the percentage of the population living in urbanized areas (URB) and each country's per capita income (INC). These control for each country's level of development. If urbanization and per capita income are higher, the country is seen as being more developed. Urban areas with higher per capita incomes tend to have better access to contraceptives and higher female education, which should both decrease fertility rates (Adsera 2003). This suggests that the coefficients for both URB and INC will be negative. Although Adsera (2003) found the opposite sign on URB, this was blamed on sampling from OECD countries that already have relatively high urbanization rates. When sampling from countries around the world at random, the negative coefficient should be strong.

Results

Data for the variables were gathered for 21 OECD countries. Each of these countries was a part of the OECD for the full time frame in order to increase similarity between the cross-sections. After the differencing technique, the periods tested were from 1982–2013. The descriptive statistics, Table 1, and the correlation matrix, Table 2, for these data are posted below.

The descriptive statistics, Table 1, for FERT, UNEMP, MAR, and URB are very low since the variables are differences of percentages, with URB even being in second differences. LEAVE, CHILD, ALLOW, and INC are all differences in per capita dollar amounts, allowing for higher descriptive statistics. FEMP is the number of employed women in thousands, and, so, it also allows for higher descriptive statistics.

Table 1: Descriptive Statistics

Variables	Mean	Standard Deviation	Maximum	Minimum	N
FERT(-1)	-0.007	0.05	0.20	-0.20	759
LEAVE(-1)	2.62	16.31	114.18	-172.16	682
CHILD(-1)	2.01	25.82	294.49	-219.49	604
ALLOW(-1)	4.01	42.25	560.70	-258.49	730
FEMP(-1)	93.83	253.09	1868.00	-1668.00	695
UNEMP(-1)	0.15	1.08	6.61	-3.37	436
MAR(-1)	-0.06	0.57	7.60	-8.10	732
URB(-2)	-.0009	0.11	1.22	-0.90	736
INC(-1)	717.64	3724.51	14180.36	-17594.83	759

Table 2 shows that the correlation between the independent variables in the regression is fairly low. Higher correlations are seen with UNEMP and other independent variables, like INC, MAR, and FEMP. Due to potential multicollinearity, UNEMP was left out of one of the models. Correlation between UNEMP and the fertility rate is also fairly high, making it a valuable explanatory variable.

Table 2: Correlation Matrix

	FERT (-1)	**LEAVE (-1)**	**CHILD (-1)**	**ALLOW (-1)**	**FEMP (-1)**	**UNEMP(-1)**	**MAR(-1)**	**URB(-2)**	**INC(-1)**
FERT(-1)	1.00								
LEAVE(-1)	0.26	1.00							
CHILD(-1)	0.11	0.08	1.00						
ALLOW(-1)	0.09	0.12	-0.11	1.00					
FEMP(-1)	0.13	0.09	0.03	0.04	1.00				
UNEMP(-1)	-0.14	-0.02	0.01	-0.004	-0.37	1.00			
MAR(-1)	0.23	-0.003	0.04	-0.005	0.03	-0.21	1.00		
URB(-2)	0.06	-0.04	0.02	-0.0004	0.15	-0.008	0.06	1.00	
INC(-1)	0.04	-0.04	0.01	0.04	0.10	-0.34	0.15	-0.03	1.00

Three models were created due to a fairly high correlation (-0.37) between the number of employed females and the total unemployment rate. Regressed together, neither are statistically significant. Regressed apart, female employment is statistically significant at a 1% level, and the total unemployment rate is statistically significant at a 10% level. Other than the change in significance between these two explanatory variables and some slight changes in the significance levels of other variables, all three of the models are fairly similar. Based on the adjusted R-squared for each model, the independent variables explain around 10% to 12.4% of the variation in the OECD countries' fertility rates between 1982 and 2013. The results of the regressions are shown in Table 3.

The primary variables of interest, the government subsidies: maternity- and parental-leave benefits, childcare subsidization, and family allowances, are all statistically significant throughout all three of the models. Leave is statistically significant at a

1% level in all of the models. Childcare subsidization is statistically significant at a 5% level in two of the models and statistically significant at a 10% level in the model without total unemployment. Family allowance is statistically significant at a 10% level in two of the models and statistically significant at a 5% level in the model without total unemployment. These results suggest a trend in the significance of government subsidies, as well as in coefficient robustness. Each of these models suggest that maternity- and parental-leave benefits have the strongest effect on fertility rates. Childcare subsidization has the second strongest impact, leaving family allowance as the weakest of the subsidies.

Table 3: OLS results (Dependent variable is Fertility Rate)

	Model 1		**Model 2**		**Model 3**	
Variable	**Coefficients**	**T-statistics**	**Coefficients**	**T-statistics**	**Coefficients**	**T-statistics**
Constant	0.002	0.67	-0.011	-3.58	0.003	1.27
LEAVE(-1)	0.001***	3.35	0.0008***	5.422	0.001***	3.41
CHILD(-1)	0.0001**	2.08	0.0001*	1.67	0.0002**	2.10
ALLOW(-1)	5.35E-05*	1.64	7.6E-05**	2.17	5.51E-05*	1.68
FEMP(-1)	1.56E-05	1.02	4.23E-05***	3.33	-	-
UNEMP(-1)	-0.003	-1.31	-	-	-0.004*	-1.91
MAR(-1)	0.034***	3.21	0.015*	1.85	0.03***	3.19
URB(-2)	0.015	0.84	0.022	1.54	0.02	1.02
INC(-1)	-1.51E-07	-0.27	4.64E-07	0.87	-1.62E-07	-0.28
R-squared	0.146		0.112		0.143	
Adjusted R-sq.	0.124		0.100		0.124	
F-statistic	6.619		9.254		7.42	

The t-statistics are calculated using robust standard errors. The robust standard errors are measured by the white (diagonal) coef covariance method. * p<0.10; ** p<0.05; ***p<0.01

Discussion

The regression results herein are similar to recent research findings (McNown 2003; Goldstein and Örsal 2010; Kalwij 2010). After the 1980s, research results have suggested a stronger income effect as compared to the substitution effect. Evidence (Model 2, after controlling for unemployment) herein support dominance of the income effect: as female employment increases, fertility rates increase.

Both of the government subsidies, maternity- and parental-leave benefits and childcare subsidization, have a statistically significant, positive impact on fertility rates, as Kalwij (2010) also found. This is an intuitive finding in the sense that government subsidies have an impact on fertility rates because they alleviate some of the financial burden of having children. These policies provide the household with additional resources, thereby, reducing the financial burden and increasing the likelihood of child-bearing in the household. This finding is in contrast to those of Kalwij (2010).

An important econometric specification issue is addressed herein that is not typically found in previous literature on fertility rates. The time-series nature of the data introduces the possibility of the presence of unit roots in the series. Failing to account for non-stationary results or effects in the data lead to inconsistent and biased coefficients in previous research wherein this issue was not addressed. Variable used in the models were tested for unit roots and were first-differenced as appropriate. Further, population living in urbanized areas required second differencing to purge the remaining non-stationarity in the first difference.

Not only does this paper shed light on potential oversights in past data, it also provides results that suggest significant implications for governments in developed countries. It further suggests that public policies have the ability to increase fertility rates. This can be very valuable information for the governments of countries confronting low fertility rates. A fertility rate of greater than 2.1 is required in many developed countries to maintain the long-term level of population and avoid population loss through time. Population loss can lead to significant negative economic impacts associated with fertility rates below the replacement ratio. Aside from immigration, other feasible public policy options such as maternity- and parental-leave benefits, childcare subsidization, and family allowances may contribute to increased fertility rates. The results herein suggest the marginal effects on fertility rates associated with maternity- and parental-leave benefits are statistically significant and larger

in magnitude as compared to childcare credits. As such, it may be beneficial for countries to consider adopting greater use of these public policy options.

References

Adsera, Alicia. 2003. "Changing fertility rates in developed countries. The impact of labor market institutions." *The Journal of Population Economics* 17: 1-7.

Bloom, David, E., Klaus Prettner, and Holger Strulik. 2012. "Declining fertility and economic well-being: Do education and health ride to the rescue?" *Labour Economics* 22, no. 84: 70-79.

Butz, William and Michael Ward. 1979. "The Emergence of countercyclical U.S. fertility." *The American Economic Review* 69, no. 3: 318-28.

d''Addio, Anna C. and Marco M. d''Ercole. 2005. "Trends and Determinants of Fertility Rates in OECD Countries: The Role of Policies." *Organisation for Economic Co-operation and Development (OECD)*, no. 27.

De Jager, A. 2013. Social welfare policy as an instrument for fertility regulation. Netspar Theses.

Ermisch, John F. 1988. "Econometric analysis of birth rate dynamics in Britain." *Journal of Human Resources* 23, no. 4: 563–76.

Evan, Tomáš and Vozárová, Pavla 2018. "Influence of women's workforce participation and pensions on total fertility rate: a theoretical and econometric study."*Eurasian Economic Review; Heidelberg,* Vol. 8, Iss. 1, 51-72.

Fitzgerald, Bridgid, 2011. "World Population Hitting 7 Billion," Earth Policy Institute, October 12, available at http://www.earth-policy.org/indicators/C40/population_2011

Goldstein Joshua and Deniz Dilan Örsal. 2010. "The Increasing Importance of Economic Conditions on Fertility." *Max Planck Institute for Demographic Research:* 1-16.

Jinno, Masatoshi and Yasuoka, Masaya. 2016. "Are the Social Security Benefits of Pensions or Childcare Policies Best Finance by Consumption Tax?" *Business and Economic Horizons; Prague*: 12, iss 3: 94-112

Kalwij, Adriaan. 2010. "The Impact of Family Policy Expenditure on Fertility in Western Europe." *Demography* 47*,* no. 2: 503-19.

McNown, Robert. 2003. "A Cointegration Model of Age-Specific Fertility and Female Labor Supply in the United States." *Southern Economic Journal* 70, no. 2: 344-58.

OECD Data, available at https://data.oecd.org/socialexp/social-spending.htm

OECD Data, available at https://data.oecd.org/emp/employment-rate.htm

OECD Data, available at https://data.oecd.org/unemp/unemployment-rate.htm

OECD. Stat, available at https://stats.oecd.org/index.aspx?queryid=68249

OECD. Stat, available at https://stats.oecd.org/index.aspx?DataSetCode=EO

The World Bank Data, available at https://data.worldbank.org/indicator/sp.dyn.tfrt.in?name_desc=false

The World Bank Data, available at https://data.worldbank.org/indicator/SP.URB.TOTL.in.zs

Data Appendix

Fertility rates- World Bank
Select: Data; Search: Fertility rate, total (births per woman)

Social policies- OECD Social Expenditure-Aggregated Date
Source: Public; Branch: Family; Type of Expenditure: Cash benefit; Type of Programme: Family- Maternity- and parental-leave and Family- Other cash benefits (Childcare subsidization); Measure: Per head, at constant prices (2010) and constant PPPs (2010), in U.S. dollars

Employed population, Aged 15 and over, females- OECD short-term labour market statistics: employed population
Units: persons, thousands

Unemployment rate Total, % of labour force- OECD

Crude marriage rate, Marriages per 1000 people- OECD Family Database
Select: 1. The structure of families, marital and partnership status, SF3.1 Marriage and Divorce Rates: Marriages per thousand

Urban population (% of total population) - World Bank Economic Indicators

GDP per capita (current US$) - OECD economic outlook

Gross Domestic Product: Implicit Price Deflator (GDPDEF) -FRED

Does Doing Well Lead to Doing Better? Evidence from ESG Impact Investing and the S&P 500

Marc J. Sardy, Rollins College
R. Paul Herman & Onindo Khan, HIP Investor Inc.

Abstract

There are contrasting views in finance regarding firms that use methods that modify their businesses using a stakeholder oriented (which can be based on ESG) rather than a shareholder-oriented approaches (which frequently ignore ESG factors). Are firms that have embraced a more responsible approach to doing business, within the firm and with outside stakeholders more profitable than those using a more traditional shareholder maximization approach? It is becoming more clear that much measurable benefit accrues to shareholders under a stakeholder-centric model, HIP ratings total shareholder returns are a good benchmark. We examine the effects through identifying that firms improve as demonstrated through Shareholder return, risk/volatility & Sharpe Ratios. We have seen that firms improve their organization through measures that lead to higher impact scores as measured through both HIP methodology and Bloomberg ESG measures. We examine the returns of firms that compose the S&P index and look at performance weighted by these factors relative to the S&P index.

Introduction

Over the last decade the stock market has hit historic lows and risen to historic highs. In 2008 when the stock market collapsed many investors began to realign their view of what defined a firm as truly successful. It became clear that financial markets could not be counted on to regulate corporate behavior and become a driving force for more productive and socially responsible behavior. Neoclassical finance in the view of Irving Fisher shows the firm solely as a black box that generates income to shareholders. Modern Finance put the focus on shareholder value and firm performance. Miller and Modigliani in the 1960's developed the idea that firm success drives from capital structure only. Most of modern finance has put the emphasis on shareholder value, stakeholders have been considered a distant secondary priority. Capital theory considers a firm's success or failure a direct outcome of their capital structure. This does not consider the various stakeholders as having a role in the ultimate performance of the firm. However, more recently investors have discovered that firms making impact on a range of measures have yielded higher returns (Snider, 2015) & (Sardy & Lewin, 2017). Investors are usually the most impactful when it comes to setting the agenda of a firm with regard to social, environmental and economic focus. Investor attraction to firms that do a better job on these measures leads to improved share price while bad news about a firm's unwillingness to address these issues may lead to investors voting with their feet (Grace, Wood and Thornley, 2012). Thus, over the last several years impact investing has developed as an approach to better address these concerns on the part of investors. Impact investing takes traditional modes of assessing a firm in the neoclassical way and extends it to looking at social impact and how this might improve the ultimate performance of the firm (Sikken, 2011) & (Sardy & Lewin, 2016). However, ESG disclosure still has a long way to go for many firms' financial directors as many are not yet convinced it is in their best interest to disclose these metrics. Perhaps some are concerned about the negative disclosure and their impacts on the firm; For instance, higher GHG emissions are a risk if politicians seriously pursue the 1.5% target, or another example are work place injuries which could lead to direct financial impact via law suits but also to a less motivated work force, which in turn may reduce productivity. However, perhaps it is not clear to many of those FD's: If you do good you have higher chance of doing well and the if you don't do good you are exposing shareholders to higher risk, a carrot and stick question.

Ultimately investors are concerned whether impact investments can make returns comparable with those not considered impact firms. One recent study shows that there may be in excess of the $4 billion impact investing market in the US. (consider including the ESG numbers as well, which are not 20% of AUM, and the fastest growing segment of assets managed). These investors have found that the returns generated by impact firms have produced consistent market returns. In many cases investors are under the impression that they may have to sacrifice any abnormal return in exchange for socially responsible corporate practices, we believe this thinking to be incorrect. Non-concessionary impact investors are especially likely to have investment impact in conditions of imperfect information—for example, in social or environmental niche markets where impact investment fund managers or other intermediaries have special expertise or intelligence on the ground. Indeed, it is becoming more a demand of investors for SRI to be part of the selection metrics used in establishing their portfolio choices. One of the unfortunate characteristics of imperfect impact investing markets is their inability to attract the large majority of socially neutral investors who demand market returns, this is compounded by the dearth of selections available to investors through 401K and other limited investment option vehicles. Anecdotally this is confirmed by the employee responses we see on the 401k

evaluation platform (www.CleanPortfolios.com)

Where such returns seem plausible, a respected institution can signal to other investors that a particular investment or an entire sector that others may have thought dubious is actually worthy of consideration. According to Strom (2011) "the main reason for investing in EcoTrust Forest (FUND) in this way is to demonstrate that sustainable forest practices can generate a profit so that mainstream investors will become more interested in it." Motivated investors may be particularly interested in identifying these opportunities and thus may be able to have impact even at non-concessionary rates. This is the most likely explanation for asserting the double-bottom-line success of firms like Elevar Equity. Elevar Equity generates "outstanding investment returns by delivering essential services to disconnected communities underserved by global networks." The forthcoming analysis of impact investing funds by Clark et al. (2016) should further illuminate the returns space.

Corporate social responsibility (CSR) has been used as a way for investors to identify firms that are more responsive to various stake holders. This has led to more focus from investors on the firms who give back to their communities. There has been much discussion on the real value of CSR, with some research showing more engagement with stakeholders lead to better corporate performance and better corporate citizenship. Other research has suggested that CSR has been used as a way to greenwash a firm and show a façade of community engagement without really committing to a more holistic approach to responsible business practices. Thus, there has been disagreement with the value of CSR as a practice within the firm. In most cases this comes without a lot of hard numbers in standardized (SASB or GIIN etc.) metrics Thus there are good way to distinguish bad CSR reporting from good CSR reporting. However, in recent years more investor interest has been directed to firms that put the emphasis on stakeholders. Stakeholders can be as limited as employees or as expansive as environmental stewardship. The scope may be determined by the firm and its reach and priorities.

While there is periodic evidence in markets of the positive performance of impact firms a major issue is the method used to measure impact and assess related returns. Neoclassical finance would hold that any effort to focus the firm on stakeholder value at the expense of shareholder value would be a breach of fiduciary responsibility of corporate governance. This view is based on historical assumptions about firm performance that are losing traction with investors in more recent decades. So, it is important to develop consistent measures to help assess whether or not the firm is actually performing better financially because of the effort of the firm to maintain social impact.

A more recent development has been various organizations that have made the effort to develop a system of analyzing firms to determine a system of measurements that better assesses whether the firm is truly committed to responsible, progressive practices. Investors have responded extremely positively to these developments. This is led to organizations like Bloomberg adding environmental, social, and governmental methods of assessing the level of commitment of the firm to the Bloomberg platform. Today many firms traded in the United States can be analyzed on Bloomberg platform with respect to their ESG metrics and evaluation. Another organization that has grown out of this need for better assessment of firm performance with respect to responsible practices grew from a book (Herman), "The HIP Investor: Make Bigger Profits by Building a Better World" via human impact + profit, or HIP. In his book Herman diverged from the traditional mindset of only examining firms which were attempting to engage responsible practices to expanding to more commercial and well-regarded firms and examining their commitment to practices that consider stakeholders their involvement in the ultimate performance of the firm.

Systems for 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 (notably companies going out of business or not producing good value products and services have lower impact than the ones which run a good business, simply in virtue of them being bigger/existent), 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 goal of these efforts is 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.

A third method, the Bloomberg product platform offers ESG as one of the measures of performance of the firm. There are three pillars to ESG the environmental as measured through various environmental analyses and disclosures of the firm, the social as measured through various behaviors towards internal and external stakeholders and governance which examines composition of corporate governance of the firm. Each of these measures are developed from a combination of firm-based financial performance measures and disclosures on file and within the firm's annual report. Ultimately, these measures are compiled into a composite score. A composite score can be looked at for most firms trading in the United States. Each of the pillars of Bloomberg's ESG rankings is based on a scale of 0 to 100 with zero being the lowest possible score suggesting there is no compliance with this category and 100 suggesting there is complete compliance with this category. Ultimately a composite score is put together for an overall rating of each firm on a scale of 0 to 100. Once again, zero is the lowest possible score and 100 is a perfect score. In practice there are no firms with perfect scores. It could be said that all firms have room for improvement under Bloomberg's ESG rankings.

Yet a fourth method is one developed by R. Paul Herman as a measure of Human Impact + Profit investing practices (HIP). This methodology includes 5 pillars of impact; health, wealth, earth, equality and trust, covering more than 20 metrics of human, social and environmental performance (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. Human-impact investing practices are similar to Bloomberg in that it also looks at several measures of firm performance through several different sources to develop five pillars of firm performance to get a better composite score to assess the corporate performance of the firm. Each of these five pillars impact; health, wealth, earth, equality and trust (Herman, 2010) have component scores with which a firm is assessed. These component scores are pulled from a combination of direct measures from financial reports and from other softer measures garnered from reports and articles of the firm. HIP firms have regularly outperformed the S&P index and have also provided stakeholder value beyond the basics of the risk-return framework. These measures are then aggregated into an overall HIP rank. The HIP ranks are rated on a scale of 0 to 100 where firms with a ranking of zero do virtually nothing with regard to HIP pillars and firms with a ranking of 100 are completely compliant with a sustainable resilient world.. In practice, no firms have a perfect score under HIP practices either; though firms can rank 100% on a relative scale by geography, market cap, and industry/sector.

In this study, our goal was to analyze ESG and HIP rankings on the Standard & Poor's 500 to determine whether the firms with high ESG or high HIP rankings within the Standard & Poor's outperform those with low HIP or ESG rankings. If firms with high rankings outperform those with low rankings this would suggest that high ESG and high HIP firms ultimately had better performance than those with low HIP and low ESG rankings. We could ultimately conclude that HIP and ESG practices lead to better firm performance. This might stand in stark contrast to basic modern finance theory which says that just the capital structure of the firm will lead to the best measures of firm performance. Anecdotal evidence of several funds which track stakeholder value such as the Parnassus endeavor fund, a fund which looks at the top 100 companies to work for and invests in those firms which are tradable and rebalances annually based on firms entering and exiting the list. The Parnassus fund has traditionally had superior performance well in excess of the S&P 500. Several other funds which track social measures as their investment strategy have also shown a strong ability to outperform traditional market measures. HIP's greater places to work have regularly outperformed the S&P 500.

Data and Methodology

In our study, we gathered results for all components of the Standard & Poor's 500, we also gathered historical rankings of each of these component firms for both the HIP and ESG measures. Our initial pass breaks our Standard & Poor's 500 data into deciles based on these rankings. For HIP rankings the highest 10% of firms by ranking would be put in the 1st decile, the next 10% of firms would be put in the 2nd decile and so on until the lowest group would be the bottom 10% of firms in the Standard & Poor's as determined by their HIP score and ESG score.

Table 1: HIP Measures

HIP Rank By Decile

Decile	1	2	3	4	5	6	7	8	9	10
2008	0.46	0.38	0.33	0.3	0.28	0.26	0.25	0.23	0.21	0.16
2009	0.48	0.4	0.35	0.32	0.29	0.28	0.26	0.23	0.21	0.15
2010	0.49	0.41	0.37	0.34	0.31	0.29	0.27	0.26	0.23	0.16
2011	0.52	0.44	0.39	0.36	0.33	0.31	0.29	0.27	0.24	0.18
2012	0.55	0.47	0.43	0.41	0.36	0.34	0.31	0.29	0.27	0.19
2013	0.56	0.49	0.44	0.41	0.37	0.35	0.32	0.3	0.26	0.18
2014	0.57	0.49	0.46	0.42	0.39	0.35	0.33	0.3	0.26	0.18
2015	0.6	0.53	0.48	0.44	0.41	0.38	0.35	0.32	0.28	0.21
2016	0.62	0.56	0.51	0.47	0.44	0.41	0.37	0.34	0.3	0.22
2017	0.62	0.55	0.51	0.47	0.43	0.39	0.36	0.33	0.28	0.21
2018	0.65	0.6	0.56	0.52	0.48	0.45	0.41	0.38	0.33	0.27

% Annual Return By Decile

Decile	1	2	3	4	5	6	7	8	9	10	S&P
2008	-29.89	-21.34	-27.23	-41.17	-25.46	-27.32	-31.27	-29.04	-25.03	-30.74	-34.48
2009	27.73	28.7	25.95	41.46	46.05	34.1	37.16	33.59	39.5	35.83	35.02
2010	11.25	17.02	21.67	22.43	25.38	24.39	30.73	15.53	24.88	26.68	17.11
2011	5.42	-1.72	0.31	5.84	7.36	1.6	8.21	4.85	16.35	6.88	-2.22
2012	19.57	21.98	15.75	19.94	20.51	19.87	15.86	21.38	17.87	21.15	8.67
2013	40.01	30.45	29.3	32.6	39.56	42.19	41.53	26.87	21.8	31.25	23.38
2014	15.75	15.65	14.36	19.6	14.27	10.87	18.27	17.9	19.51	12.57	15.5
2015	-0.99	0.01	5.48	3.13	-1.61	2.1	-0.58	7.39	8.58	-10.75	2.45
2016	14.71	19.43	19.52	16.46	14.55	15.18	16.2	12.3	8.92	22.39	15.39
2017	22.66	23.93	17.02	26.33	21.68	20.76	10.77	15.12	15.43	13.48	17.32
2018	-3.09	-3.99	-11.62	-6.37	-6.32	-5.65	-5.37	-11.84	-12.95	-9.07	-11.22

Sharpe Ratio by Decile

Decile	1	2	3	4	5	6	7	8	9	10
2008	-1.2	-0.79	-1.25	-1.79	-0.67	-1.07	-1.67	-1.37	-1.14	-1.07
2009	0.79	0.94	1.01	0.85	0.55	0.66	1.23	1.05	0.58	0.84
2010	0.48	0.8	1	0.69	0.91	1.22	1.31	0.82	1.02	0.88
2011	0.21	-0.09	0.01	0.28	0.34	0.08	0.33	0.25	0.64	0.24
2012	0.78	0.9	0.75	0.93	1.17	0.96	0.78	0.75	0.48	0.57
2013	1.73	1.39	1.06	1.33	1.63	1.01	1.6	0.97	0.75	1.62
2014	1.21	0.95	0.82	0.78	0.93	0.66	0.92	0.71	0.82	0.48
2015	-0.11	-0.03	0.21	0.12	-0.11	0.05	-0.05	0.3	0.38	-0.5
2016	0.8	0.8	0.46	0.96	0.74	0.46	0.56	0.56	0.41	0.89
2017	0.88	0.93	0.87	1.01	0.82	0.82	0.4	0.59	0.6	0.43
2018	-0.24	-0.38	-0.68	-0.52	-0.45	-0.4	-0.42	-0.69	-1.18	-0.44

% Annual Abnormal Return By Decile

Decile	1	2	3	4	5	6	7	8	9	10
2008	4.59	13.14 *	7.25 +	-6.69 +	9.02	"7.16 +	3.21	5.44 +	9.45 *	3.74
2009	-7.29	-6.32	-9.07 *	6.44	11.03	-0.92	2.14	-1.43	4.48	0.81
2010	-5.86+	-0.09	4.56	5.32	8.27 +	7.28 *	13.62 **	-1.58	7.77 *	9.57 *
2011	7.64 +	0.5	2.53	8.06 *	9.58 *	3.82	10.43 *	7.07 *	18.57 **	9.1 *
2012	10.9 *	13.31 **	7.08 *	11.27 **	11.84 **	11.2 **	7.19 *	12.71 *	9.2 +	12.48 *
2013	16.63 *	7.07 **	5.92	9.22 *	16.18 **	18.81 **	18.15 **	3.49	-1.58	7.87 *
2014	0.25	0.15	-1.14	4.1	-1.23	-4.63 *	2.77	2.4	4.01	-2.93
2015	-3.44 +	-2.44	3.03	0.68	-4.06	-0.35	-3.03	4.94 +	6.13 *	-13.2 **
2016	-0.68	4.04	4.13	1.07	-0.84	-0.21	0.81	-3.09	-6.47 *	7 *
2017	5.34 +	6.61 *	-0.3	9.01	4.36	3.44	-6.55 *	-2.2	-1.89	-3.84
2018	8.13 *	7.23 **	-0.4	4.85 *	4.9 *	5.57 *	5.85 *	-0.62	-1.73	2.15

+=.05<Alpha>.10, *.01<Alpha<.05, **Alpha<.01

Table 2: Other Measures

					Alpha					
Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2008	-1.41	9.05	6.7	-13.83	-2.6	2.18	-2.57	-7.53	-7.07	-14.63
2009	-1.48	-0.02	-3.23	4.1	16.14	3.19	2	1.96	4.79	9
2010	13.2	19.2	23.6	24.1	28	26.1	33	17.1	27.2	28.7
2011	7.43	0.3	2.37	7.67	9.29	4.15	10.1	7.27	18	9.6
2012	14.3	10.6	20.4	17.4	21.5	15.5	14.3	11.3	16.5	18.9
2013	16.09	12.73	8.06	15.09	17.13	18.27	9.42	9.33	5.18	3.44
2014	1.08	0.2	-0.2	6.34	-2.8	-5.5	-2.5	7.06	3.32	-4.7
2015	-3.5	-2.2	3.31	0.76	-4.2	-0.5	-3.4	5.02	6.45	-13
2016	-1.3	4.78	7.15	0.96	-0.9	-1.7	-2.3	-2.5	-4.9	5.61
2017	6.51	6.9	2.43	8.22	4.51	3.5	-11.6	0.55	-3.81	-0.38
2018	8.3	8.81	3.75	6.12	11.19	4.34	8.21	-3.69	1.32	-1.29
					Beta					
Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2008	0.83	0.88	0.98	0.8	0.67	0.86	0.83	0.63	0.53	0.47
2009	0.83	0.82	0.83	1.07	0.85	0.88	1	0.9	0.99	0.76
2010	0.89	0.98	0.86	0.79	1.14	0.77	0.98	0.69	1.02	0.86
2011	0.91	0.91	0.94	0.83	0.88	1.14	0.84	1.09	0.75	1.22
2012	0.86	0.9	0.95	0.75	0.88	1.08	0.91	1.07	0.6	1.32
2013	1.02	0.76	0.91	0.75	0.96	1.02	1.3	0.73	0.66	1.13
2014	0.92	1	0.91	0.81	1.1	1.02	1.27	0.66	0.99	1.12
2015	1.01	0.88	0.84	0.95	1.08	1.1	1.21	0.96	0.83	1.06
2016	1.04	0.95	0.79	1.01	1.01	1.1	1.2	0.96	0.91	1.08
2017	0.89	0.98	0.69	1.05	0.9	0.94	1.32	0.82	0.98	0.7
2018	0.94	1.11	1.22	1.09	1.4	0.88	1.19	0.74	1.22	0.65

There are several other practices in our initial pass which will be changed going forward but may have some odd impacts on the data. Each portfolio was formed based on the year-end ESG or HIP ranking. The deciles are then based on year-end performance from the preceding year of the rankings. The returns data which we use to analyze the firm is based on April to April stock performance. For example, the annual performance of the stock goes from April, and HIP ranking for December 2012 would then be used to choose the portfolio whose returns would be assessed from April 2013 through April 2013. In addition, our first pass did not always include all of the firms which had dropped out of the Standard & Poor's 500. This data can sometimes be challenging to work with. However subsequently this data has been gathered and in our next pass, we will correct this issue using January to December data. Thus, portfolio choice and performance will be very closely related. In addition to looking at the overall HIP and ESG scores we also examined each of the HIP pillars and their performance. This approach needs revision and will be addressed in a later version of this paper.

In successive years the performance of the hip portfolios is consistent with our expectations. However, in some off years' portfolios are somewhat inconsistent. In addition, oddly the middle portfolios seem to outperform some of the extreme high and low HIP firms. This may result from several issues. Firms on either end of the spectrum may be facing different challenges. One clear indication based on 2018 Standard & Poor's weights suggest that the firms in the 1st decile are considerably larger than those in the 10th decile. In addition, even at points where the lower HIP firm's portfolio might do better Sharpe ratios suggest that it is more of a risk-based return rather than a lower risk return coming from the larger firms in the 1st decile.

Bloomberg ESG ratings seem to be consistent with HIP ratings in that they show returns associated with the smallest decile (decile 10) are largely based on small firm returns which are less likely to have implemented impact measures within the firm. Another shortcoming of Bloomberg ESG data is our window was one year shorter than our HIP window. Here we can see that the high ESG firms outperform the low ESG firms. Oddly, Bloomberg seems to have discontinued reporting ESG data for the most current year for firms in the Standard & Poor's 500.

Ultimately, the HIP index seems to do a better job on predicting performance based on predetermined success pillars. However, both Bloomberg and HIP data suffer from survivorship bias with regard to the S&P 500. This is highlighted by the fact that almost half of S&P firms have turned over since 2000.

Table 3: Bloomberg ESG Measures

ESG Rank by Decile											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
2014	58.61	48.02	42.92	37.47	32.1	26.62	21.13	17.51	15.48	12.81	
2015	59.5	49.3	43.48	37.58	32.8	27.54	21.89	18.38	16.07	12.7	
2016	59.56	50.26	43.93	38.25	33.86	28.6	22.9	19.22	16.16	12.93	
2017	58.27	49.03	43.59	38.66	33.03	27.18	22.49	18.91	15.61	12.3	
ESG Return by Decile VS 2018 Portfolio Weight											
Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**S&P**
2014	14.46	17.53	26.5	16.6	20.6	24.77	20.08	18.79	22.82	11.92	12.73
Avg 2018 weight	0.43	0.32	0.22	0.21	0.24	0.14	0.14	0.18	0.08	0.13	
2015	0.86	4.36	3.33	2.38	-4.14	1.55	3.13	-0.9	2.05	-0.52	-1.41
Avg 2018 weight	0.41	0.3	0.24	0.2	0.23	0.17	0.11	0.12	0.18	0.11	
2016	26.3	18.9	16.9	19.7	14.4	13.4	23.9	15.9	14.3	16.7	13.72
Avg 2018 weight	0.41	0.33	0.19	0.23	0.13	0.26	0.19	0.11	0.1	0.13	
2017	12.11	13.26	12.36	17.28	12.65	14.85	12.8	15.91	7.2	10.83	10.97
Avg 2018 weight	0.48	0.25	0.28	0.15	0.18	0.22	0.14	0.12	0.1	0.14	
% Annual Abnormal Return By Decile											
Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
2014	1.73	4.8 *	13.77 **	3.87	7.87 *	12.05 **	7.36 *	6.06 *	10.1 **	-0.81	
2015	2.27	5.77 *	4.74 *	3.8 +	-2.73	2.96 +	4.54 +	0.51	3.47	0.89	
2016	12.56 **	5.21	3.13	5.96 *	0.68	-0.29	10.18 *	2.14	0.6	3.01	
2017	1.14	2.29	1.39	6.31	1.68	3.88	1.83	4.94 +	-3.77	-0.14	
ESG Sharpe Ratio by Decile											
Decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
2014	0.81	0.81	1.1	0.69	0.85	1.21	0.74	0.75	0.82	0.63	
2015	0.02	0.21	0.16	0.1	-0.22	0.08	0.11	-0.07	0.08	-0.05	
2016	1.21	0.53	0.76	0.96	0.54	0.63	0.57	0.78	0.55	0.66	
2017	0.5	0.47	0.58	0.29	0.44	0.42	0.37	0.56	0.25	0.43	

+=.05<alpha<.1, *= .01<Alpha<.05, *=.01<alpha

References

Alsford, Jessica, Victoria Chapelow, Richard Felton, Eva Zlotnicka, 'Embedding sustainability into valuation', Institute for Sustainable Investment, Morgan Stanley (2015).https://www.yumpu.com/en/document/view/36603848/embedding/103.

Barton, Dominic, 'Capitalism for the long term', *Harvard Business Review*, March 2011.

Bonini, Sheila, Schwartz, Stephen, 'Profits with Purpose: How Organising for Sustainability can benefit the bottom line', McKinsey on Sustainability & Resource Productivity, July 2014.

Brest, Paul, Born Kelly, 'When Can Impact Investing Create Real Impact?', *Stanford Social Innovation Review*, Fall 2013.

Clark, Cathy, Jed Emerson and Ben Thornley, 'Success in Impact Investing Through Policy Symbiosis' Huffington Post article 2014. http://www.huffingtonpost.com/ben-thornley/success-in-impact-investi_b_4849740.html

Clark, Cathy, Jed Emerson and Ben Thornley, The Impact Investor: Lessons in Leadership and Strategy for Collaborative Capitalism, Jossey-Bass, October 13, 2014.

Clark, Gordon et al., 'From the Stockholder to the Stakeholder: How Sustainability can drive financial performance,' Smith School of Enterprise & University of Oxford & Arabesque Asset Management (2015)

Clark, Gordon, Andreas Feiner, and Michael Viehs. "How Sustainability Can Drive Financial Outperformance." 2014.

Cohen, Ronald, William. A. Sahlman, 'Social Impact Investing Will Be the New Venture Capital', *Harvard Business Review*, January 17, 2013.

El Ghoul, Sadok, Omrane Guedhami, Chuck C.Y. Kwok, and Mishra, Dev R., 'Does Corporate Social Responsibility Affect the Cost of Capital?'. *Journal of Banking & Finance*, Vol. 35, Issue 9, pp. 2388-2406, September 2011. Available at SSRN: http://ssrn.com/abstract=1546755

Fitzgerald, Patrick, 'Sustainable capitalism: there is no trade-off', working paper, 2016. http://www.slideshare.net/DanFitzgerald7/sustainable-capitalism-there-is-no-tradeoff-january-13-2016-dan-fitzgerald.

Herman, R. Paul, (2010) The HIP Investor: Make Bigger Profits by Building a Better World, John Wiley & Sons (April 2010).

Sardy, M, Lewin, R. (2016),Towards A Global Framework for Impact Investing,*Academy of Economics & Finance Journal* 7.

Sardy, M, Lewin, R. (2017),Does impact investing lower the cost of capital? *Academy of Economics & Finance Journal* 8.

Long-run Equilibrium Shifts and Short-run Dynamics of National Stock Markets during Financial Crisis

John Tarwater, Cedarville University

Abstract

I employ a vector error correction model to examine the interdependence between price stock indexes of NAFTA countries that have been segregated into tiers based on market capitalization. In each set of NAFTA countries (US – MEX; US – CAN; MEX – CAN), the returns of the tiered indexes reflect a long-run relationship within the same tier. Using a rolling vector error correction approach, I find a shift in the long run equilibrium during the most recent global financial crisis. The cointegrating parameter that ties the tiers together is greater in the absolute during the crisis period compared to the pre- and post-crisis periods. Despite showing that the stock indexes of the three NAFTA countries exhibit a cointegrating relationship, tests do not confirm that the relationship is the result of the NAFTA accord.

Introduction

According to the World Bank, an explosion of regional trade agreements (RTA) has occurred in the past twenty years, increasing from 50 in 1990 to more than 300 in 2018.[i] Consequently, the proliferation of RTAs has led to greater linkages between various goods and equity markets. In addition to regional trade agreements, the mere globalization of securities has introduced increased linkages between various goods and equity markets (Hamao, Masulis, & Ng, 1990; Meric, Leal, Ratner, & Meric, 2001; Okada, 2013; Phylaktis & Ravazzolo, 2002).

Consequently, academicians and practitioners alike have investigated the possible effects of these tightening relationships. While some of these studies have examined the long-run relationship between the stock markets of various major countries, such as the US, Japan, England, and Germany (Masih & Masih, 2002), many have focused on the potential cointegration of markets within a particular region, such as the countries of the North American Free Trade Agreement (NAFTA) (Aggarwal & Kyaw, 2005; De Hoyos & Iacovone, 2013; Bradley T. Ewing, Payne, & Sowell, 1999; Lahrech & Sylwester, 2013) or Gulf Cooperation Council (GCC) (Alotaibi & Mishra, 2017). As a result, we have a better understanding of how the strength of market linkages might change based upon how developed the country or market is (Al Nasser & Hajilee, 2016), as well as its relationship to a local economic crisis (Climent & Meneu, 2003; In, Kim, Yoon, & Viney, 2001).

Identifying these relationships plays an important role for various stakeholders. Investors are interested in market diversification in order to reduce risk. If national equity markets are cointegrated, then the benefits of diversification are limited (Byers & Peel, 1993; De la Torre, Gozzi, & Schmukler, 2007). Managers, on the other hand, might find the financial integration increases market efficiency through the flow of information and adjustments (Darrat & Zhong, 2005). Additionally, managers may be interested in how integrated markets affect expected returns and the cost of capital. As foreign investors participate in the local market, the source of systematic risk shifts to the world stock market, thus affecting expected returns and stock market prices (Chari & Henry, 2004). Moreover, a reduction in risk should also result in a reduction in the cost of capital (Bekaert & Harvey, 2000). Finally, these increased relationships may also be of significant interest to policy makers and economic planners who discover that integration can foster development through more efficient allocation of capital (Umutlu, Akdeniz, & Altay-Salih, 2010) and a lower the probability of asymmetric shocks (Yu, Fung, & Tam, 2010). Policy makers may utilize regulatory policy to restrict foreign equity investments in order to limit exposure to adverse volatility effects. Such actions, however, may limit the ability of firms to raise capital for projects, thus stifling economic growth.

In this essay, I study the long term relationship of the equity markets between the three countries that participate in NAFTA during the period from January 2005 through December 2016. In contrast to previous studies, I examine this relationship via stock returns that have been stratified by tiers based upon market capitalization. This approach follows a recent study of the long-run relationship in the housing market in which the authors evaluated long-run equilibrium shifts in price tiers (Damianov & Escobari, 2016). Each index tracks the appreciation and depreciation rates of the stock prices in segments over time.

While I employ a variety of econometric methods, my major results flow from two techniques in particular. I estimate a vector error correction model, which allows us to evaluate both long-run and short-run dynamics. Through this test, I am able to discern the strength of a linkages between the tiers of the various countries. More importantly, however, I can use these results to gain insight on how the short-term dynamics influence the long-term relationship. Afterwards, I utilize a rolling regression approach to allow the cointegrating parameters to vary over time. Using this procedure, I create a panel of estimated cointegrating parameters, which allows me test the stability of the relationship before, during, and after the global financial crisis of 2008.

Thus, I believe that this study fills three major gaps in the research regarding cointegration among NAFTA countries. First, previous research on cointegration among NAFTA countries reached conflicting results. Early studies showed no cointegration (Atteberry & Swanson, 1997; Bradley T. Ewing et al., 1999; Bradley T. Ewing, Payne, & Sowell, 2001), while later studies were split. In later studies, the researchers divided the samples between pre- and post-NAFTA periods. Following this method, some found for cointegration among the countries in the post-NAFTA period (Aggarwal & Kyaw, 2005; Darrat & Zhong, 2001; Gilmore & McManus, 2004), and others found evidence of no cointegration in the same period (Ciner, 2006; Phengpis & Swanson, 2006). Consequently, I believe that there is a gap to be explored here.

Secondly, I believe my decision to investigate cointegration among NAFTA countries by means of stratifying the markets based on market capitalization yields important information previously left unexplored. Various studies support the use of segregating firms based on market capitalization or size. In his groundbreaking article on the size effect, Banz (1981) observed that smaller firms tend to have higher returns than larger firms, which later studies seem to support (Fama & French, 1995; Reinganum, 1983). Financial theory suggests that firm size (i.e., market capitalization) is a proxy for risk, and that smaller firms tend to have greater risk than larger firms. Indeed, in their study on the earnings during the recession of 1981-82, Fama and French (1992) show that the small firms are exposed to cyclical risk factors in a fundamentally different way than for large firms, perhaps because of their access to credit markets.[ii] In this regard, smaller firms have a higher return due to their liquidity risk (Liu, 2006) or informational uncertainty (Zhang, 2006). Consequently, market forces exert downward pressure on the prices of stocks for small firms which lead to higher returns for investors.

Just as the Zillow Home Value Index (ZHVI) and the Case-Shiller index (CSI) capture changes in home prices over time by separating the prices into tiers (high tier, middle tier, and low tier), I attempt to acquire similar pricing information by dividing stock prices into tiers based upon market capitalization. As Damianov and Escobari (2016) used the price tiers in home prices to evaluate long run shifts in home prices, I aim to imitate this technique to gain information about the cointegrating relationship between the economies of the NAFTA countries.

Lastly, previous studies on cointegration have not investigated parameter stability of long run links. I believe that these three gaps allow me to explore and address meaningful gaps in the literature.

As a result of my analysis, I ascertain five meaningful insights. First, not dissimilar from previous studies of the returns on indexes for national stock markets, I find that the log of the price indexes for each country in each tier is non-stationary in levels but is stationary in first differences (i.e., they are integrated of order one).

Secondly, I show that for various pairs of NAFTA countries (e.g., US – MEX; US – CAN; and MEX – CAN) the stock returns of one country appreciate faster than the returns of its counterpart prior to the global financial crisis of 2008. After the crisis, however, the same countries depreciate faster than their counterparts. These comparisons are made within a particular market tier. For instance, I examine $\text{LowTier}^{\text{Mexico}}$ with $\text{LowTier}^{\text{USA}}$. Using cointegration tests, I quantify the extent of this phenomenon for each pair of countries. I find that each pair in each tier is bound by a long-run relationship. Consequently, the stock returns for each country in the pair are driven by the same factors that drive the long-run relationship to be cointegrated. Stated differently, if markets are segmented, then each market's assets are priced according to factors particular to that domestic market. If the market is integrated, then the market's assets are priced and experience returns according to international factors (Taylor & Tonks, 1989). Cointegration in this context implies that each national stock price series contains valuable information on the common stochastic trends which bind each pair of countries' returns together. Therefore, one might be able to use this information to predict another country's stock returns, which would be evidence against the market efficiency hypothesis (Granger, 1986). Nevertheless, others argue that predictability does not imply inefficiency (Masih & Masih, 2002). To be sure, violation of the market efficiency hypothesis assumes that one could use the information to earn risk-adjusted excess returns.

Thirdly, I show that in various instances the short-run dynamics between country pairs exhibits strong correlation with the long-run equilibrium parameters. Consequently, the research suggests that lagged short-term dynamics influences the long-run equilibrium of stock returns within a common priced tier.

Fourthly, I show that for each of the three-country pairs in both high and low tiers, the cointegrating parameter (β) is statistically significant. In four out of the six pairings, furthermore, the cointegrating parameter is negative. For example, the index of returns for stocks in the United States in the low tier are consistently above the index values for Mexico in the low tier. Moreover, the value of the cointegrating parameter is larger than negative one in the absolute. The negative cointegrating parameter suggests that for every dollar of price appreciation in the United States in the lower tier, the stock prices of Mexico in the low tier will appreciate at a greater amount in the long run. To my knowledge, I am the first to study the short run and long run dynamics of the time series of segmented prices during the most recent recession period.

Finally, I utilize a rolling regression approach to allow the cointegrating parameters to vary over time. Using this procedure, I create a panel of estimated cointegrating parameters, which lets us test the stability of the relationship before, during, and after the global financial crisis of 2008. Consequently, I estimate the cointegrating factor (β) for a window of 72 months for six groups (three country pairs in two tiers). This technique allows me to obtain a panel of betas for every set of countries in each tier in which the price tiers are cointegrated. Using this panel, I find a statistically significant shift in the long run

equilibrium. The tests show that for each country pair, the cointegrating parameter increases in absolute value reaching its peak at the height of the recession and afterwards decreasing. Therefore, the lower priced stocks in each tier are relatively more volatile compared to the higher priced stocks of the comparable country in the same tier.

Data

In this study, I examine stock price indexes for the three NAFTA countries (USA, Mexico, and Canada). For the US stock indexes, I use the S&P 500 Largecap index to capture high market capitalized companies. To qualify for this index, a company must have a minimum market value of $5.3 billion. For the low tier companies, I use the S&P Smallcap 600 index, which is comprised of 600 companies. The mean market capitalization for the S&P Smallcap 600 is about $1.2 billion, with market values ranging between $450 million to $2.1 billion. I access these indexes through Compustat via the Wharton Research Data Services (WRDS). The sample includes US high tier data from January 1988 to December 2016, and the low tier data covers the period January 1989 through December 2016.

For Mexico stock prices, I use data obtained directly from S&P Dow Jones Indices LLC. The main benchmark on the Mexican stock market, which is the Bolsa de Valores de Mexico (BMV), is called IPC, which stands for *índice de precios y cotizaciones*. The S&P/BMV IPC seeks to measure the performance of the largest and most liquid stocks in the Mexican market. This index captures the high tier stock prices for Mexico. For the low tier stocks, I use the S&P/BMV/IPC Smallcap index. The sample for both tiers covered the period from December 2004 through December 2016.

For Canada, I utilize information on the Toronto Stock Exchange (TSX) through TMX Group Inc. For the high tier stock, I use S&P/TSX 60 Largecap, and for the low tier, I use S&P/TSX Smallcap. While the high tier index covers the period from January 1988 through December 2016, the low tier data covers only the period from May 1999 through December 2016.

The indexes for USA and for Mexico are based on the US dollar. The Canadian indexes are based on the Canadian dollar. I use the monthly average exchange rates from the Federal Reserve Board to convert the Canadian dollars to the US dollar equivalent. Therefore, all three indexes are based on a common currency.[iii]

In order for unit root and vector error correction tests to have the same sample, I limit the data to the period from January 2005 to December 2016. I convert all data to this same base year, and convert all prices to reruns by taking the natural log of the daily price relatives. Table 1 reports the summary statistics for the three NAFTA countries in both the low tier and high tier indexes.

Table 1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
				LOWTIER					HIGHTIER		
Country	Obs.	Mean	S.D.	RSD	Min	Max	Mean	S.D.	RSD	Min	Max
United States	144	4.938	0.317	6.411	4.161	5.564	4.797	0.253	5.281	4.131	5.245
Mexico	144	5.272	0.460	8.725	4.218	5.979	5.557	0.388	6.982	4.554	6.020
Canada	144	4.662	0.234	5.011	3.959	5.048	5.064	0.177	3.486	4.554	5.359

Notes: The sample contains monthly observations of the log of prices from 2005m1 through 2016m12 from price indexes of three countries: USA, Mexico, and Canada. There are 144 times-series observations of LOWTIER and HIGHTIER for each of the three countries. The base period for this essay is 2005m1. All series are in a common currency: the US dollar.

I have 144 monthly returns for three countries in two tiers. Thus, I have 864 tiered index observations for my initial statistics. The standard deviations are reported in columns 3 and 8, and the relative standard deviations are reported in columns 4 and 9. Each of these statistics shows that the LOWTIER returns are more volatile than the HIGHTIER returns. This finding lends further credence to the size effect suggested by previous research (Banz, 1981; Basu, 1997; Liu, 2006; Reinganum, 1983; Zhang, 2006), as well as my decision to segment the returns based on market capitalization. Moreover, the summary statistics also suggest that stock returns in Mexico indexes are more volatile than the returns of the United States and Canada in the same tiers.

Defining the Existence of Long Run Equilibrium

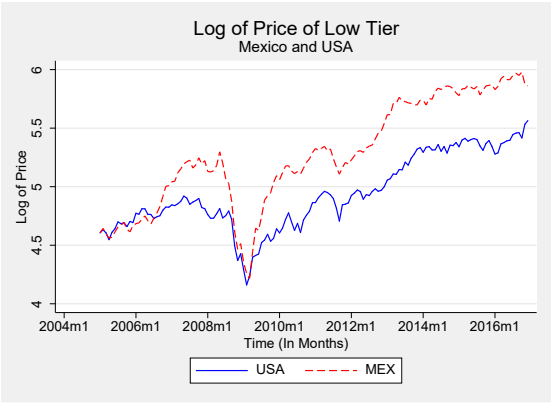
In the study, I have two series of data (Low and High) for each of the three NAFTA countries, thus giving us six series in all. I call the series of index returns for the small capitalized firms LOWTIER and the series of index returns for large capitalized firms HIGHTIER. In this study, I do not focus the investigation on the relationship between the high and low tiers of a single country, such as $\text{LOWTIER}^{\text{Mexico}}$ and $\text{HIGHTIER}^{\text{Mexico}}$. Rather, I examine the relationship between NAFTA countries within a common tier, such as $\text{MEXICO}^{\text{LowTier}}$ and $\text{USA}^{\text{LowTier}}$.

Therefore, for any NAFTA country (NC) during month (t), I have NC_t^{Tier}. If I combine the price tiers of any two different NAFTA countries in the following manner,

$$NC1_t^{Tier} + \beta NC2_t^{Tier} = d_t^{Tier}, \quad (1)$$

where NC1 and NC2 represent the log of prices for any two NAFTA countries, I say that a long run equilibrium exists if the difference d_t^{Tier} is stationary for a given constant β. If a long run equilibrium exists, then $d_t^{Tier} = I(0)$, and $NC1_t^{Tier} = -\beta NC2_t^{Tier}$. Figure 1, for example, shows the graphs of the log of prices for Mexico and USA in the low tier (i.e., MEXICO$^{\text{LowTier}}$ and USA$^{\text{LowTier}}$). While the graph is not determinative, it at least suggests that a long run relationship exists between the two series beginning shortly after 2009.

Figure 1: Mexico - USA (Low Tier)



Investors across firms and countries possess a myriad of preferences with respect to risk and future price changes. Moreover, firms across the region react differently to various investment opportunities. Because of the varied preferences of investors and reactions by firms, shocks to the stock returns occur, which temporarily cause the returns to deviate from the long run equilibrium. These short run deviations from the long run equilibrium are present when $d_t^{Tier} \neq I(0)$. Because the differenced series is stationary, any deviations are necessarily temporary and are corrected in subsequent time periods. To be sure, if NAFTA has strengthened the market linkages within the region, then the markets should digest the information being communicated in the prices more quickly and consequently return to its long run equilibrium values. By stating that a long run equilibrium exists, I mean that the log of the prices of the countries in each tier are cointegrated with a vector $[1, \beta]$. The same conclusion is true for high tier returns of each country. If the long run relationship exists, then the tiered returns between countries are integrated of order one, $I(1)$, and d_t^{Tier} is integrated of order zero, $I(0)$.

I test cointegration between countries within a tier by restricting β in Eq. 1 to be equal to negative one. A necessary condition for cointegration is that the variables in the system must be integrated of the same order. I employ two types of unit root tests in order to determine whether or not the variables are stationary. First, I use the generalized least squares Augmented Dickey-Fuller (ADF) type test (Dickey & Fuller, 1979), which has the null hypothesis that a unit root exists. The ADF test attempts to account for any temporarily dependent or heterogeneously distributed errors by including lagged sequences of first differences of the variable in its set of regressors. For this reason, I employ the Akaike Information Criterion (AIC) to determine the appropriate number of lags.

Despite the popularity and ease of using the ADF test, subsequent studies suggest that the test lacks power and too frequently accepts the null hypothesis (DeJong, Nankervis, Savin, & Whiteman, 1992). Consequently, I also use a test proposed by Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) that takes as its null hypothesis that the series is stationary (Kwiatkowski, Phillips, Schmidt, & Shin, 1992). The primary statistic is based on a partial sum of the residuals from the regression, producing an LM statistic:

$$LM = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\sigma_\epsilon^2} \text{ where } S_t = \sum_{i=1}^{t} e_i, t = 1,2,\ldots,T. \quad (2)$$

In this context, e_i is the residual term from the regression of y_t on an intercept; $\widehat{\sigma_\epsilon^2}$ is a consistent long run variance estimate of y_t; and T represents the sample size. The KPSS statistic has a nonstandard distribution and critical values. If the KPSS statistic is large, then the null of stationarity is rejected.

I test each series in the log of levels ($NC_t^{LowTier}$ and $NC_t^{HighTier}$) and in differences in the log of prices ($\Delta NC_t^{LowTier}$ and $\Delta NC_t^{HighTier}$). I compute these statistics for each series in the sample. The results for the unit root tests show all three countries are non-stationary in levels in the low and high tiers. The ADF test suggests that the returns in first differences for Mexico are not stationary in the high tier. However, the ADF is known to have less power and too frequently accepts the null. Hence, the KPSS test I believe accurately reflects that it is stationary at the 1% level. Combining these results, I find that all series are I(1) in levels and I(0) in first differences.

Estimating Long Run Equilibrium

Because variables are not stationary (integrated of order 1), I test for the existence of a long run relationship between them.

Tests for Cointegration

Early studies on cointegration show that a long run relationship may be ascertained through a two-step procedure (Engle & Granger, 1987). However, subsequent studies suggest a far superior procedure using a maximum likelihood estimator (ML) (Johansen, 1988; Johansen & Juselius, 1990). This technique not only avoids the possibility of carrying over errors from the first step of estimation to later steps, it also allows greater flexibility for testing hypotheses. For this reason, I employ methods proposed by Johansen (1988) and Johansen and Juselius (1990). The ML estimator of the parameters of a cointegrating relationship is

$$\Delta y_t = \alpha \xi' y_{t-1} + \sum_{t-1}^{p-1} \Gamma_i \Delta y_{t-1} + \epsilon_t \tag{3}$$

where y is a $(K\ x\ 1)$ vector of I(1) variables, α and ξ are $(K\ x\ r)$ parameter matrices with rank $r < K, \Gamma_1, \ldots, \Gamma_{p-1}$, and ϵ_t is a $(K\ x\ 1)$ vector of normally distributed errors that is serially uncorrelated but has contemporaneous covariance matrix Ω. From this ML estimator, Johansen derives two likelihood-ratio (LR) tests for inference on r—the number of cointegrating equations. The two LR tests are known as the trace statistic and the maximum eigenvalue statistic.

The null hypothesis of the trace statistic is that there are no more than r cointegrating relations. Thus, the remaining eigenvalues are zero. The alternative hypothesis to the trace statistic is the maximum eigenvalue statistics, which has as its null that there are exactly r cointegrating equations. In Table 2, I present results for the existence of cointegrating relationships between different countries based upon pricing tiers. The trace and maximum eigenvalue statistics provide no conclusive answer regarding cointegration. For example, I report the maximum eigenvalue statistic in columns 1 and 2 of Panel B. For each country pair (USA-MEX, USA-CAN, and CAN-MEX), I can reject the null hypothesis that there are no cointegrating equations in both tiers (column 1) Panel B). However, the critical values for the maximum eigenvalue test do not correspondingly support that there is at least one cointegrating factor (column 2). In columns 3 and 4, I similarly do not find support for a cointegrating relationship between any country pair in either tier via the trace statistic.

In contrast to the Johansen tests, a third method for determining the number of cointegrating equations involves minimizing an information criterion, such as the Schwarz Bayesian information criterion (SBIC) or the Hannan Quinn information criterion (HQIC) (Aznar & Salvador, 2002; Gonzalo & Pitarakis, 1998). The final column in Panel A shows that there are at least two cointegrating equations among the three NAFTA countries in both tiers, suggesting that the price tiers in the three countries are integrated. Similarly, in Panel B, the Hannan-Quinn statistic reported in column 8 shows that all three country pairs are cointegrated.

Because the Johansen methods and the Hannan-Quinn criterion supply inconsistent statistics, I explore further whether or not the lack of cointegration might stem from regime or trend shifts, shifts for which the Johansen methods do not account. Research shows that the power of the Johansen test falls drastically when a structural break exists in the data, which affects unit root and cointegration tests (Perron, 1989; Perron & Vogelsang, 1992; Život & Andrews, 1992). These breaks can contribute to a loss of power in standard residual-based testing (Gregory & Hansen, 1996b). Consequently, Gregory and Hansen (1996b) proposed cointegration tests that account for a break in the cointegrating relationship (Gregory & Hansen, 1996a). Table 3 presents three test statistics for the three country pairs in both price tiers: the ADF-type test, the Z_a type test, and the Z_t type test. All three tests are significant at the 1% level, providing additional evidence of a cointegrating relationship. Columns 2, 5, and 7 reflect the estimated date of the shifts.[iv]

Tests for Vector Error Correction

While I recognize that Eq. 1 theoretically represents the long run equilibrium, I also anticipate that true data generating process is far more complex than such a simple equation. As a result, I estimate a more robust and flexible vector-error correction model not dissimilar from the model proposed by Damianov and Escobari (2016). The following equations shows the model:

$$\Delta NC1_t^{Tier} = a_1 + \gamma_1\left(NC1_{t-1}^{Tier} + \beta NC2_{t-1}^{Tier}\right) + \sum_{j=1}^{k} a_{11}(j)\Delta NC1_{t-j}^{Tier} + \sum_{j=1}^{k} a_{12}(j)\Delta NC2_{t-j}^{Tier} + \epsilon_{NC_1 t} \quad (4)$$

$$\Delta NC2_t^{Tier} = a_2 + \gamma_2\left(NC1_{t-1}^{Tier} + \beta NC2_{t-1}^{Tier}\right) + \sum_{j=1}^{k} a_{21}(j)\Delta NC1_{t-j}^{Tier} + \sum_{j=1}^{k} a_{22}(j)\Delta NC2_{t-j}^{Tier} + \epsilon_{NC_2 t} \quad (5)$$

Table 2: Cointegration Tests (Johansen and Hannan-Quinn) (Log of Prices)

Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Maximum Eigenvalue			Trace Statistic			Minimum Hannan-Quinn			
	r=0	r=1	r=2	r=0	r<=1	r<=2	r=0	r=1	r=2	Eq.
USA – Mex - Can										
Low Tier	30.58	21.22	15.96	67.75	37.18	15.96	-8.718	-8.824	-8.908	2
High Tier	30.00	24.08	15.26	69.34	39.34	15.26	-10.77	-10.87	-10.98	2

Panel B	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Maximum Eigenvalue		Trace Statistic		Minimum Hannan-Quinn			
	r=0	r=1	r=0	r<=1	r=0	r=1	r=2	#Eq.
USA – MEX								
Low Tier	29.68	15.61	45.29	15.61	-5.875	-6.021	-6.111	1
High Tier	25.14	17.87	43.01	17.87	-6.883	-6.996	-7.102	1
USA – CAN								
Low Tier	26.68	19.75	46.43	19.75	-5.591	-5.714	-5.835	1
High Tier	24.30	13.67	37.97	13.67	-7.200	-7.307	-7.383	1
CAN – MEX								
Low Tier	21.94	17.75	36.69	17.75	-5.466	-5.556	-5.661	1
High Tier	31.66	20.33	52.00	20.33	-6.578	-6.738	-6.862	1

Notes: Critical values for the maximum eigenvalue for r = 0 are 14.07 (5%) and 18.63 (1%); and r = 1 are 3.76 (5%) and 6.65 (1%). Critical values for the trace statistic for r = 0 are 15.41 (5%) and 20.04 (1%). For $r \leq 1$ are 3.76 (5%) and 6.65 (1%).

Table 3: Cointegration with Regime Shifts (Log of Prices)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ADF-type test			Z_t -type test		Z_a -type test	
USA – Mex- Can	t-stat	Break	Lag	Z_t-stat	Break	Z_a- stat	Break
Low Tier	-12.88	2011m10	0	-12.93	2011m10	-154.7	2011m10
High Tier	-9.649	2007m6	1	-13.56	2010m10	-161.2	2010m10
USA - MEX							
Low Tier	-11.16	2008m.11	0	-11.77	2011m7	-141.3	2011m7
High Tier	-9.562	2010m9	1	-13.68	2010m10	-162.5	2010m10
USA - CAN							
Low Tier	-13.98	2011m7	0	-14.03	2011m7	-166.2	2011m7
High Tier	-10.38	2014m7	0	-10.42	2014m7	-123.9	2014m7
CAN - MEX							
Low Tier	-12.22	2011m10	0	-12.26	2011m10	-147.1	2011m10
High Tier	-10.62	2007m2	1	-13.90	2007m1	-164.6	2007m1

Notes: Optimal lag for the ADF-type test chosen by the Akaike criterion. All specifications model a change in regime and trend. Critical values for ADF-type test are -6.02 (1%), -5.50 (5%), and -5.24 (10%). Critical values for the Z_a-type test and Z_t-test are different for the different countries. All statistics are significant at the 1% level except for USA – CAN (High Tier), at 5%.

In this model, NC1 and NC2 represent any two NAFTA countries under comparison. The term in parentheses signifies the long run equilibrium, which is equivalent to the difference formula in Eq. 1. The left-hand side of equations 4 and 5 captures the short run dynamics. When variations from the long run equilibrium occur in the short run, they must be "corrected." The speed of this correction is captured by the parameters γ_1 and γ_2. If the deviations from equilibrium are positive (i.e., $d_t^{Tier} > 0$), then β is negative and the price index in NAFTA country 1 is relatively larger than the price index of NAFTA country 2 within the same tier. As a result, it is not unreasonable to anticipate that NC1 will decrease in the short run (i.e., $\Delta NC1_t^{Tier} < 0$) and NC2 will increase (i.e., $\Delta NC2_t^{Tier} > 0$). In this situation, therefore, I expect $\gamma_1$to be negative and $\gamma_2$to be positive.

An important component in the model is its capacity to provide information about short run dynamics. Consider the parameters on the lags of the right-hand part of the equation: a_{11} and a_{22}. If an increase (or decrease) in this lag term is associated with a corresponding increase (or decrease) in the current period, then the parameters would capture momentum effects. And although each parameter technically derives from its own particular equation, its presence within the system of equations within the model allows us to test for the strength of momentum within a particular country.

Table 4 presents the maximum likelihood estimation of Equations 4 and 5 for each set of NAFTA countries within both tiers. I used the AIC to determine the optimal lag length. According to the model, a vector of $[1, \beta]$ depicts the long run equilibrium. Column 9 presents the calculated β for each NAFTA pair. The results show considerable variation in the values of the cointegrating statistic. For each NAFTA country in both tiers, the $\beta's$ are all statistically significant.

Table 4: Vector Error Correction Model (Log of Prices)

Panel A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mexico - USA	Change in Mexico				Change in USA				
	a_1	γ_1	a_{11}	a_{12}	a_2	γ_2	a_{22}	a_{21}	β
Low Tier	-0.030	0.129**	-0.536***	0.325***	0.009	0.429***	-0.139*	0.130	-3.072***
	0.367	0.0538	0.087	0.109	0.340	0.0498	0.0804	0.101	0.279
High Tier	-0.016	-0.123***	-0.517***	0.533***	0.015	-0.125***	-0.0068	0.0120	6.728***
	0.317	0.0196	0.082	0.120	0.256	0.0158	0.0660	0.0969	0.892
Panel B									
Canada–USA		Change in Canada				Change in USA			
Low Tier	0.023	0.101	-0.550***	0.312**	-.005	0.474***	-0.247***	0.164	-2.738***
	0.469	0.081	0.0943	0.145	0.350	0.0603	0.070	0.109	0.234
High Tier	-0.018	0.129	-0.508***	0.300**	0.0034	0.662***	-0.398***	0.234*	-1.888***
	0.320	0.115	0.113	0.147	0.264	0.095	0.093	0.122	0.142
Panel C									
Mexico-Canada		Change in Mexico				Change in Canada			
Low Tier	-0.031	-0.488***	-0.206**	0.188***	0.026	-0.577***	0.489***	-0.347***	0.517***
	0.329	0.0742	0.0842	0.063	0.417	0.094	0.107	0.080	0.159
High Tier	-0.003	0.586***	-0.144	-0.122	-0.008	-0.205*	-0.240**	-0.157	-1.463***
	0.328	0.112	0.106	0.112	0.321	0.110	0.103	0.109	0.132

Note: The figures in the second row are the standard errors. The symbol *** indicates significance at 1%; the symbol ** signifies significance at 5%; the symbol * signifies significance at 10%.

Four of the six estimates of the cointegrating parameter are signed negative and two are positive. Not only is the parameter for Mexico – USA in the high tier signed positive, it is extremely large. This variation from the other estimates may be related to the regime shifts identified in the cointegration tests in Table 3. If the VECM were able to account for the shift, the value of the estimate might not appear as an outlier. Indeed, Table 5 reflects the value of the cointegrating factor when the parameter is allowed to vary over time. In each instance, the value of the parameter is negative and greater than one in the absolute.

The statistically significant β in Table 4 suggests a long run relationship between two countries within the pricing tiers. For Mexico – USA, for example, $Mexico_t^{Low} = 3.072 \cdot USA_t^{Low}$. Thus, a one point increase in the prices of stocks in the low tier of the USA are associated with 3.072 increase in the stock prices in Mexico within the low tier, suggesting returns for USA stocks in the low tier are appreciating slower than low tier stocks in Mexico. Perhaps this finding reflects the volatility of the Mexico stocks. With greater volatility comes increased risk, which potentially leads to greater returns.

Table 4 also shows that there is considerable evidence for momentum in both tiers. Only Mexico in the high tier in the Mexico – Canada pairing and USA in the high tier in the USA – Mexico pairing failed to have a statistically significant result. This finding suggests that the immediate past returns for any country in any tier have a significant impact on the current short-term returns.

Stability of Long Run Equilibrium

In Eqs. 4 and 5, the estimate of β is assumed to be time invariant. In light of the global financial crisis around 2007- 08 however, this assumption may be too strong. While any two NAFTA markets may be cointegrated within the tiers, one might suspect that the markets might respond to the shocks differently during the time of the crisis—the long run equilibrium might shift. For this reason, I investigate the stability of the long run equilibrium ($NC1_t^{Tier} + \beta NC2_t^{Tier}$) by using a rolling regression approach. To execute a rolling vector error correction model, I allow β to change over time.

First, I establish a window of months over which I estimate Eqs. 4 and 5. For my study, I chose 72 months. This size was big enough to get a solid estimate, but also small enough that I could roll it forward another 72 times. For my first estimate, I evaluate observations $[1, w]$.Subsequent estimates move from $[2, w+1]$ until I reach the final month: $[2, w+1]$, $[3, w+2]$, … $[N-w+1, N]$. From this procedure, I ascertain a series of (β_t) that captures the dynamics of the cointegrating equation. I repeat this process for each of the six NAFTA pairs, allowing us to obtain a panel of estimates: β_{it}.[v]

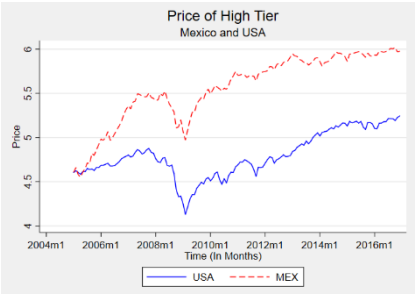
Using this panel, I analyze how the financial crisis affected the long run equilibrium between countries within a particular price tier. The model I estimate is:

$$\beta_{it} = \delta \cdot \mathbb{I}_{|\tau_i - t| \leq \theta} + \mu_i + \eta_{it} \tag{6}$$

where $\mathbb{I}$ is an indicator function that captures the period of the financial crisis from its formation to its end. The statistic τ_i represents the time at which the crisis for a particular grouping of NAFTA countries (i) reached its height, and θ denotes a positive integer that captures the time distance away from this period. For the purposes of this essay, I calculated the time at which the crisis "bubble" burst as the date at which the lower series in a grouping reached its maximum. The statistic of greatest interest is δ, which captures any shift in the long run equilibrium (β_{it}). Two final variables capture time invariant specific effects to the NAFTA groupings (μ_i) and any remaining stochastic error (η_{it}). Because I do not know the actual length of the crisis for each grouping, I use different window sizes (i.e., θ) equal to 3 months, 6 months, and 12 months.[vi]

Figure 2 presents the graph of the returns for high tier indexes of Mexico and the United States. From the graph, one can easily detect an ordered relationship, with the returns for stocks in Mexico being on the top and the returns for the stocks in the USA being on bottom. The question under discussion is whether or not the long run relationship between the two return series, which was previously identified in the VECM, is stable. To be sure, I explore this question via Eq. 6. Table 5 presents the results of this analysis.

Figure 0: Mexico - USA (High Tier)



I explore the possible parameter instability from various time and econometric perspectives. Table 5 shows the three "window" options for θ (3 months, 6 months, and 12 months). Moreover, the table also presents the estimates for β prior to the height of the crisis (β_{Pre}), at the height of the crisis (β_{Dur}), and the difference between the two $(\beta_{Dur} - \beta_{Pre})$. Lastly, I analyzed the parameter using both a pooled and fixed effects regression. Column 4, for example, with a $\theta = 6$, indicates that a one percentage increase (decrease) in the lower series of any two NAFTA countries was associated with a corresponding 1.609 % increase (decrease) in the upper series of the group during the period prior to the financial crisis. This elasticity increases to 1.90 during the 12 months around the height of the global financial crisis. From Figure 2 one can see an example with USA and Mexico, with returns for the USA on bottom and Mexico having the returns shown on top. This finding suggests that the rate of appreciation (or depreciation) of Mexico was greater than the rate of appreciation (or depreciation) of the United States during the crisis. Stated more generally, the returns of NAFTA countries on the upper portion of the graph appreciate (or depreciate) at a higher rate than those whose returns are below and cointegrated with them.

The results show that the estimates for β are significant at the 1% level for both periods before and during the crisis. More importantly, the final row of Table 2.6 explores the difference between the two periods. Columns 1 and 2 show that for $\theta = 3$, the difference between β_{Pre} and β_{Dur} is statistically significant at the 5% level. Columns 4, 5, and 6, on the other hand, show that for $\theta = 6$ and $\theta = 12$ the difference between β_{Pre} and β_{Dur} is statistically significant at the 1% level The results presented in Table 5 indicate support for the conclusion that β shifted during the crisis. Indeed, the first column indicates that the appreciation (or depreciation) rate of the log of price is .380 smaller in the years leading to the crisis than it was at the height of the crisis.

As a robustness check, I rerun all of the same statistical tests on the non-transformed variables; I test prices. As before, all of the unit root tests, cointegration tests, and the vector error correction models show significant results. The summary statistics

for the prices show that that prices in the lower tier are more volatile than the prices in the upper tier, and the prices of Mexico are considerably more volatile than those in the USA or Canada. Similarly, the unit root tests show that the price indexes are non-stationary in levels and stationary in first differences, and the cointegrating tests support the conclusion that the price indexes are cointegrated. The VECM in prices show long run equilibrium estimates of β are significant for four of the six country pairings. However, the evidence for momentum is largely non-existent.

Table 5: Long Run Equilibrium (Log of Prices)

	(1)	(2)	(3)	(4)	(5)	(6)
	$\theta = 3$		$\theta = 6$		$\theta = 12$	
VARIABLES	Pooled	FE	Pooled	FE	Pooled	FE
β_{pre}	-1.627***	-1.627***	-1.609***	-1.609***	-1.536***	-1.550***
	(0.0552)	(0.0359)	(0.0579)	(0.0377)	(0.0636)	(0.0412)
$H_0: \beta_{pre} = 0$	[0]	[0]	[0]	[0]	[0]	[0]
$H_0: \beta_{pre} = -1$	[0]	[0]	[0]	[0]	[0]	[0]
β_{Dur}	-2.006***	-2.006***	-1.900***	-1.900***	-1.915***	-1.887***
	(0.180)	(0.117)	(0.127)	(0.0828)	(0.0919)	(0.0595)
$H_0: \beta_{Dur} = 0$	[0]	[0]	[0]	[0]	[0]	[0]
$H_0: \beta_{Dur} = -1$	[0]	[0]	[0]	[0]	[0]	[0]
$\beta_{Dur} - \beta_{Pre}$	-0.380**	-0.380***	-0.291**	-0.291***	-0.379***	-0.337***
	(0.188)	(0.123)	(0.140)	(0.0910)	(0.112)	(0.0724)
$H_0: \beta_{Dur} - \beta_{Pre} = 0$	[0.045]	[0.002]	[0.038]	[0.002]	[0.001]	[0]
Obs	420	420	420	420	420	420

Note: Figures in parentheses are standard errors. Figures in brackets are p-values. The symbol *** indicates significance at 1%; the symbol ** signifies significance at 5%; the symbol * signifies significance at 10%.

Table 6: Long Run Equilibrium (Prices)

	(1)	(2)	(3)	(4)	(5)	(6)
	$\theta = 3$		$\theta = 6$		$\theta = 12$	
VARIABLES	OLS3	FE3	OLS6	FE6	OLS12	FE12
β_{pre}	-0.628***	-0.618***	-0.643***	-0.615***	-0.792***	-0.720***
	(0.174)	(0.164)	(0.182)	(0.171)	(0.199)	(0.187)
$H_0: \beta_{pre} = 0$	[0]	[0]	[0]	[0]	[0]	[0]
$H_0: \beta_{pre} = -1$	[0.034]	[0.020]	[0.050]	[0.025]	[0.297]	[0.137]
β_{Dur}	-1.739***	-1.869***	-1.113**	-1.282***	-0.502	-0.685**
	0.617	0.579	(0.444)	(0.418)	(0.317)	(0.301)
$H_0: \beta_{Dur} = 0$	[0.005]	[0.001]	[0.013]	[0.002]	[0.113]	[0.023]
$H_0: \beta_{Dur} = -1$	[0.231]	[0.134]	[0.799]	[0.500]	[0.117]	[0.296]
$\beta_{Dur} - \beta_{Pre}$	-1.111*	-1.251**	-0.470	-0.667	0.290	0.0352
	(0.641)	(0.602)	(0.480)	(0.452)	(0.374)	(0.356)
$H_0: \beta_{Dur} - \beta_{Pre} = 0$	[0.084]	[0.038]	[0.328]	[0.141]	[0.438]	[0.921]
Obs	432	432	432	432	432	432

Note: Figures in parentheses are standard errors. Figures in brackets are p-values. The symbol *** indicates significance at 1%; the symbol ** signifies significance at 5%; the symbol * signifies significance at 10%.

I regenerate a new panel of estimates for β_{it} via a rolling regression technique. Lastly, I analyze Eq. 6 based on this new panel, which reflects prices rather than returns. The results are reported in Table 6. Column 4, for example, with $\theta = 6$, indicates that a one point increase (decrease) in the lower series of any two NAFTA countries was associated with a corresponding $.615$ point increase (decrease) in the upper series of the group during the period prior to the financial crisis. At the height of the crisis, this number increases to 1.282. For the three month window size, the difference between β_{Pre} and β_{Dur} is statistically significant at either the 5% or 10% level. In light of these robustness tests, therefore, I can say with considerable confidence that the long run parameter shifted during the time of the crisis.

Finally, I attempt to determine whether or not the long run relationship among the tiered returns of the three countries was the result of the regional trade agreement—is there a NAFTA effect? In order to investigate this question, I perform the same

tests on a non-NAFTA country and compare them with the returns from the USA, Mexico, and Canada. I chose Japan as my non-NAFTA country. Table 7 shows the results of the cointegration tests between Japan and each NAFA country. The results show that the returns of Japan in each tier are cointegrated with the returns from each NAFTA country. In other words, the results suggest that while the returns from the three NAFTA countries may show evidence of tightening market linkages, they may not be necessarily due to a NAFTA effect.

Conclusion

In this essay, I investigate the interrelationships between the stock markets of three North American Countries (Mexico, North American, and Canada) which have been stratified into tiers based on market capitalization. In particular, I examine the long run relationship during the period of 2005 through 2016, thus covering the global financial crisis of 2007. An understanding of this relationship between markets provides important information about the economic significance of NAFTA and for investors and managers who depend on stock market behavior.

In this essay, I investigate not only if the stock markets of the three North American countries are cointegrated, I examine whether or not NAFTA was the cause of that long run relationship. More broadly, I asked if RTAs potentially lead to economically and financially linked economies. The results of this analysis suggest that the stock markets of the United States, Mexico, and Canada not only are linked, but that they are also linked within tiers based upon market capitalization. The study also suggests the existence of structural breaks in the data, which might account for previous researchers not finding a cointegrating relationship among the stock markets of the three North American countries. In addition, the study suggests the long run equilibrium links between the stock markets of the NAFTA countries may not necessarily be the result of NAFTA.

My results come with some caveats. The cointegration may result from increased correlations resulting from the financial crisis which dominates the period under investigation. And even though the results are robust, the current study controls may not capture the full impact of factors not considered. Still, my robustness tests find that the markets are cointegrated not only between countries, but also within countries. And, the robustness tests show that the cointegration not only stems from series based on returns, but also to series based on stock prices.

These findings are significant for investors, who are seeking to diversify risk. If the stock markets are linked, and if they are linked within tiers, then one cannot diversify the risk of a stock by investing in a similarly-sized stock from a neighboring country's stock market. Company managers, likewise, may use this information when making investment decisions tied to the cost of capital. If the potential investment lies within a cointegrated market, then the systematic risk shifts to the world market, which affects expected returns.

It is difficult to predict the long term effect of the agreement meant to replace NAFTA – United States – Mexico – Canada Agreement (USMCA). Although some individual markets, such as the automotive market between US and Mexico and the dairy market between US and Canada, experienced some significant changes, most of the changes in the new agreement appear more cosmetic. In fact, the changes appear to reflect that fact that the United States is more important to the economies of Mexico and Canada than each of the other countries are to the United States. Because the changes to NAFTA are minimal, it is doubtful that they will significantly affect the long term stock market relationship of those countries. Indeed, as the world becomes increasingly linked, I anticipate that those markets will reflect the same tightening reality.

Table 7: Cointegration Tests (Japan) (Log of Prices)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Maximum Eigenvalue		Trace Statistic		Minimum Hannan Quinn			
	r=0	r=1	r=0	r<=1	r=0	r=1	r=2	#Eq.
USA – JAP								
Low Tier	30.94	21.90	52.84	21.90	-5.763	-5.918	-6.053	1
High Tier	24.55	17.97	42.53	17.97	-6.371	-6.480	-6.587	1
JAP – CAN								
Low Tier	23.88	15.21	39.09	15.21	-5.635	-5.738	-5.825	1
High Tier	20.00	14.61	34.60	14.61	-5.989	-6.065	-6.147	1
JAP – MEX								
Low Tier	23.23	18.96	42.18	18.96	-5.332	-5.431	-5.545	1
High Tier	25.02	14.54	39.57	14.54	-5.997	-6.109	-6.191	1

Notes: Critical values for the maximum eigenvalue for r = 0 are 14.07 (5%) and 18.63 (1%); and r = 1 are 3.76 (5%) and 6.65 (1%). Critical values for the trace statistic for r = 0 are 15.41 (5%) and 20.04 (1%). For $r \leq 1$ are 3.76 (5%) and 6.65 (1%).

From a methodological standpoint, one might question whether or not segregating stocks based on market capitalization is appropriate. The results from this analysis suggest that it not only is appropriate, but that it provides significant information

about risk. Additional research could extend this by exploring volatilities based on company size and market capitalization.

Finally, I used a rolling vector error correction technique to evaluate the stability of the long run relationship. I find that during the time of the most recent global financial crisis, the cointegrating parameter was unstable. While the finding in and of itself does not violate the efficient market hypothesis, it does nevertheless present an alert investor to possible arbitrage opportunities in crisis situations. In this essay, I do not explore reasons for this instability, but rather, I leave that for another researcher or time.

References

Aggarwal, R., & Kyaw, N. A. (2005). Equity market integration in the NAFTA region: Evidence from unit root and cointegration tests. *International Review of Financial Analysis, 14*(4), 393-406.

Al Nasser, O. M., & Hajilee, M. (2016). Integration of emerging stock markets with global stock markets. *Research in International Business and Finance, 36*, 1-12.

Alotaibi, A. R., & Mishra, A. V. (2017). Time varying international financial integration for GCC stock markets. *The Quarterly Review of Economics and Finance, 63*, 66-78.

Atteberry, W. L., & Swanson, P. E. (1997). Equity market integration: The case of North America. *North American Journal of Economics & Finance, 8*(1), 23.

Aznar, A., & Salvador, M. (2002). Selecting the rank of the cointegration space and the form of the intercept using an information criterion. *Econometric Theory, 18*(4), 926-947.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics, 9*(1), 3-18.

Basu, S. (1997). Why institutional credit agencies are reluctant to lend to the rural poor: A theoretical analysis of the Indian rural credit market. *World Development, 25*(2), 267-280.

Bekaert, G., & Harvey, C. (2000). Foreign speculators and emerging equity markets. *The Journal of Finance, 55*(2), 565-613.

Byers, J. D., & Peel, D. A. (1993). Some evidence on the interdependence of national stock markets and the gains from international portfolio diversification. *Applied Financial Economics, 3*(3), 239-242.

Chari, A., & Henry, P. B. (2004). Risk sharing and asset prices: evidence from a natural experiment. *The Journal of Finance, 59*(3), 1295-1324.

Chen, G., Firth, M., & Rui, O. M. (2002). Stock market linkages: evidence from Latin America. *Journal of Banking & Finance, 26*(6), 1113-1141.

Ciner, C. (2006). A further look at linkages between NAFTA equity markets. *The Quarterly Review of Economics and Finance, 46*(3), 338-352.

Climent, F., & Meneu, V. (2003). Has 1997 Asian crisis increased information flows between international markets. *International Review of Economics & Finance, 12*(1), 111-143.

Damianov, D. S., & Escobari, D. (2016). Long-run equilibrium shift and short-run dynamics of US home price tiers during the housing bubble. *The Journal of Real Estate Finance and Economics, 53*(1), 1-28.

Darrat, A. F., & Zhong, M. (2001). *Equity Market Integration and Multinational Trade Agreements: The Case of NAFTA.* Presentation To The 2001 Annual Meeting Of Financial Management Association International, Toronto, Canada.

Darrat, A. F., & Zhong, M. (2005). Equity market linkage and multinational trade accords: The case of NAFTA. *Journal of International Money and Finance, 24*(5), 793-817.

De Hoyos, R. E., & Iacovone, L. (2013). Economic Performance under NAFTA: A Firm-Level Analysis of the Trade-productivity Linkages. *World Development, 44*, 180-193.

De la Torre, A., Gozzi, J. C., & Schmukler, S. L. (2007). Stock market development under globalization: Whither the gains from reforms? *Journal of Banking & Finance, 31*(6), 1731-1754.

DeJong, D. N., Nankervis, J. C., Savin, N. E., & Whiteman, C. H. (1992). The power problems of unit root test in time series with autoregressive errors. *Journal of Econometrics, 53*(1-3), 323-343.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association, 74*(366a), 427-431.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.

Ewing, B. T., Payne, J. E., & Sowell, C. (1999). NAFTA and North American stock market linkages: an empirical note. *The North American Journal of Economics and Finance, 10*(2), 443-451.

Ewing, B. T., Payne, J. E., & Sowell, C. (2001). Transmission of conditional stock return volatility across North American markets: evidence from pre-and post-NAFTA. *The International Trade Journal, 15*(4), 409-427.

Fama, E., & French, K. (1992). The cross‐section of expected stock returns. *The Journal of Finance, 47*(2), 427-465.

Fama, E. & French, K. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance, 50*(1), 131-155.

Fama, E. F., & French, K. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics, 105*(3), 457-472.

Gilmore, C. G., & McManus, G. M. (2004). The impact of NAFTA on the integration of the Canadian, Mexican, and US equity markets *North American Economic And Financial Integration* (pp. 137-151): Emerald Group Publishing Limited.

Gonzalo, J., & Pitarakis, J.-Y. (1998). Specification via model selection in vector error correction models. *Economics Letters, 60*(3), 321-328.

Granger, C. W. J. (1986). Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economics and Statistics, 48*(3), 213-228.

Gregory, A. W., & Hansen, B. E. (1996a). Practitioners corner: tests for cointegration in models with regime and trend shifts. *Oxford Bulletin of Economics and Statistics, 58*(3), 555-560.

Gregory, A. W., & Hansen, B. E. (1996b). Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics, 70*(1), 99-126.

Hamao, Y., Masulis, R. W., & Ng, V. (1990). Correlations in price changes and volatility across international stock markets. *The Review of Financial Studies, 3*(2), 281-307.

Hansen, H., & Johansen, S. (1993). Recursive estimation in cointegrated VAR-models. 1-25.

In, F., Kim, S., Yoon, J. H., & Viney, C. (2001). Dynamic interdependence and volatility transmission of Asian stock markets: Evidence from the Asian crisis. *International Review of Financial Analysis, 10*(1), 87-96.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control, 12*(2), 231-254.

Johansen, S., & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and Statistics, 52*(2), 169-210.

Koch, P. D., & Koch, T. W. (1991). Evolution in dynamic linkages across daily national stock indexes. *Journal of International Money and Finance, 10*(2), 231-251.

Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*(1-3), 159-178.

Lahrech, A., & Sylwester, K. (2013). The impact of NAFTA on North American stock market linkages. *The North American Journal of Economics and Finance, 25*, 94-108.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of Financial Economics, 82*(3), 631-671.

Masih, A. M. M., & Masih, R. (2002). Propagative causal price transmission among international stock markets: evidence from the pre- and postglobalization period. *Global Finance Journal, 13*(1), 63-91.

Meric, G., Leal, R. P. C., Ratner, M., & Meric, I. (2001). Co-movements of U.S. and Latin American equity markets before and after the 1987 crash. *International Review of Financial Analysis, 10*(3), 219.

Okada, K. (2013). The interaction effects of financial openness and institutions on international capital flows. *Journal of Macroeconomics, 35*, 131-143.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society*, 1361-1401.

Perron, P., & Vogelsang, T. J. (1992). Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business & Economic Statistics, 10*(3), 301-320.

Phengpis, C., & Swanson, P. E. (2006). Portfolio diversification effects of trading blocs: The case of NAFTA. *Journal of Multinational Financial Management, 16*(3), 315-331.

Phylaktis, K., & Ravazzolo, F. (2002). Measuring financial and economic integration with equity prices in emerging markets. *Journal of International Money and Finance, 21*(6), 879-903.

Ratanapakorn, O., & Sharma, S. C. (2002). Interrelationships among regional stock indices. *Review of Financial Economics, 11*(2), 91-108.

Reinganum, M. R. (1983). The Anomalous Stock-Market Behavior of Small Firms in January - Empirical Tests for Tax-Loss Selling Effects. *Journal of Financial Economics, 12*(1), 89-104. doi:Doi 10.1016/0304-405x(83)90029-6

Taylor, M. P., & Tonks, I. (1989). The internationalisation of stock markets and the abolition of UK exchange control. *The review of Economics and Statistics*, 332-336.

Umutlu, M., Akdeniz, L., & Altay-Salih, A. (2010). The degree of financial liberalization and aggregated stock-return volatility in emerging markets. *Journal of Banking & Finance, 34*(3), 509-521.

Yu, I.-W., Fung, K.-P., & Tam, C.-S. (2010). Assessing financial market integration in Asia – Equity markets. *Journal of Banking & Finance, 34*(12), 2874-2885.

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance, 61*(1), 105-137.

Život, E., & Andrews, D. (1992). Further evidence on the Great Crash, the oil price shock, and the unit root hypothesis. *Journal of Business and Economic Statistics, 10*(3), 251-270.

How Efficient Can R&D Activities Be? A Study Based on High-tech Industry in China

Shuangjin Wang, Tianjin University of Commerce, China
Richard Cebula, and Maggie Foley, Jacksonville University

Abstract

In this study, the influence of environmental factors on regional technological collaborative innovation is to be examined. This evaluation is done using the tracking data developed for the advanced technology industry in Beijing area from 1995 to 2015. The empirical findings based on the three-stage DEA-Windows model reveal that the input and output for the R&D activities are strongly related to external environmental factors. The technological infrastructure, the regional economic relevancy, the regional trade dependency, the technological infrastructure investment intensity, and the regional comparative labor productivity have significantly negative influences on funding for R&D activities and new product development.

Introduction

Information gathered from the China Innovative and Development Report (2015), the National Innovation Blue Book, indicates that China is ranked behind the U.S and Japan in the area of advanced technology in terms of the number of patents per 1,000 R&D personnel and in terms of the total number of patents. The most significant difference is observed in the value of innovation efficiency. In China, the urban agglomeration for the three industrial areas of Beijing, Tianjin, and Hebei is less developed when compared to the agglomerations of the Yangtze River Delta and the Pearl River Delta. Regarding economic development, the Hebei province is ranked behind Beijing and Tianjin, especially in areas such as the advanced technology industry. These conclusions have been drawn from the values obtained after measuring the input-output efficiency of R&D activities, which is the most significant indicator of development in the advanced technology industry. Furthermore, it has been observed that the efforts aimed at attaining high agglomeration economies of scale effects in Beijing, Tianjin, and Hebei are limited by the existence of endogenous restrictions; hence, a minimal value is obtained from their independent creative elements, which reduces the impact of the intra-regional synergies, increases the intra-regional transaction costs and impedes the effective spillover of prevalent factors in the region.These issues have to be addressed in order to attain overall developmental progress (Zhang & Wen, 2017).The remarkable impact of the advanced technology industry in promoting development in the local economy has resulted in this industrybeing held in high regard by the authorities who realize the value it adds to both the investment sector and total factor productivity (TFP), which are two main areas that stimulate development in the high-tech industry. However, the stakeholders tend to overlook the efficiency related to the application of over-stretched innovative resources. An analysis of the latter perspective is also notably absent in the research literature. We fill up this gap in the literature by exploring the influence of the environmental factors on innovation efficiency by considering the data sourced from the Beijing, Tianjin, and Hebei region. Further evaluations are used to identify better ways to carry out the efficient allocation of resources during the process of innovation. The answers provided to the question above reveal ways in which to improve the utilization of inputs and outputs so as to promote R&D activities in the high-tech industry in China.

The structure of this paper is as follows: Section 2 includes the literature review, Section 3 analyzes the theory and model, Section 4 provides the empirical results and Section 5, the conclusion.

Literature Review

In recent years, the Chinese economy has gradually developed so as to become driven by innovation rather than certain other factors. Naturally, there are challenges arising from the use of innovation to drive regional economic development (Bai & Jiang, 2015; Wang, 2017, Sun et al. 2017). In the literature, the roles of a variety of factors related to innovation have been documented, among which are included input and output factors such as human resource investment, the level of R&D investment, the number of patent applications and new products sales, and government funding, and so forth. However, the findings regarding the effects of these factors are mixed. On the other hand, it has been established that the intensity of R&D expenditure and the extent of scientific research activities could be regarded as the most significant factors which limit the output of technological innovation (Zhu, 2016; Dang, 2015; Li, 2014). The number of patent applications and new product sales (Cooper et al., 2007), as well as the firm size and industrial characteristics (Horta et al., 2012; Akgun et al., 2007; Wang, 2011), also exercise significant roles in technological innovation. Li et al. (2015) find that there has been a positive impact on

the advancement of technological innovation by actions such as the provision of subsidies by the government and financial institutions. Sipos et al. (2014) discovered that factors such as innovative opportunities and market information are significant for industries that possess strong innovative dimensions. And for the industries with weaker innovative capacities, there might be limitations in their ability to secure adequate funding, which is a significant factor that promotes growth. Zhu and Wang (2015) argues that the huge financial requirements needed to enter the industry are part of the principal considerations in the aerospace industry, when compared with the proportion of R&D personnel which has the highest impact and the export ratio which stands out as the least relevant factor in the computer industry. Yung-Ching et al. (2011) argue that the impact of front-end innovation is crucial to promoting innovation. Hence, they propose the advancement of sustainable,cutting-edge technology. Zheng (2014) points out that R&D investments, organizational capacity, manufacturing levels, decision-making abilities and marketing efforts also promote technological innovation. But the influence of these factors tends to be eroded over the years.

Model Selection and Data Description

Model selection

To evaluate the effectiveness of innovation in high-tech industry, several methods have been applied in the literature. The most popular methods are the BC^2-DEA model (Wang & Zhou, 2016), the series DEA model (Yu, 2015), the BCM-DEA model (Chen, 2014), and the SBM and DEA model (Chen et al., 2014). Lin & Tan (2016) uses the meta-frontier DEA model to measure the four energy-intensive sub-sectors with energy-saving and energy-saving potential in China (Fan, 2014; Gao, 2010). In addition, other approaches such as a two-stage DEA model (Kao & Hwang, 2008; Halkos et al., 2014), a three-stage DEA model (Liu & Li, 2016; Qiao et al., 2017), a gray relational analysis model (Pei Xiao and Chen Junliang, 2013), and a structural equation model (Yu et al., 2015) are highlighted in studies where innovation performance in the high-tech industry have been evaluated. The major issue with certain of these approaches involves the attributes of the model, which are all specified as linear. However, the determinants of the innovation performance may not have linear correlations with innovation. From this perspective, a three-stage DEA model, which is a non-linear model, is used in the present study. Since we use tracking data in a small region in China, this model is suitable in that the dynamic path of the regional synergistic development is not decided yet. In addition, we use the DEA-Windows and the SFA methods for the evaluation of the external environment determinants for collaborative innovation.

The Three Stage DEA Model

An issue that arises with the DEA model is the difficulty in estimating the efficiency of the decision-making unit (DMU) due to the impacts of the environmental variables and other factors which have not been included in the model. Fried H.O. et al. (2002) proposed a three-stage DEA model which integrates the DEA model with the SFA model. While it resolved the issues regarding the classical DEA model, it is not applicable to the time series data for dynamic innovation development processes. In particular, the DEA window analysis is suitable to evaluate the dynamic evolution trend of efficiency through mobile smoothing methods (Nguyen, 2014; Cheng, 2011). Based on this model, different DMUs can be developed alongside the same DMU at different stages. In summary, we have combined a three-stage investment oriented DEA model with a DEA-Windows analysis method. We subsequently proceed to separate the evaluation process for the efficiency of DMU into three stages.

In the first stage, we develop a DEA-Windows model. The DEA is characteristically a non-parametric method, which can be used to evaluate similar decision-making units having multiple inputs and outputs. In this study, we used a Window-I-V method for which the time window was set at 3. For the input, we first calculate the initial DEA efficiency for the remaining environmental factors, and the next step is the addition of a slack variable to the second stage of the dependent variable.

In the second stage, we develop an SFA model adopting a similar approach. Fried (2002) observes that the performance of the decision-making unit is simultaneously affected by the management efficiency, environmental factors, and statistical noise. Fried (2002) proposed that the three influences should be separated. Hence, we decomposed the first stage input relaxation variables into three functional variables: the environmental factors, the random interference, and the management inefficiency factors. The equation is represented as follows:

$$S_{ni} = f(Z_i;\beta_n) + \nu_{ni} + \mu_{ni} \tag{1}$$

where; S_{ni} is the slack value of the *i*-th DMU ； Z_i is the environment variable; β_n is the coefficient of the environment variable; $\nu_{ni} + \mu_{ni}$ is the mixed error term; ν_{ni} is the random interference; μ_{ni} is the management inefficiency; and $\nu \sim N(0,\sigma_v^2)$ is the random error term which indicates the effect of the random disturbance factor on the slack variable; μ is the management

inefficiency, indicating the effect of the management factor on the slack variable, assuming that it follows a normal distribution with a mean of zero, i.e.., $\mu \sim N^{+}(0,\sigma_{\mu}^{2})$.

We use the SFA regression to eliminate the influence of the environmental factors and other stochastic factors on efficiency measurement. First, we separate the management inefficiency μ . We use Jondrow(1982) and Liu (2016) to estimate the random interference v_{ni} .

$$E\left[v_{ni}\left|v_{ni}+\mu_{ni}\right.\right]=s_{ni}-f(z_i;\beta_n)-E\left[u_{ni}\left|v_{ni}+\mu_{ni}\right.\right] \tag{2}$$

$$E\left(\mu\left|\varepsilon\right.\right)=\sigma_{*}\left[\frac{\phi(\lambda\frac{\varepsilon}{\sigma})}{\Phi(\frac{\lambda\varepsilon}{\sigma})}+\frac{\lambda\varepsilon}{\sigma}\right] \tag{3}$$

where $\sigma_{*}=\frac{\sigma_{\mu}\sigma_{v}}{\sigma}$, $\sigma=\sqrt{\sigma_{\mu}^{2}+\sigma_{v}^{2}}$, $\varepsilon=s_{ni}-f(z_i;\beta_n)$, $\lambda=\sigma_{\mu}/\sigma_{v}$. $\phi(\bullet)$ and Φ （$\bullet$） are the density functions and distribution functions of the standard normal distribution, respectively.

Second, we use the following model to adjust the external environment for the inputs of each DMU:

$$X_{ni}^{A}=X_{ni}+[\max(f(Z_i;\hat{\beta}_n))-f(Z_i;\hat{\beta}_n)]+[\max(\hat{v}_{ni})-\hat{v}_{ni}] \quad i=1,2,\cdots,I;n=1,2,\cdots,N \tag{4}$$

where X_{ni}^{A} is the adjustedinput; X_{ni} is the input before the adjustment; $[\max(f(Z_i;\hat{\beta}_n))-f(Z_i;\hat{\beta}_n)]$ is the case where the *n*-th item is adjusted to the most affected external environmental factors ; and $\max(\hat{v}_{ni})-\hat{v}_{ni}$ is the amount of input for the DMU to increase the maximum random interference. The above approach is based on the assumption that each DMU is placed in a homogeneous external environment in a natural state, thereby the environmental factors and the impact of random factors on efficiency are eliminated.

In the third stage, we adjust the DEA-Windows model. Each DMU is adjusted for the input data X_{ni}^{A} to replace the original input data X_{ni} . For the original output data, we use the input-oriented DEA-Windows model to re-calculate the efficiency of each DMU. Notably, the efficiency measure is at the management level due to the above treatment of the environmental factors and the stochastic factors. The results based on such methods can reflect the technical efficiency of each DMU more objectively. The results derived from the SFA regression will be adjusted to accommodate the input slack variable for the environnent variable.

Variable Selection and Data Description

Sample Selection and Data

We retrieved the innovation data and the relevant environmental variables from the Statistical Yearbook of China's High-Tech Enterprises (1996-2016), the China Science and Technology Statistical Year book (1996-2016), China Torch Statistical Year book (1996-2016), the China Highway Information Network, and other high-tech industry science and technology statistics reports for the areas of Beijing-Tianjin-Hebei from 1996 to 2016. Our sample has seven R&D cycles and 945 observations. Each cycle has a two-year gap, with the first cycle beginning in1995 and ending in 1997.

The Collaborative Environmental Variables

The environmental variables represent external factors that can determine the influence of the efficiency of input and output of R&D activities in a high-tech industry. These values can neither be controlled nor modified within a limited time. Regarding the situation in the Beijing, Tianjin and Hebei provinces, our focus is directed at the following five areas: technology acquisition structure; regional economic relevance; regionals (domestic) trade dependence; the intensity of scientific and technological infrastructural investment; and regional comparative labor productivity. Theseare the environmental variables to be considered in the SFA analysis.After exploring these issues, we selected ten more variables. They are, as follows: the introduction of technology, the consumption of funds, the total population, the regional GDP, the distance between the core cities, the net exports of products and services, the differences in international trade, the number of investments in science and technology infrastructure, the geographic area, and the number of employees. The calculations for the main environmental variables from the perspective of synergistic development are as follows:

(1) Technology Acquisition Structure (TAS).

TAS is quickly applied in introducing, digesting, and absorbing of funds for innovation. Following the literature, we

calculate TAS as follows.

$$TAS_{ij} = DAF_{ij} / TTF_{ij} \tag{5}$$

where TAS_{ij} represents the technical acquisition structure of region *i* in year *j*; DAF_{ij} and TTF_{ij} are the digestible funds and technology introduction funds for the region*i* in year *j*, respectively. This index is less than 1. The lower it is, the more it reflects how much it places emphasis on hardware but less on software. It also shows how important the transferring of a complete set, keyequipment,and explicit knowledge are in this region. Of note, it is far from sufficient to only transfer the tacit knowledge, such as the experience of the management and the skills of the talented employees.

(2) Correlation of regional economy (CRE).

CRE represents the close economic ties among regions. It is expressed as follows.

$$CRE_{ij} = \sqrt{P_i G_i} \times \sqrt{P_j G_j} / D_{ij}^2 \tag{6}$$

CRE_{ij} is the economic correlation strength among regions; G_i and G_j are the population in regions *i* and *j,* respectively; P_i and P_j are the gross product values in regions *i* and *j,* respectively; D_{ij} represents the shortest driving distance between main cities in each region. If the latter value is very high, it will tend to cause a corresponding increase in the degree of the coordinated economic development.

(3) Interregional trade dependence (ITD).

Interregional trade dependence (ITD) reflects the degree of trade dependence and the level of regional economic integration within the region, which is extremely important to allow for (Cebula, 1994). According to the current national economic accounting system, the GDP of the net exports of products and services include the import and export values for foreign trades, and the interregional trade values for products and services. The interregional (domestic) trade dependency can be expressed as follows.

$$ITD = \frac{B_r}{GDP} = \frac{N - B_i}{GDP} \tag{7}$$

Where B_r is the interregional trade; N is the net export of products and services and B_i is the international trade balance. The greater the absolute value of $N - B_i$, the more accurate the regional economic exchanges and the higher the level of cooperation.

(4) Technical infrastructure investment intensity (TIII).

The scope of infrastructural investment includes the investment in developing general and technological infrastructure (Tong and Wu, 2016). They include the investment in science and technology infrastructure, investment in information technology and scientific research infrastructure and the investment in fixed assets regarding the investments in scientific research and technical services and information dissemination, software and information technology services. In addition, the investment intensity to promote scientific and technological infrastructure is expressed by the ratio of investment in science and the technologicalinfrastructure as well as the geographical size of the various areas. With this variable, a proper measurement of the efficiency in the use of land in high-tech development zones is attainable. The equation is represented as:

$$TIII_{ij} = \frac{TII_{ij}}{Area_i} \tag{8}$$

where $TIII_{ij}$ represents the investment intensity of the technology infrastructure in area *I* and year *j*; TII_{ij} represents the investment in science and technology infrastructure in area *i* and year *j*; The greater the gap of $TIII_{ij}$ among Beijing and Tianjin and Hebei, the greater the distribution of innovative resources in the high-tech industry in this overall region. A higher gap also indicates the regional imbalance and the urgent need for synergies of science and technology resources in this area.

(5) Regional comparative labor productivity (RCLP):

RCLP refers to the proportion of the output of different production sectors or industries in one region and the proportion of labor force in this sector or industry. It is usually used to reflect the level of labor productivity in different industries. The equation is as follows.

$$RCLP_i = \frac{GDP_i / GDP}{Labor_i / Labor} \tag{9}$$

Where GDP_i and GDP represent the sum of the GDP of the region i and the GDP of all regions, respectively; $Labor_i$ and $Labor$ represent the number of labors in the area i and total labor. The greater the numerical gap of RCLP among Beijing and Tianjin and Hebei, the greater the productivity differences in interregional employment. It also points out the imbalance in labor productivity. The existence of a significant spatial differentiation indicates the need for more focus on regional cooperation in science and technology and regional differences in labor effectiveness.

Input-output Indicators

The previous studies were known to consider the number of employees and the R&D expenditure as measures for the internal expenses. However, this approach could not be used to objectively consider the time invested by the R&D personnel. For this reason, the R&D activities were considered instead. Furthermore, the aspects of R&D funding for new products were included as well as the R&D expenditure and the total R&D costs for new products. We also applied patent and the sales of new products as the output indicators since they could be used as an objective representation of technological innovation output capacity and the outcome of the R&D, respectively. In addition, the possible influence of R&D on the methods used by firms to raise capital, reorganize industrial structure, set up industrial ties, generate new trades, and make institutional changes suggests that R&D inputs and outputs can potentially last for a period of two to three years. Therefore, we set the value of the development cycle at three years. In Table 1, a report of the R&D input and output for the different cyclesis is shown.

Table 1: Input and Output for R&D Activities

		Input Indicators			**Output Indicators**	
Innovation Cycle	**Region**	**R&D Staff Full Time Equivalent (person years)**	**R&D Expenses (10,000 yuan)**	**New Product Development Expenses (10,000 yuan)**	**New Product Sales Revenue (10,000 yuan)**	**Number of Patent Applications (Pieces)**
1995-1997	Beijing	12864	88367	93942	1554243	79
	Tianjin	7044	21905	24172	762396	69
	Hebei	1747	18439	20358	20358	29
1998-2000	Beijing	13658	175097	240440	5699934	123
	Tianjin	6553	88409	95207	3263246	191
	Hebei	4201	42912	33521	33521	53
2001-2003	Beijing	17355	585791	388835	10891475	2044
	Tianjin	6909	228879	88246	13057870	578
	Hebei	7257	75501	57433	57433	77
2004-2006	Beijing	22567	808564	751724	12795377	2400
	Tianjin	9092	305359	203808	27878031	1229
	Hebei	9509	100070	87062	87062	360
2007-2009	Beijing	29918	1044052	1224548	43716736	7232
	Tianjin	16510	570219	671691	25173328	4244
	Hebei	12160	168857	157483	157483	910
2010-2012	Beijing	46897	2032608	2752842	41656269	19001
	Tianjin	28893	934097	979501	27997558	8094
	Hebei	20903	363352	360882	360882	1497
2013-2015	Beijing	69792	3375278	4562974	50473833	25051
	Tianjin	53964	1785155	1566324	54488453	10479
	Hebei	33610	901688	835018	835018	3283

Source: from China Science and Technology Statistical Yearbook, 1996-2015

Empirical Results
First stage of the DEA-Windows Model

Before using the DEA model, the conditions for the "co-directional" input and output indicators must be met. This means there cannot be a decrease in the output while the input increases. The assessment of this requirement is done by carrying out a Pearson correlation test of the selected input-output indicators. The results of this test are shown in Table 2. From these results, it is inferred that the Pearson correlation coefficients which exist between the input index and output index are statistically significant at the 1% level. This finding shows the "co-directional" requirement of the DEA efficiency analysis has been met.

Table 2: Pearson Correlation Coefficient of Input and Output Indexes of R&D Activity

Index	R&D Full Time Equivalent	R&D Expenses	New Product Development Costs
Sales of New Products	0.802*	0.825*	0.774*
	(0.000)	(0.000)	(0.000)
Number of Patent Applications	0.919*	0.972*	0.985*
	(0.000)	(0.000)	(0.000)

Note: * indicates statistical significance at 1% level, and the value in parentheses is the P value of the significance test.

Next, we focus on the input and output data for the seven research cycles and the necessary modifications are necessary to enable us to use the DEA-Solver-Pro 11. For this part, the time window is also set at 3. We use the DEA and the DEA-Windows to illustrate the robustness of the results. The results are shown in Table 3.

Table 3: Comparison of Efficiency of R&D Activities Based on DEA Methods (Prior to Adjustment)

Class	Region	Innovation Cycle						
		1995-1997	1998-2000	2001-2003	2004-2006	2007-2009	2010-2012	2013-2015
Classical DEA	Beijing	0.463	0.777	0.866	0.526	1.000	1.000	0.886
	Tianjin	0.529	0.453	1.000	1.000	1.000	1.000	0.830
	Hebei	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DEA-Windows	Beijing	0.483	0.575	0.942	0.678	0.952	1.000	0.887
	Tianjin	0.932	0.827	1.000	0.996	0.996	1.000	0.985
	Hebei	0.701	0.533	0.699	0.774	0.807	0.626	0.521

The contents of Table 3 indicate that the use of a DEA model renders the R&D activity in Hebei province equals to 1. However, this representation does not agree with what is commonly applicable. Moreover, the use of the DEA model hinders our attempts to study the changes in efficiency over the seven innovation cycles. Thus, the DEA model consequently is used to carry out a weak dynamic assessment. Furthermore, the DEA-Windows method enables us to find the efficiency of R&D activities and the dynamic evolution of the efficiencyof R&D activities in the three areas. This is a better assessment of the dynamic adjustment of R&D resources.

By using the input-oriented Window-IV approach, we discover that the measurement of efficiency in the three areas has values above 0.7 as regards the seven research cycles within the period of 1995-2015. Beijing is ranked second in the three places. Its average efficiency of R&D activities is 0.788. The highest efficiency innovation was recorded between 2010-2012, while the lowest innovation efficiency was within the period 1995-1997 which is averaged at 0.483. Tianjin ranks the first with respect to the efficiency of R&D activities. Its highest efficiency reaches 0.962from 1995 to 1997and its lowest efficiency is 0.827 from 1998 to 2000. Hebei Province has the lowestefficiency of R&D activities as reported for the three areas. Its highest value is 0.810from 2013 to 2015,and the lowest efficiency is 0.667 from 1998 to 2000. In this overall region, Beijing falls behind Tianjin regarding the evaluation of innovation efficiency by using the patents and new product sales. These findings could be due to inefficiency in allocating resources in Beijing since Beijing is the capital city of China and thus it is the center for scientific and technological innovation resources.

In addition, we find that there is a big difference in the efficiency values of R&D activities in the Hebei province between the traditional DEA model and the DEA-Windows model. The results based on DEA-Windows method are more reliable because this method addresses the issues with the window width, and it can accommodate multiple decision-making units

simultaneously. It also is a good choice when the sample size is too small. In comparison, DEA model's time window is set at 1, resulting in large errors at the end of the process.

Second Stage of the SFA model

Due to the inability of the classic DEA model to isolate the environmental factors, our focus is directed at the expenses generated by the full-time R&D personnel, and other R&D expenditures, as well as new product development funds. In addition, we included the three input variables separately for the slack and radial aspects. The five environmental variables include the technological acquisition structure, regional economic relevancy, inter-regional (domestic) trade dependency, the investment infrastructure intensity in science and technology, and the regional comparison in labor productivity. These variables are included to examine the impact of the three input slack variables. Of note, the environmental variables are dummy variables. The reports are illustrated in Table 4.

Table 4: SFA Regression Estimation Results in the Second Stage

Variables	**R&D Investment Slack Variable**	**R&D Expenditure Slack Variable**	**New Product Development Fund Slack Variable**
TAS	-0.5398 (0.0286)	-1.0133*** (0.3111)	-0.2526 (0.0396)
CRE	-0.4729 (0.1686)	-0.8903 (0.2211)	-0.0323 (0.0101)
ITD	-16.9284*** (1.6866)	-12.5730*** (73.3872)	-11.6087*** (3.7130)
TIII	6.6818 (2.3493)	-10.1333*** (3.6667)	-9.2481*** (3.7284)
RCLP	-0.2098 (0.0296)	-0.2409*** (0.0319)	-0.1677* (0.0720)
Constant Term	2.3429*** (0.0766)	1.8948*** (0.3516)	1.3662*** (0.3179)
σ^2	0.1342*** (0.0413)	0.1730 (0.0136)	1.4733 (0.1591)
γ	0.0562* (0.0197)	0.5549*** (0.0503)	0.9596*** (0.0491)
Logarithmic Likelihood Function	-8.7091	-4.9877	-5.7473

Note: *, *** respectively indicate the significance level of 10% and 1% levels, respectively. Standard errors are reported as shown in the brackets.

The content in Table 4 shows that the five environment variables on the three input variables as redundancy value coefficient are most statistically significant. Furthermore, the γ value of the total variance, used for characterization of management inefficiency of variance, is statistically significant at the 10% and 1% levels. This shows the significant role of the environmental factors which influence the R&D input slack variable for the regional coordinated development of the high-tech industry in the Beijing-Tianjin-Hebei region. Further analysis reveals that the γ values of the three kinds of input relaxation variables have different features, among which, γ value for the R&D personnel input slack variable is 0.0562.This finding is an indication that the random disturbance factors dominate the slack variable. Moreover, the γ value of the slack variable in R&D expenditure is 0.5549, an indication that its slack variable is influenced by the management inefficiency and other random factors. The γ value of the slack variable for new product development funding is 0.9596, indicating the impact of the management inefficiencies on its relaxation variables. The analysis presented above implies the importance of using the SFA model to effectively eliminate the effect of the environmental factors on R&D activity efficiency. The following is a brief analysis which reveals the impact of the environmental variables on the R&D investment slack variable in the Beijing-Tianjin-Hebei high-tech industry.

(1) Technology Acquisition Structure (TAS):

A positive coefficient is derived for this variable while using the R&D personnel in the model. A comparison also shows that it is negative when the presentation is made for the relaxation variable of the R&D expenditure and new product development. This evidence is an indication of the need to improve efficiency in utilizing funds and reducing the learning curve for the new technology in order to attain a better management funding for the R&D activities and the new product development. In addition, an increase in the cost of technology absorption and the value of TAS will remarkably increase the work needed to be done by the researchers regarding the relevant technical applications. It is commonly acknowledged that the total absorption fees include the import costs. Currently, the ratio of funds attributed to the introduction of new technology and the technology absorption fees is given as 0.362 in the Beijing-Tianjin-Hebei region. This ratio is averaged as 3 to 1 globally, and it reaches 7 to 1 in developed countries. The evidence above is an indication that the introduction of advanced technology could to a large extent limit the developmental processes and the promotion of new technology in the Beijing-Tianjin-Hebei region.

(2) Correlation of regional economy (CRE):

Table 4 shows the coefficients for the slack variables and the R&D variables. Negative values have been obtained for the R&D funds and new product development funding. This indicates that an increasing regional economic cooperation causes a free flow of the talents needed to develop resource allocation efficiency while eliminating any trace of redundant R&D personnel. The advancement in the regional integration also improves the positive impact on the efficiency of science and technology innovation, which also promotes the inter-regional impact of innovation factors. Achieving a better method to allocate talents and funding increases the ability to improve the elements of innovation attributed to the allocation structure in the inflow areas such as the funds, talents, improvement in the marginal returns from elements of industrial and regional innovation, enhancement of the agglomeration results, and the increase in productivity. Moreover, the spillover effects have caused a free flow of innovation factors within the region, which influences the results attained after technology sharing and collaborative innovation in this area.

(3) Interregional trade dependence (ITD):

From the results, it is observed that the ITD has a statistically significant negative impact on the slack time variables of R&D personnel, the R&D expenditure and the new product development funds. Assessing the level of integration in the overall Beijing-Tianjin-Hebei region will include considering the level of inter-regional trade dependence. From one perspective, areas such as industrial specialization can be improved by exportation resulting in comparative advantages in the area of resource endowment and industrial chain layout. From another perspective, the regional collaboration could potentially remove the boundaries and restrictions which exist in the three places to promote the unhindered access to scientific and technological resources as well as innovative elements, thereby alleviating constraints in accessing natural resources, markets, and human capital hence a reduction in the R&D personnel redundancy. Achieving the promotion of interregional trade can provide a way to resolve issues attributed to insufficient market demand in the region, thus creating a larger market where production will be increased along with other effects such as increasing economies of scale, reduced R&D funding and lower new product development costs for slack variables related to the enhancement of innovation efficiency in the technology industry.

(4) Technical infrastructure investment intensity (TIII):

The use of advanced information resources in the Beijing-Tianjin-Hebei region to increase the benefits thereof will require the creation of better scientific research infrastructure and an advanced interconnected “information highway” with the use of innovative technology. In Table 4, it is revealed that TIII shows a positive effect on the slack variables of R&D personnel total time equivalence, and the regression coefficients of R&D and new product development slack variables are negative. The best explanation for this is that from one perspective there exists a huge gap between the incubator stage during product development, which requires technological infrastructure, and the stages for promotion and market recognition. Furthermore, the inclusion of technology infrastructure will increase the workload of the technical personnel, causing an avoidable redundancy of the resources. In addition, the enhancement of information technology infrastructure can reduce communication costs, particularly those related to the R&D personnel, where information is quickly accessible to users and there is an unlimited communication line among users without territorial boundaries. The “space compression effect” in the Beijing-Tianjin-Hebei region promotes the spillover effects attributed to the collaborative innovation technology which was caused by the interregional development based on the variant elements. Hence, it can improve the factors of interregional development related to the competitive incentive effect, the economies of scale which will cause further improvement in the regional R&D activities and efficiency with the use of available funds to promote new product development.

(5) Regional comparative labor productivity (RCLP):

RCLP has a negative impact on the total time equivalent of R&D personnel, R&D funds input, and slack expenditure of new product development. It should also be noted that the slack expenditure of new product development has a statistically significant impact on R&D personnel and R&D funds input. This is an indication that via reducing the slack of labor equivalent of R&D workers, when the regional comparative labor productivity is increased, there will be a corresponding increase in the quality of the workers' scientific research. The inclusion of highly skilled workers will significantly improve the utilization of

investments in material, leading to a decrease in R&D funding and funding redundancy for new product development. Hence, optimizing the human resources structure can enhance labor productivity in the region to fully utilize human capital, promote innovation, and increase the supply of the public technology.

Empirical Results for the Adjusted DEA-Windows Model

The second stage is based on the coordinated development of the environmental factors. The parameters are estimated with the following formula:

$$X_{ni}^{A} = X_{ni} + [\max(f(Z_i;\hat{\beta}_n)) - f(Z_i;\hat{\beta}_n)] + [\max(\hat{v}_{ni}) - \hat{v}_{ni}] \tag{10}$$

The original inputs were modified based on DEA-Solver-Pro11.0. The new input-output data are estimated using DEA-Windows approach. Figure 1 shows the efficiency of the R&D activities for the innovation cycle of the high-tech industry in the Beijing-Tianjin-Hebei area under both the new and the old approaches.

Figure 1: Comparison between old and new approaches for efficiency of R&D activity for the innovation cycle of the high-tech industry in the Beijing-Tianjin-Hebei area

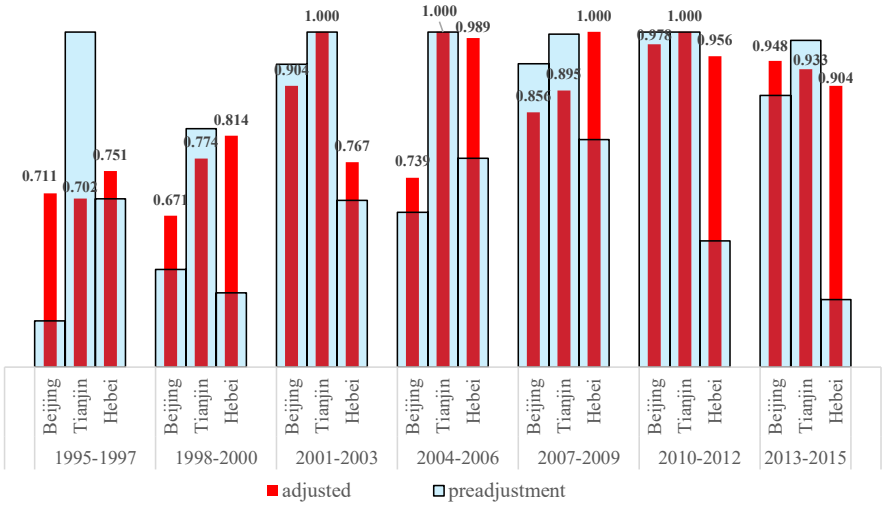


Figure 1 shows that input and output efficiency for R&D activities has remarkably evolved in the Beijing-Tianjin-Hebei high-tech sector. The Wilcoxon symbol rank test indicates that there are significant differences for R&D activity efficiency in the first and the third stages before and after the modifications. The evidence thus reveals that the efficiency of R&D activities in Beijing-Tianjin-Hebei high-tech industry cannot be neglected when the influence of environmental factors and random errors are ignored. Therefore, it is not possible to simulate the actual situation.

Furthermore, the efficiency of R&D activities tends to change with time due to the changes in innovation policies and the environmental innovation factors. However, the efficiency of R&D activities in the Beijing-Tianjin-Hebei region is relatively constant. In Beijing, its lowest point is 0.483 in the 1995-1997 innovation cycle; the highest point is 0.956 as recorded in the 2010-2012 cycle. In Tianjin, the highest point of the efficiency value is 0.596 within the period 1995-1997. Hebei province's lowest efficiency is 0.565 in the 1995-1997 cycle; its highest point is 1,as recorded in the 2007-2009 cycle.

From the perspective of spatial evolution, Tianjin city has witnessed the highest level of R&D activities, followed by Beijing and then by Hebei. Furthermore, there is a big difference in R&D activities between Beijing and Tianjin. But, there is an imbalance regarding the distribution of innovation resources in science and technology among the three places. After the adjustment for input factors, the average efficiency of Beijing is recorded as 0.769, with a reduction of 0.019 after the adjustment. The average efficiency of Tianjin is reduced from 0.962 to 0.832, while that of Hebei is increased from 0.665 to 0.752. This pattern of results shows that the development of high technology industry varies greatly in this region. Beijing and Tianjin have more capital,and theyare better equippedwith human resourcesand technology thanthe Hebei province. The gap is widening as time passes due to the siphoning effect. Once the environmental factors for collaborative innovation are eliminated, the efficiency of science and technology innovation in Beijing and Tianjin will experience a decrease. The results indicate that

separating the administrative divisions in the Beijing-Tianjin-Hebei area make the different places become solely interested in their benefits. The unbalanced innovation chain structure in science and technology makes the entire situation worse. Efforts aimed at improving innovation in this region should be more collaborative, and policies which could be beneficial should be adopted in the areas to promote the efficient management of resources.

Conclusion

In this study, the three-stage DEA-Windows model and the SFA analysis method have been used to evaluate the efficiency of the R&D activities in the high tech industry in three areas, namely, Beijing, Tianjin, and Hebei province from 1995 to 2015.Whileprevious studies used the DEA model to evaluate the values of innovation efficiency, we use the DEA-Windows analysis method to analyze the influence of the environmental factors on the collaborative innovation of R&D activities from the perspective of efficiency, with a particular focus on the dynamic path of the innovation efficiency. Furthermore, the SFA method was used to eliminate the interference factors from the external environment and the random errors. To a certain extent, this approach overcomes the problem mentioned in the literature regarding the provincial innovation efficiency and the associated dynamic evolution and stability as reflected objectively in the input-output efficiency of R&D activities which characterize the high-tech industry in one city. The main conclusions are as follows:

(1) The three-stage DEA-Windows model can be used to effectively eliminate the random disturbance when it is used to measure the input and output efficiency of environment factors for R&D activities in the high-tech industry in the Beijing-Tianjin-Hebei area.

(2) The SFA regression results indicate that the interregional trade dependency can be used to create a better time management system for R&D activities in the high-tech industry in the overall Beijing-Tianjin-Hebei region. The technological acquisition structure, the interregional trade dependency, the technological infrastructure investment intensity, and the effective utilization of labor productivity can significantly reduce the redundancy for funding for both R&D activities and new product developments. Furthermore, the degree of interregional trade dependency and the investment in the infrastructure in science and technology can significantly improve the utilization efficiency for new product development funds.

(3) The results of the adjusted DEA-Windows analysis show that after eliminating the environmental factors for the coordinated developments and the random errors, the input-output efficiency of R&D activities in Beijing and Tianjin tends to decrease, while the efficiency in Hebei province tends to increase.

(4) Regarding innovations, each place has witnessed significant changes due to changes in innovation policies and the environment. There are still on-going efforts to improve the efficiency of R&D activities in the Beijing-Tianjin-Hebei region, which can be shown to have much potential value.

References

Akgun A.E., Keskin H., Byrne J.C. et al. 2007. “Emotional and Learning Capability and Their Impact on Product Innovativeness and Firm Performance”. *Technovation* 9: 501-513.

Bai, J.H., Jiang, F.X. 2015. “Collaborative Innovation, Spatial Correlation and Regional Innovation Performance’. *Economic Research* 7: 174-187.

Cebula, R.J. 1994. “The Pedagogy of Teaching Interregional Income Determination for Interdependent Economies”. *Journal of Regional Analysis and Policy* 24(1): 72-81.

Chen, J.L., Meng, L.J., Jiang, C.L. 2014. “Research on Technological Innovation Efficiency of High Tech Industry from the Two Stage Perspective: Based on Network SBM Model and DEA Window Analysis”. *Science and Technology Management Research* 34 (11): 11-16.

Cheng, L.W., Sun, W., Wang, J.Y. 2011. “Elements of the Incomplete Market from the Perspective of the Innovation Efficiency of High Tech Industry -- Based on the Three-Stage DEA-Windows Funded Scale and Allocation Efficiency”. *Science Research* 29(6): 930-938.

Cooper, W. W., Seiford, L. M., & Tone, K. 200. *Data Envelopment Analysis*. New York: Springer.

Fan, Y.Q., Zhou, G. 2014. “Research on the Influencing Factors and Regional Linkage Effect of Technological Innovation Efficiency of China's High Tech Industry”. *Science and Technology Management Research* 21: 1-5.

Fried H.O., Lovell C.A.K., Schmidt S.S., 2002. “Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis”. *Journal of Productivity Analysis* 17(1-2): 157-174.

Gao, N., YU, W.C., Liang, P.H. 2017. “Structures, Legal Environment and Regional Innovation Activities”. *Journal of Scientific Research* 38 (2): 26-34.

Halkos G.E., Tzeremes N.G., Kourtzidis S.A. 2014. “A Unified Classification of Two-Stage DEA Models”. *Surveys in Operations Research & Management Science* 19(1): 1-16.

Horta I.M., Camanho A.S., Costa J.M.D. 2012. "Performance Assessment of Construction Companies: A Study of Factors Promoting Financial Soundness and Innovation in the Industry". *International Journal of Production Economics* 137(1): 84-93.

Jondrow, J., Lovell, C. A. K., Materov, I. S., & Schmidt, P. 198. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model". *Journal of Econometrics* 19(2-3): 233-238.

Kao, C.H., Wang, S.N. 2008. "Efficiency Decomposition in Two-Stage Data Envelopment Analysis: An Application to Non-Life Insurance Companies in Taiwan". *European Journal of Operational Research* 185(1): 418-429.

Li, H.D., Ma, W. 2014. "Research on the Factors Influencing the Innovation Performance of High-tech Industry from the Perspective of Investment". *Science and Technology Management Research* 10: 126-130.

Li, L., Wu, S. 2014. "Dynamic Evaluation and Comparison of Coordinated Development Level of Regional Economy in China Based on DEA". *East China economic management* 1: 65-68.

Li, X.D., Li, N., Liu, D.H. 2015. "Analysis of Influencing Factors on Innovation Efficiency of High Tech Industry". *Statistics and Decision Making* 6: 109-113.

Lin, B.Q., Tan, R.P. 2016. "Ecological Total-Factor Energy Efficiency of China's Energy Intensive Industries". *Ecological Indicators* 70: 480–497.

Liu, M.F., Li, S.H. 2016. "Research on the Innovation Efficiency of China's Hi-tech Development Zone Based on Three Stage DEA Model". *Management Review* 28(1): 42-52.

Nguyen T. P. T., Roca E., Sharma P. 2014. "How Efficient is the Banking System of Asia's Next Economic Dragon? Evidence from Rolling DEA Windows". *Applied Economics* 46(22): 2665-2684.

Qiao, Y.B., Wang, Y.Y. 2017. "Evaluation of Innovation Efficiency of Provinces in China Based on Three Stage DEA-Windows Analysis". *Science and Technology Management* 38(1): 88-97.

Tong, J., Wu, K.P. 2016. "The Change of Infrastructure Investment Structure in the Process of Economic Development". *Quantitative and Technical Economic Research* 12: 61-77.

Sipos G.L., Bzoi G., Ionescu A. 2014. "The Impact of Hampering Innovation Factors on Innovation Performance-European Countries Case". *Procedia-Social and Behavioral Sciences* 124: 415-424.

Sun, H., Wang, H. 2017. "Government Subsidy, R&D Investment and Enterprise Innovation performance-An Empirical Study Based on High-tech Enterprises on GEM". *Science and Technology Management Research* 37(12): 111-116.

Wang, L.J., Zhang, B.W., Zhu, C.L. 2011. "Industrial Characteristics, Macro-Control and Technological Innovation Strategy Selection-Evidence from China's Industrial Level". *World Economic Research* 3: 3-8.

Wang, L.P., Zhou, L. 2016. "Technical Innovation Efficiency Evaluation and Resource Allocation of High Tech Industry in Beijing, Tianjin and Hebei Province". *Science and Technology Management Research* 8(1): 7-13.

Wang, T., Wang, H. 2017. "The Influence of Geographic Neighbors on the Innovation Performance of Enterprises: Based on the Perspective of Enterprise Competence". *Business Economics and Management* 2: 25-32.

Yung-Ching Ho, Ching-Tzu Tsai. 2011. "Front End of Innovation of High Technology Industries: The moderating Effect of Front-End Fuzziness". *Journal of High Technology Management Research* 22: 47-58.

Zhang, G., Wen, K. 2017. "Cooperation Innovation, Regional Integration and Innovation Performance". *Science and Technology Progress and Countermeasures* 3: 35-39.

Zheng, X. 2014. "Research on the Influencing Factors of Technological Innovation Capability of High Tech Enterprises in China". *Finance and Economics Subject Research* 11: 127-132.

Zhu, J.L., Wang, Y.Y., Mei, J.X. 2015. "Differences in Innovation Performance and Influence Factors of High Tech Industry-Based on a Comparative Analysis of the Panel Data of the Semi-Parametric Model". *Systems Engineering Theory and Practice* 35(12): 3135-3143.

Acknowledgement

This project is supported by China Ministry of Education of Humanities and Social Science (Project number: 15YJA790059).