2016

Empirical studies of financial and labor economics

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

EMPIRICAL STUDIES OF FINANCIAL AND LABOR ECONOMICS

by

MENGMENG LI

B.S., Wuhan University, 2008
M.S., Boston University, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2016

Approved by

First Reader

Jianjun Miao, Ph.D.
Professor of Economics
Boston University

Second Reader

Zhongjun Qu, Ph.D.
Associate Professor of Economics
Boston University

Third Reader

Stephen Terry, Ph.D.
Assistant Professor of Economics
Boston University

DEDICATION

To my parents for all their love and support, and for enabling me to obtain the best education possible. I appreciate their sacrifices and would not have been able to reach this achievement without them.

To my patient spouse, Chong.

ACKNOWLEDGMENTS

I am truly indebted to the members of my committee for their support during various stages of writing this dissertation. I would like to extend my deepest appreciation to my main advisor, Prof. Jianjun Miao, for his invaluable advice and guidance throughout my entire graduate education. I am also very grateful to Profs. Zhongjun Qu and Stephen Terry for their constructive comments and suggestions, and for having an open door whenever I needed their help. My personal and intellectual debt to them goes far beyond what I can express here.

In addition, I wish to acknowledge the assistance I received from Profs. Stefania Garetto, Adam Guren, Pierre Perron, Hiroaki Kaido, and other faculty and staff members of the Economics Department at Boston University. I also benefitted from numerous discussions with my fellow graduate students, especially Felipe Cordova.

Finally, yet importantly, I would like to thank my family for supporting me throughout this process. I am also grateful for having the endurance and determination to meet the challenges of the past five years.

EMPIRICAL STUDIES IN FINANCIAL AND LABOR ECONOMICS

MENGMENG LI

Boston University, Graduate School of Arts and Sciences, 2016

Major Professor: Jianjun Miao, Professor of Economics

ABSTRACT

This dissertation consists of three essays in financial and labor economics. It provides empirical evidence for testing the efficient market hypothesis in some financial markets and for analyzing the trends of power couples' concentration in large metropolitan areas.

The first chapter investigates the Bitcoin market's efficiency by examining the correlation between social media information and Bitcoin future returns. First, I extract Twitter sentiment information from the text analysis of more than 130,000 Bitcoin-related tweets. Granger causality tests confirm that market sentiment information affects Bitcoin returns in the short run. Moreover, I find that time series models that incorporate sentiment information better forecast Bitcoin future prices. Based on the predicted prices, I also implement an investment strategy that yields a sizeable return for investors.

The second chapter examines episodes of exuberance and collapse in the Chinese stock market and the second-board market using a series of extended right-tailed augmented Dickey-Fuller tests. The empirical results suggest that multiple "bubbles" occurred in the Chinese stock market, although insufficient evidence is found to claim the same for the second-board market.

The third chapter analyzes the trends of power couples' concentration in large metropolitan areas of the United States between 1940 and 2010. The urbanization of college-educated couples between 1940 and 1990 was primarily due to the growth of dual-career households and the resulting severity of the co-location problem (Costa and Kahn, 2000). However, the concentration of college-educated couples in large metropolitan areas stopped increasing between 1990 and 2010. According to the results of a multinomial logit model and a triple difference-in-difference model, this is because the co-location effect faded away after 1990.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADF.....Augmented Dickey Fuller

API ..... Application Programming Interface

ARIMA .....Auto Regressive Integrated Moving Average

DDP..... Discount Dividend Pricing Model

CFTC..... Commodity Future Trading Commission

EMH.....Efficient Market hypothesis

ES ..... Expert System

GEM..... Growth Enterprise Market

GSADF ..... Generalized Sup Augmented Dickey Fuller

IMF ..... International Monetary Fund

LFP..... Labor Force Participation

JSON..... JavaScript Object Notation

MAPE .....Mean Absolute Percentage Error

NLTK.....Natural Language Toolkit

NYSE .....New York Stock Exchange

RADF.....Rolling Augmented Dickey Fuller

RMSE..... Root Mean Square Error

SADF .....Sup Augmented Dickey Fuller

SSE.....Shanghai Stock Exchange

CHAPTER ONE

Big Data in Testing the Efficient Market Hypothesis of the Bitcoin Market

1 Introduction

Bitcoin was booming. The price of a Bitcoin soared to above $1,000 in December 2013; only 10 months before it was less than $15. This recent price surge, driven by Chinese investors stashing money offshore, looks like a typical bubble.[1] Hoarding behavior by investors also means that Bitcoin is currently more of a speculative asset than a currency. In addition to the huge arbitrage possibilities, some investors are attracted by the potential for anonymous transfers or by the fixed upper limit on the number of bitcoins in circulation. But whatever the reason, Bitcoin is attracting surprisingly numerous fans. It is fascinating to delve more deeply into the Bitcoin phenomenon.

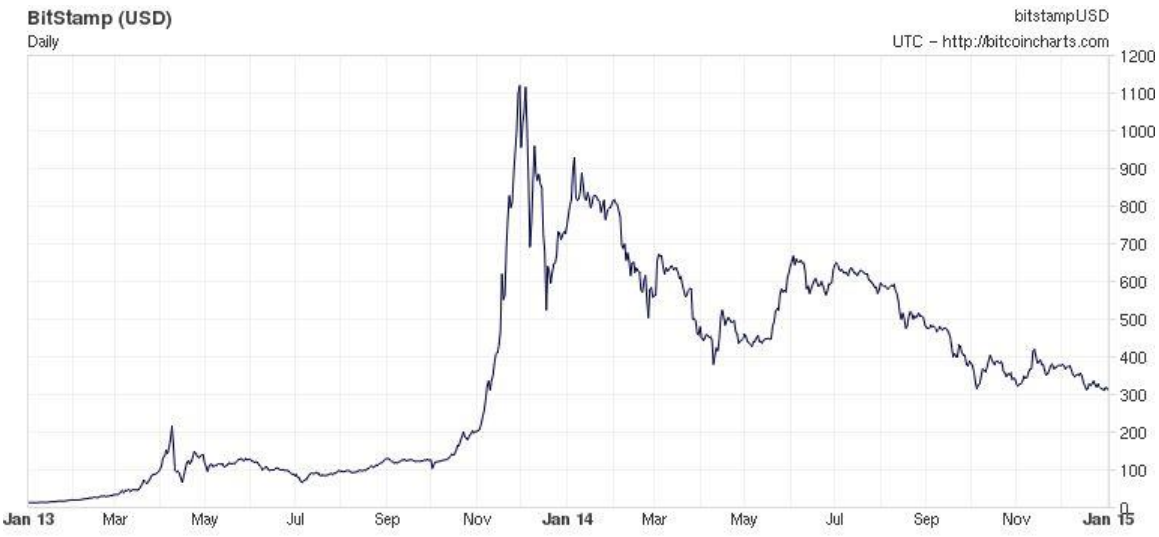

Figure 1.1.1: Historical Bitcoin prices between January 2013 and January 2015

[1] The Bitcoin bubble: http://www.economist.com/news/leaders/21590901-it-looks-overvalued-even-if-digital-currency-crashes-others-will-follow-bitcoin

Figure 1.1.1 presents the historical prices of Bitcoin between January 2013 and January 2015. As indicated by the price swings shown in the figure, Bitcoin lost half its value over the course of 2014 after the huge spikes at the beginning of the year. The frequent fluctuations in prices lead to thought-provoking questions such as whether the Bitcoin market is efficient and what kind of factors impact Bitcoin price movements. If the market is inefficient, then we can assume that irrational investors exist in the presence of arbitrage opportunities. And if we can identify the factors that affect Bitcoin price movements, we could conceivably take advantage of this information to "beat the market."

This chapter proposes that market sentiment information extracted from social media data impacts Bitcoin price movements, further proving the inefficiency of the market. With the help of text data mining methods, I apply textual analysis on social media (Twitter) data to extract Twitter sentiments from more than 130,000 Bitcoin-related tweets. Granger causality tests confirm that Twitter sentiments affect Bitcoin returns in the short run. Furthermore, utilizing the sentiment information, I provide a powerful forecasting model and a portfolio management strategy at the end of the research. This chapter applies the presently known practice of Twitter sentiment analysis to the case of Bitcoin, and may be among the first such applications in this field.

Before discussion of the research procedure, a brief introduction of the Bitcoin is in order. Bitcoin is a digital store of value and payment system invented by Satoshi Nakamato. It is a peer-to-peer electronic cash system and a leading global open-source cryptocurrency (Kroll et al., 2013). It uses cryptography to control the creation and

transfer of money, while transactions are broadcast as digitally signed messages to the shared public network, the "block chain."[2] The reward of solving a block is automatically adjusted so that roughly after every four years of operation of the Bitcoin network, half the amount of bitcoins created in the prior four years is created.[3] Also, the total number of bitcoins in existence will never exceed 21,000,000. Figure 1.1.2 plots the total amount of bitcoins in circulation from 2009 to 2039. As such, we know that by sometime around 2040 the entire supply of bitcoins will have been " mined" and that no new incremental supply will emerge from that point—so there is no way a central bank can inflate their value away by issuing more.

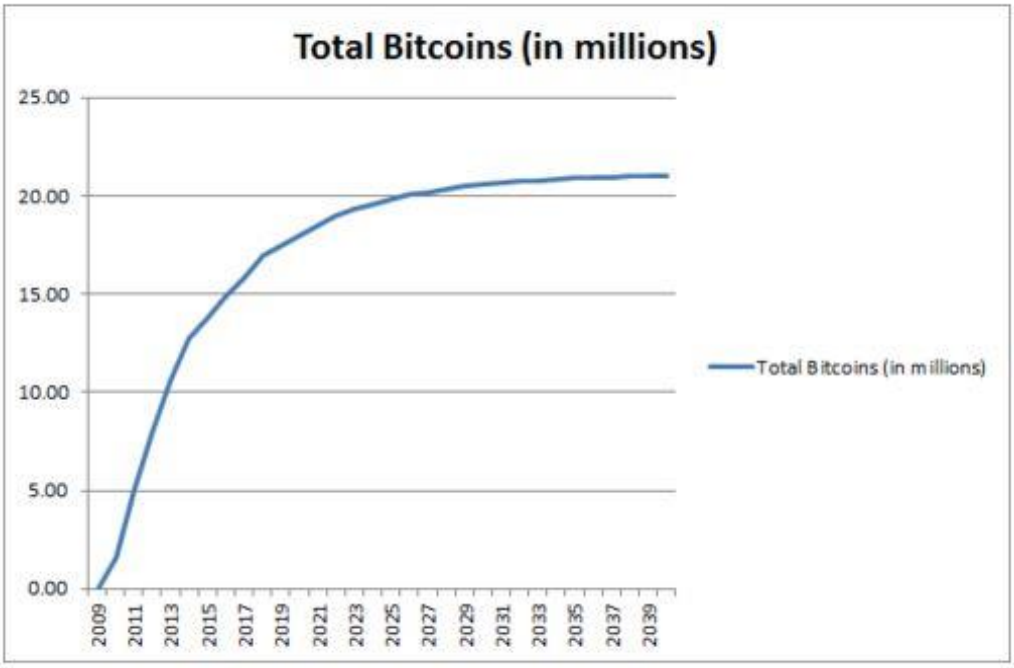

Figure 1.1.2: Total Bitcoins in circulation (compundingmyinterests.com)

[2] See "A Short Introduction to Bitcoin"; https://1btcxe.com/what-are-bitcoins.php
[3] See "Mining Bitcoin": http://ca.newsbtc.com/mining-bitcoin/

Although Bitcoin is a relatively new concept, there are signals of traction in the online trading business and elsewhere, where Bitcoins are increasingly accepted in transactions. A growing number of shops and websites are accepting payments in Bitcoin, and Bitcoin ATM machines even have been installed in Canada and the US since 2013. While the value of all Bitcoins totaled about $3.5 billion in September 2015, the highest market capitalization of Bitcoin reached $13.9 billion (USD) in December 2013. Although this is relatively small when compared with the New York Stock Exchange (NYSE), which had a market capitalization of around $19.7 trillion in September 2015, it is almost equally as active as the NYSE in terms of the daily trading volume. The highest daily trading volume reached $72 million just in the USD currency market, which accounts for less than one-third of the worldwide Bitcoin market. In September 2015, the daily trading volume of the Bitcoin market averaged around $10 million. In comparison, the NYSE's daily trading volume is around $33.5 billion. It is noteworthy that a great many investors have shown an interest in the Bitcoin market, and it is growing steadily. Moreover, on September 17, 2015, Bitcoin was officially designated as a commodity by to the Commodity Futures Trading Commission (CFTC).[4] By this action, the CFTC asserted its authority to provide oversight of the trading of cryptocurrency futures and options, which will now be subject to the agency's regulations. This action will undoubtedly increase the confidence of investors in the safety of trading Bitcoins, and attract a growing number of people into the Bitcoin market.

[4] "Bitcoin Is Officially a Commodity": http://www.bloomberg.com/news/articles/2015-09-17/bitcoin-is-officially-a-commodity-according-to-u-s-regulator

Following this discussion of the general concept of the Bitcoin market, the next step in this chapter starts to focus on testing for the Bitcoin market's efficiency. Section 2 provides a theoretical background of the efficient market hypothesis test, as well as model structures for examining the correlation between Bitcoin returns and market sentiments. If I can successfully prove that market sentiments help us better predict future Bitcoin returns, we can conclude that the Bitcoin market is inefficient. In other words, the sentiment information collected from social media data is a key factor, which makes a significant contribution toward predicting Bitcoin price movements, inasmuch as the extensive use of social media has had pervasive impacts in various fields. Before elaborating how market sentiments impact Bitcoin returns in sections 4 and 5, section 3 discusses the procedures for extracting sentiment information from social media data using text data mining methods.

While research in behavioral economics suggests that emotions can affect individual behavior and decision-making (Akerlof and Shiller, 2009; Scott and Loewenstein, 2008), the sentiment information extracted from social media data has always been considered a key factor for financial markets. Twitter, as one of the top ten most visited sites,[5] has drawn more and more attention from different disciplines as a laboratory to study large sets of social media data. For instance, Jermain et al. (2014) elaborated a method of forecasting the Bitcoin market using signal words found on Twitter. In this chapter, I generate sentiment variables after processing relevant data

[5] "Top Sites". Alexa Internet. Retrieved May 13, 2013

using data mining strategies. Bitcoin prices can be acquired on BitcoinChart.com,[6] while the Twitter data are downloaded from the Internet Archive.[7] Later, a sentiment analysis tool—TextBlob[8]—is adopted to analyze the filtered Twitter data, in order to extract the sentiment information. , The relationship between the derived Bitcoin and sentiment variables is then revealed by econometric method in section 4.

More specifically, in section 4, I make use of Granger Causality analysis to reveal the causal effect of sentiment features on Bitcoin returns. I successfully reject the null hypothesis, i.e., the sentiment does not Granger-cause Bitcoin returns, which also rejects the market's efficiency. Moreover, I explore an Auto Regressive Integrated Moving Average (ARIMA) model with the help of the sentiment features to forecast future Bitcoin prices. Later, in section 5, based on the previous models, I implement a pseudo real time exercise to predict Bitcoin prices, as well as an investment strategy that yields a sizeable return for investors. Lastly, section 6 discusses some extensions and future prospects of this research. In sum, this chapter provides an approach to apply the presently known practice of Twitter sentiment analysis to the case of Bitcoin at a "big data" scale, as well as providing useful econometric methods and practical ways to manipulate the Bitcoin market.

---

[6] www.bitcoinchart.com, data from this website are public free accessible.

[7] Data are downloaded from API, which is the application program interface. Twitter offers two APIs. The REST API allows developers to access core Twitter data and the Search API provides methods for developers to interact with Twitter Search and trends data http://www.webopedia.com/TERM/A/API.html

[8] Textblob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. https://textblob.readthedocs.org/en/dev/

2 Theoretical Background

In financial economics, the efficient-market hypothesis (EMH) states that it is impossible to "beat the market" because market efficiency causes existing share prices to always incorporate and reflect all relevant information.[9] According to the EMH, financial assets always trade at fair value, making it impossible for investors to either purchase undervalued assets or sell assets for inflated prices. Therefore, it should be impossible to outperform the overall market by advanced portfolio selection, and the only way to make higher profits is through taking greater trading risks.

There are normally three versions of EMH: weak, semi-strong, and strong. The strong version of EMH states that prices instantly reflect all past publicly available information, as well as hidden information. The semi-strong form of EMH claims both that prices could reflect all public information and that prices instantly change to reflect new public information. Meanwhile, the weak form of EMH states that in an efficient market the prices of traded assets (e.g., stocks, bonds, or property) already reflect all past publicly available information. In other words, in the efficient market no information in the past should contribute to predicting the future returns.

In this chapter, I focus on testing using the weak version of the EMH. Significant research has been done on the efficient market hypothesis. As noted above, under the weak form of the efficient market hypothesis, stock prices are able to reflect all the past public information (Fama, 1970). Moreover, when facing a wide variety of publicly available information, based on rational expectation theory, people will make full use of

[9] In financial economics, EMH states that asset prices fully reflect all available information https://en.wikipedia.org/wiki/Efficient-market_hypothesis

the past available information to generate expectations regarding the volatility of stock prices (Muth, 1961).. Furthermore, the aggregate expectations, namely, the market expectation, have been demonstrated to be embodied in stock prices (Lane and Jacobson, 1995). Excess returns cannot be earned in the long run by using investment strategies based on historical share prices or other historical data. That implies that future price movements are determined entirely by information.

Researchers and investors have criticized the efficient market hypothesis both empirically and theoretically. Grossman and Stiglitz (1980) argued that prices cannot perfectly reflect the publicly available information because the information is costly. If it did, those who spent resources to obtain it would receive no compensation, leading to the conclusion that it is impossible for the market to be efficient. Even in theory, if there are costs of gathering and processing information, abnormal returns will exist. Those returns are necessary to compensate investors for their information collection and processing expense, and then they are no longer abnormal. Alternatively, Lo and MacKinlay (1999) argued that the degree of market inefficiency determines the effort investors are willing to expend to gather and trade on information. Henceforth, a non-degenerate market equilibrium will arise only when there are sufficient profit opportunities, i.e., inefficiencies, to compensate investors for the costs of trading and information collection.

Based on all the criticism above, we need to figure out how to test for the efficient market hypothesis formally. The following is a basic formula stating the weak form of EMH: in an efficient market no information in the past should contribute to predicting

future returns. Here y_t represents the assets' returns at time t. The weak version of EMH is stated as:

$$E(y_t|y_{t-1}, y_{t-2}, \dots) = E\ (y_t) \quad (1.2.1)$$

Therefore, a simple model to test the EMH is to have the null hypothesis as: H_0: $\beta_1 = 0$

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \quad (1.2.2)$$

If we can successfully test that $\beta_1$is significantly different than 0, we can reject the null hypothesis, which leads to the conclusion that the market is inefficient.

Once we have successfully tested that the market is inefficient, naturally the next step is to think about how to realize the arbitrage opportunities afforded by the inefficient market. It has been proven by Bollen, Mao, and Zeng (2011) that prediction of the Dow Jones Industrial Average can be made more accurate by taking the mood of Twitter into consideration. Outside of financial markets, the possible impacts of social media data have been studied in multiple areas, such as forecasting movie box-office grosses (Asur and Huberman, 2010), tie strength among friends (Gilbert and Karahalios, 2009), and other areas. Obviously, we can see that social media information possesses a certain power to impact several fields, including the stock market. Therefore, it comes very naturally to think about the following question:

"Does social media information impact Bitcoin returns?"

If we try to formulize the above question, we need to test the following equation:

$$E(y_t|Z_{t-1} = G) \neq E\ (y_t|y_{t-1}, y_{t-2}, \dots) \quad (1.2.3)$$

The left-hand side represents the predicted Bitcoin returns conditional on past values and social media information, while G includes all previous values of Bitcoin, as well as

sentiment information. The right-hand side represents the predicted Bitcoin returns conditional only on past values. If it can be proven that the left-hand side is statistically significantly different from the right-hand side, then we can conclude that the sentiment information contributes to predicting future Bitcoin returns.

The underlying idea of the above test assumes that social media information impacts the expectation of the asset values. According to a study by Lemmon and Portniaguina (2006), sentiment from social media can be used to measure expectation. This is also verified in the research of Souleles (2004). Thus, if we can extract sentiment from social media data, it may be feasible to reveal the impact of social media sentiment on asset returns. From a psychological perspective, according to the theory of planned behavior proposed by Ajzen (1991), people's behaviors are affected by their perceived behavioral control, subjective norms, and attitudes toward the behaviors. The behavior could be either buying or selling the asset; therefore, it will affect the price movements.

In light of the above discussion, the following logical sequence can be constructed. A wealth of social media information is available that drives investors' expectations, which help to further shape their attitudes. These attitudes can be parsed by people's posts on social media, and sentiment analysis is the tool that can be used to expose these attitudes (O'Connor, Balasubramanyan, Routledge, and Smith, 2010; Liu and Zhang, 2012). Meanwhile, perceived behavior controls depend on the perceived degree of difficulty of their Bitcoin trading, while subjective norms depend on the perception of other people's opinions of their trading behaviors. The intention is influenced by these three elements (attitudes, behavior controls and subjective norms) in

aggregate, and intentions precipitate investors' behaviors, either to buy or sell Bitcoin. Furthermore, cumulative trading behaviors affect asset prices via the principle of supply and demand. The entire process explains why sentiment on social media can impact Bitcoin price movements.

Some scholars claim that any test of the efficient market hypothesis is a joint test of an equilibrium-returns model and rational expectations. According to Fama (1970), an efficient market will always "fully reflect" available information, but in order to determine how the market should "fully reflect" this information, we need to determine investors' risk preferences. Therefore any test of the EMH is a test of both market efficiency and investors' risk preferences. For this reason, the EMH, by itself, is not a well-defined and empirically refutable hypothesis. However, I do not base any assumption on rational expectations of the Bitcoin market. Instead, I validate that the sentiment information helps us to better predict the market than if only using market instruments (e.g., daily price, trading volume, etc.). Since the market price itself should contain all past information according to the EMH, I can therefore claim that the Bitcoin market is "inefficient"[10].

[10] In this chapter, claiming the market is "inefficient" only means some information could help us predict future asset returns. It doesn't prove inefficiency theoretically.

3 Data Collection & Sentiment Analysis

3.1 Social Media Data

With nearly one-fourth of the entire global population using social media, its impacts on different aspects of society have become more and more prominent. The first modern social media platform, Open Diary, was released in 1998 (Kaplan and Haenlein, 2010). According to their study, social media is an Internet-based application built on Web 2.0 in which users are allowed to create, exchange, and maintain information. Social media is pervasive in modern society, including in such areas as politics, housing markets, and financial markets. As a consequence, most people have already gained some understanding of its importance. By 2012 (Cai et al., 2012), it was projected that the population on social media would grow to around 1.3 billion. In order to explore the impacts of social media on the Bitcoin market, the selection of a particular object for experimentation should possess representative characteristics. I chose to utilize Twitter to carry out the concrete study.

Twitter is an online social networking service that enables users to send and read short (140 character) messages called “tweets.” It launched in July 2006 in the United States and has gained tremendous attention and attracted a massive user base (Kwak et al., 2010). Twitter rapidly gained worldwide popularity and now has more than 500 million users[11], who in 2015 posted around 500 million tweets per day. In 2013, Twitter was one of the ten most-visited websites, and has been described as “the SMS of the

[11] "Twitter MAU Were 302M For Q1, Up 18% YoY - Twitter (NYSE:TWTR) | Benzinga]". April 28, 2015. Retrieved May 2, 2015.

Internet" (Alexa Internet, 2013).[12]. I chose to study Twitter due to its data's public availability. Twitter data can be downloaded from the stream Application Programming Interface (API), which offers good services for researchers and developers to conduct experiments. I will explain the API process in detail below.

Figure 1.3.1 is an example of a tweet that mentions Bitcoin. The character hashtag "#" is assigned in the text message to signify a certain topic by attaching a topic name behind. And the sign "@" mentions a specific user account that is followed by the account name. All tweets are accessible for both registered and non-registered users. But only the registered user can re-post (retweet) the existing tweet. I downloaded the Twitter data from Internet Archive,[13] which is an online non-profit digital library. It provides free access to archived Twitter data generated since January 2012. The Twitter data on Internet Archive were collected through the Stream API, and is in JSON (JavaScript Object Notation) format. Figure 1.3.2 shows an example of a JSON file's structure. It not only contains the tweet text, but also plenty of other useful information such as the retweet count, geographic information, reply status, and the user's personal information.

[12] "Top Sites". Alexa Internet. Retrieved May 13, 2013
[13] Internet archive website: https://archive.org/index.php

Figure 1.3.1: A tweet with "#," "@," and picture from the Twitter account of Gift Off and retweeted by Twitter account Bitcoin

Figure 1.3.2: JSON file structure of one tweet (Cao, 2014)[14]

[14] Cao(2014)'s Figure 3.2: Data structure of single tweet json data.
http://dare.uva.nl/cgi/arno/show.cgi?fid=544733

In this chapter, I use a simple collection of the API stream, also known as the "Spritzer" version.[15] It is a light and shallow record (stream) of Twitter grabs, designed mainly aiming for the purposes of research, history, testing, and memory. The Twitter Spritzer provides a 1% random sample of public tweets. For each day we have an average of around 4 million tweets after the random selection, which is large enough for representing the aggregate of tweet data. By using this Spritzer version, around 952 million tweets have been collected within the timeframe from February 1, 2014 to September 30, 2014, which is 238 days[16] in total. The data were downloaded from the website: https://archive.org/details/twitterstream, and data size is around 0.5 TB for each month. Given the huge data set that exists even for a single day, filtering the relevant data becomes the top priority. I discuss the procedures of data filtering using the single tweet data example in JSON format presented in Figure 1.3.3.

In figure 1.3.3, the text following the highlighted word "text" is the content of this tweet, which is "First US bitcoin ATMs to open soon in Seattle, Austin http://t.co/D3AxMNMgKB." In the first line, ""lang': 'en'" means this tweet was written in English. At this moment, most natural language processing mechanisms for the sentiment analysis of texts are only well developed for some languages such as English, and English is the dominant language on Twitter (Fox, 2013). Therefore, the first step is to filter out non-text tweets and non-English tweets. The second step is to filter out the

[15] Spritzer stream is sampled stream. https://blog.gnip.com/tag/spritzer/
[16] The whole sample should cover 242 days, but there are 4 days have missing data, so I deleted those 4 days.

non-Bitcoin-relevant tweets through a toolkit called Google Keyword Planner,[17] which is designed to offer critical keywords associated with specific issues. This free toolkit is a product of Google Adwords that provides data on search queries in Google and other resources for planning a specific advertising campaign. Through the ranking of search frequencies, keywords about Bitcoin's popular products, service, and online exchange platform were selected. Figure 1.3.4 shows the keyword set for filtering the Twitter data.


```
{"geo": null, "id_str": "435671944688463872", "filter_level": "medium", "lang": "en", "entities":
{"urls": [{"expanded_url": "http://www.ndtv.com/article/world/first-us-bitcoin-atms-to-open-soon-in-
seattle-austin-484860", "indices": [54, 76], "url": "http://t.co/D3AxMNMgKB", "display_url":
"ndtv.com/article/world/\u2026"}], "hashtags": [], "user_mentions": [], "symbols": []},
"retweet_count": 0, "user": {"profile_link_color": "0084B4", "protected": false, "id_str": "37034483",
"favourites_count": 7, "listed_count": 8037, "screen_name": "ndtv", "lang": "en", "default_profile":
false, "profile_background_image_url_https":
"https://pbs.twimg.com/profile_background_images/662611415/inwd4bauu5m1vp5wc18r.png",
"profile_background_color": "C0DEED", "profile_image_url":
"http://pbs.twimg.com/profile_images/2341726876/ndtv240_240_normal.png", "utc_offset": 19800,
"verified": true, "followers_count": 1679745, "created_at": "Fri May 01 20:34:48 +0000 2009",
"is_translation_enabled": false, "following": null, "profile_sidebar_border_color": "FFFFFF",
"time_zone": "New Delhi", "profile_background_image_url":
"http://pbs.twimg.com/profile_background_images/662611415/inwd4bauu5m1vp5wc18r.png", "url":
"http://www.ndtv.com/", "profile_image_url_https":
"https://pbs.twimg.com/profile_images/2341726876/ndtv240_240_normal.png", "profile_text_color":
"333333", "description": "Breaking news alerts from India", "profile_sidebar_fill_color": "DDEEF6",
"contributors_enabled": false, "follow_request_sent": null, "geo_enabled": false, "notifications":
null, "default_profile_image": false, "profile_banner_url":
"https://pbs.twimg.com/profile_banners/37034483/1347983190", "location": "India", "statuses_count":
120653, "profile_use_background_image": true, "is_translator": false, "profile_background_tile": false,
"id": 37034483, "name": "NDTV", "friends_count": 23}, "retweeted": false, "in_reply_to_user_id": null,
"truncated": false, "in_reply_to_screen_name": null, "favorited": false, "source": "<a href=
\"http://twitter.com/tweetbutton\" rel=\"nofollow\">Tweet Button</a>", "place": null,
"in_reply_to_status_id": null, "in_reply_to_user_id_str": null, "possibly_sensitive": false, "text":
"First US bitcoin ATMs to open soon in Seattle, Austin http://t.co/D3AxMNMgKB", "created_at": "Tue Feb
18 07:07:40 +0000 2014", "coordinates": null, "favorite_count": 0, "id": 435671944688463872,
"in_reply_to_status_id_str": null, "contributors": null}
```


Figure 1.3.3: Single tweet data example in JSON format

```
bitcoin_keywords_set=set(["bitcoin","bitcoins","bitcoins'","#bitcoin","bit coin","bit coins","bitcoin's",
                          "bitmines","bitmine","bitmining","mtgox","mt.gox","bitstamp","bitcoin.org","blockchain",
                          "bitpay","cryptocurrency","btcvert.com","btcvert","coindesk","bitcoind","coinbase","coinbase's",
                          "blockchain.info","bitcoinexpo","bitcoin-qt", "bitinstant","bitspark","cointellect", "bitpesa",
                          "coinfire","bitshare","bitshares","dogecoin","litecoin","bitfinex","bitx","bitcurex","bittrex",
                          "c-cex","btce","btc-e","campbx","coinjar","cyptonit","cryptsy","fyb-se","btcinstant","virtex"])
```

Figure 1.3.4: The keyword set for Bitcoin

[17] Google keyword planner website: https://adwords.google.com/KeywordPlanner

After the two steps of filtering, the number of Bitcoin-related tweets is around 500 per day. The next section discusses how to apply text analysis to those filtered tweets.

3.2 Sentimental Analysis

Text data mining, roughly equivalent to text analytics, refers to the process of deriving high quality information from text data. A typical text mining[18] task includes text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling. Sentiment analysis refers to the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information in source materials. In this chapter, I have applied sentiment analysis on text messages of tweets, in order to derive the sentiment behind the text. To be more specific, the input is simple text while the output is a definite sentiment measurement, or more precisely, a numeric value.

Among various sentiment analysis tools, I selected TextBlob[19] to perform the text-mining job, because of its ease of implementation and powerful features. TextBlob is a library written in Python, designed for natural language processing and distributed for free. It consists of two leading and featured natural language processing tools—Pattern and Natural Language Toolkit (NLTK)—for calculating numerical sentiment. The default

[18] Text mining: https://en.wikipedia.org/wiki/Text_mining

[19] Textblob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. https://textblob.readthedocs.org/en/dev/

analyzer is Pattern, which I include in section 6 for a robustness check. In this section, I adopt the NLTK sentiment analyzer, which has a float number output from 0 to 1. The larger the value, the more positive the sentiment is in the text. Any tweet has an output score bigger than 0.5 is defined as a positive tweet, whereas any tweet that has an output score of less than 0.5 is defined as a negative tweet; any tweet has an output score equal to 0.5 is defined as a neutral tweet. For instance, if I input the tweet example in Figure 1.3.3, which says: "First US bitcoin ATMs to open soon in Seattle, Austin," with the help of NLTK, the result was presented to be 0.48 (Figure 1.3.5). In other words, this is a negative tweet, although very close to neutral.

```
>>> from textblob import TextBlob
>>> from textblob.sentiments import NaiveBayesAnalyzer
>>> tweet=TextBlob("First US bitcoin ATMs to open soon in Seattle, Austin",analyzer=NaiveBayesAnalyzer())
>>> tweet.sentiment.p_pos
0.48409153888813666
```

Figure 1.3.5: Tweet example processed by TextBlob in Python

Although it is easy to utilize, the reliability of this NLTK analyzer is guaranteed by a sophisticated process and well-documented mining procedures. The Natural Language Toolkit is a Python library suite, which supports classifying text messages into various categories by utilizing a Naive Bayes classifier (Bird, Klein, and Loper, 2009). As a leading tool for sentiment analysis, NLTK has been widely adopted within a large number of research projects (Bird, 2006; Bird et al., 2009). NLTK offers a large data set with a definite positive or negative tag on each sentence, which is also called the training set, to train the Naive Bayes classifier. I adopted the movie reviews corpus, which is the

most popular training set (Pang and Lee, 2005). The accuracy, based on the movie reviews corpus, is around 81% (Bird et al., 2009). During the training, any word in the original sentences containing less than two characters was removed at first due to typical meaninglessness of these kinds of words. Second, a list of all different words, also known as a features list, was generated, in which words were reordered based on frequency.

Next, I used the feature list to extract the input text words that intersected with this features list. The resulting set is called the feature set, which is associated with the sentiment tag from the training set. In other words, the training set is utilized to instruct the classifier, telling the classifier which word is more likely to be associated with which tag. In this way, if more words are found to be linked with the positive tag, the input will be associated with the positive tag, and vice versa. Since the classifier was trained by a large training set, which includes an incredibly large possibility number of records, it becomes intelligent enough to recognize new input with known tags automatically. Henceforth, any tweet can be classified into positive, negative, or neutral categories.

After the sentiment analysis, each tweet was classified into a positive, negative, or neutral category. Next, I constructed five variables for the sentiment information. First, I have N_t^{Pos} and N_t^{Neg} represent the total number of positive or negative tweets on a particular date t. Second, I have carried forward the work of Antweiler, Copeland, and Taylor (2001) for defining bullishness(B_t) for each day given as:

$$B_t = \ln\left(\frac{1+N_t^{Pos}}{1+N_t^{Neg}}\right) \quad (1.3.1)$$

Bullishness is also referred to as the sentiment index in many relevant studies. I include

"1+" in equation 1.3.1 to avoid the possibility that any N_t^{Pos} or N_t^{Neg} might be zero. The logarithm of bullishness measures the share of surplus positive signals and also gives more weight to larger numbers of messages communicating a specific sentiment.

Next, I define message volume for a time interval t simply as a natural logarithm of the total number of tweets for a specific stock/index, which is $ln(N_t^{Pos} + N_t^{Neg})$ (Rao and Srivastava, 2014). In the end, the agreement between positive and negative tweets is defined as:

$$A_t = 1 - \sqrt{1 - \frac{N_t^{Pos} - N_t^{Neg}}{N_t^{Pos} + N_t^{Neg}}} \tag{1.3.2}$$

For example, if all tweets on day t about Bitcoin are positive or negative, the agreement would be 1, meaning that the social opinions are in absolute agreement. I did not take neutral tweets into account, since the neutral tweets only account for less than 1% of the total tweets data per day. Besides the above five variables, I also have carried[20] terminologies for all those five features (Positive, Negative, Bullishness, Message Volume, Agreement) remain the same for each day with a lag of one day. For instance, carried Positive for day t is written as $Positive_{t-1}$. I listed the correlation matrix of both sentiment variables and carried sentiment variables in Appendix A.1.

[20] Here carried terminology means the lag variable of each feature. In other words, we are interested in the impacts of these features from previous day.

3.3 Bitcoin Data

I collected Bitcoin market instruments such as daily weighted price and daily trading volumes to examine their relationships with market sentiments extracted from tweets. The daily trading volumes and the daily weighted Bitcoin prices were collected from Bitcoin Charts (www.bitcoincharts.com). And I also downloaded opening (O_t) and closing (C_t) price, as well as highest (H_t) and lowest (L_t) price of Bitcoin between February 1, 2014 and September 30, 2014. Returns are calculated as the difference of the logarithm to the base e between the closing values of the stock price of a particular day and the previous day.

$$R_t = \left\{ln\, weighted\_price_{(t)} - ln\, Weighted\_price_{(t-1)}\right\} * 100 \qquad (1.3.3)$$

And I estimated the daily volatility based on intra-day highs and lows using Garman and Klass (1980)[21] volatility measures given by the formula:

$$\sigma = \sqrt{\frac{1}{2} * \frac{1}{T}\Sigma[ln\frac{H_t}{L_t}]^2 - (2ln2 - 1)[ln\frac{C_t}{O_t}]^2} \qquad (1.3.4)$$

Here T stands for the whole time duration. The Garman and Klass volatility estimator was created in late 1980, and is an extension of the Parkinson (1980) estimator, which includes opening and closing prices. As overnight jumps are ignored, the Garman and Klass measure underestimates volatility, and is known as the best analytic scale-invariant estimator.

There are many online trading platforms for Bitcoin all over the world. I denote each online trading platform as a "Bitcoin market." Figure 1.3.6 shows the Bitcoin

[21] Garman and Klass (1980) and Parkinson (1980) estimators belongs to the same type of volatility estimator, which makes use of information like opening, closing, high and low prices of each day.

market volume distribution worldwide in August 2015. The left pie chart shows the volume distribution by market, and the right pie chart shows the volume distribution by currency. We can tell from the pie chart that China accounts for over 60% of the entire market, whereas the U.S. only takes around 20%.

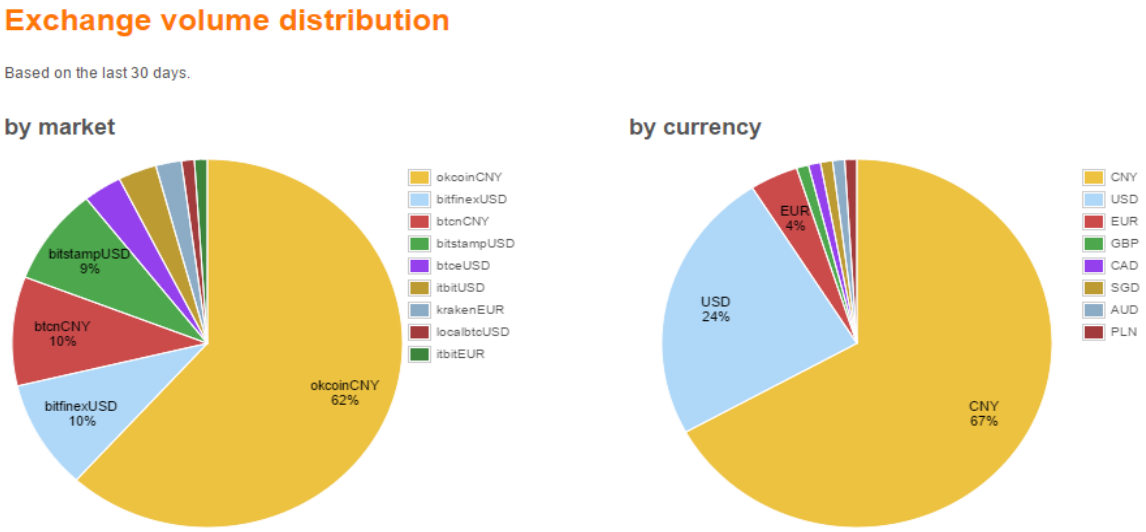

Figure 1.3.6: Market volume distribution in August 2015 (bitcoincharts.com)

Further, market data from the top five leading Bitcoin markets were obtained from bitcoinchart.com, namely BitStamp, Bitfinex, BTC-e, btcnCNY, and okcoinCNY. For each market, I collected open price, closing price, high price (intraday), low price (intraday), volume by bitcoin, volume by currency, and the weighted daily price. Figure 1.3.7 plots the volume distribution of the top five Bitcoin markets by USD from January 2014 to January 2015. Three of them (BitsStamp, Bitfinex, and BTC-e) are from the US and other two (btcnCNY and okcoinCNY) are from China. Next, we need to think carefully about choosing the appropriate market data for this study. It comes naturally to

three choices: (1) the largest market in the US (BTC-e); (2) the weighted composite market of the three largest markets in the US; and (3) the weighted composite market of the five largest markets in the US. Figures 1.3.8, 1.3.9, and 1.3.10 depict the data collection results: daily weighted prices, daily returns, and daily volatilities for the above three choices. We can tell from the figures that the movements of the daily weighted prices, returns, and volatilities were quite similar for each choice. The volatilities vary in figure 1.3.10 due to the fact that the volatility measure includes both the high and low price of the day. For the composite market, the intra-day difference is higher than a single market, leading to larger volatility for the composite US and whole market.

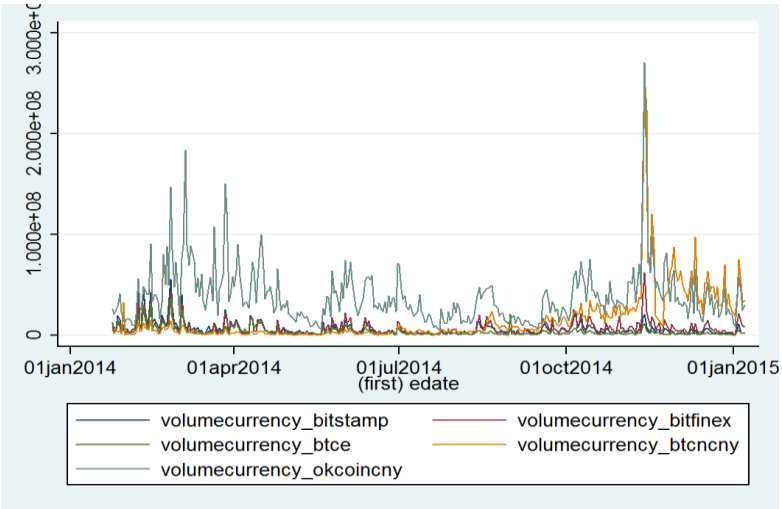

Figure 1.3.7: Volume distributions of top five markets in U.S.

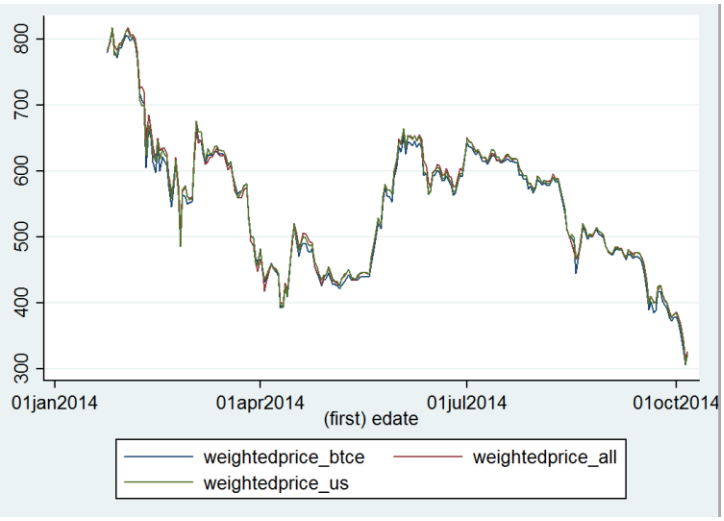

Figure 1.3.8: Daily weighted prices of the three markets all over the world

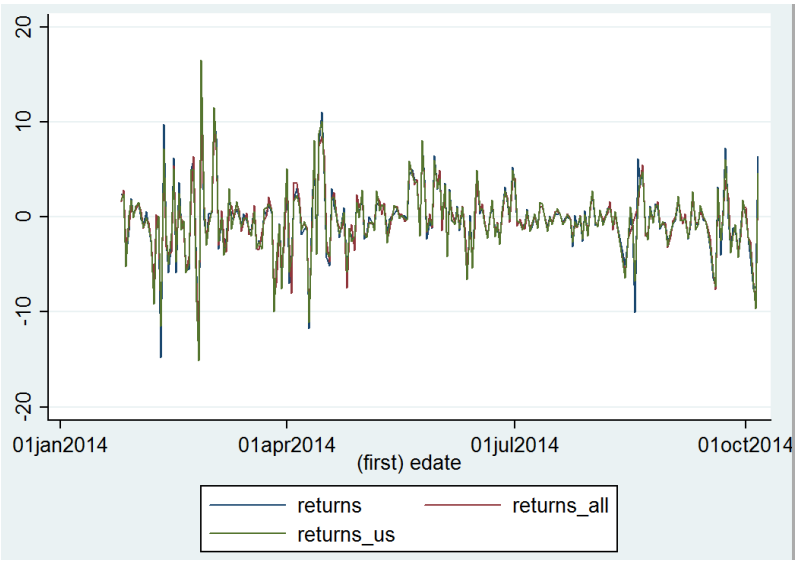

Figure 1.3.9: Daily returns of the three markets all over the world

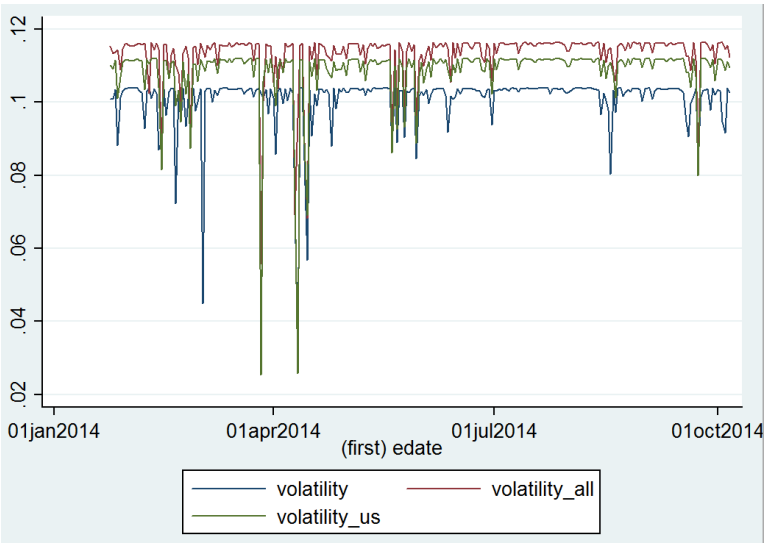

Figure 1.3.10: Daily volatilities of the three markets all over the world

Additionally in Appendix A.1, I present the correlation matrix for the Bitcoin market variables and the sentiment variables for each market in Tables A.1.1, A.1.2, and A.1.3. The correlations between those three Bitcoin markets and the sentiment features show almost identical behavior in terms of magnitude and sign. Therefore, I have chosen the composite US Bitcoin market for the research that follows.

4 Models

4.1 Bivariate Granger Causality Analysis

The previous section has shown the correlations between Bitcoin market parameters and sentiment variables. However, it does not answer the question of whether sentiment information affects the Bitcoin returns. And we also need to investigate the efficiency of the Bitcoin market. In order to achieve these research goals, I applied a

Granger Causality model to the time series averaged to daily windows to Bitcoin returns with the sentiment information (positive, negative, bullishness, tweet volume, and agreement).

Granger Causality is a statistical concept of causality that is based on prediction. It is not used to establish causality, but as an econometric tool to investigate a statistical pattern of lagged correlation. Granger Causality analysis rests on the assumption that if a variable X causes Y, then changes in X will systematically occur before changes in Y. In other words, Granger Causality tests whether X predicts Y, rather than whether X causes Y. Formally, this tests whether one time series is significant in predicting the other time series. Let R_t denote the returns of Bitcoin at time t, and X_t denotes the sentiment variables. To test the impacts of sentiment information on changes in returns, I perform the Granger Causality analysis in equation (1.4.1) and equation (1.4.2):

$$R_t = \alpha + \sum_{i=1}^{J} \beta_i R_{t-i} + \epsilon_t \quad (1.4.1)$$

$$R_t = \alpha + \sum_{i=1}^{J} \beta_i R_{t-i} + \sum_{i=1}^{J} \gamma_i X_{t-i} + \varepsilon_t \quad (1.4.2)$$

J here represents the lags for prediction. I choose a week-long window, setting J equal to 7. In other words, I am interested in the impact of sentiment information for the previous 7 days on predicting future Bitcoin returns.

If the variance of ε_t is reduced by the inclusion of the X_t terms in the second equation, then it is said that X_t Granger-causes R_t. In other words, X_t Granger-causes R_t if the coefficients in γ_i are jointly significantly different from zero. This can be tested by performing an F-test of null hypothesis $H_0: \gamma_i = 0$, i.e., the sentiment information does not Granger-cause Bitcoin returns, given assumptions of covariance stationary on R_t and

X_t. Based on the F-statistics, if γ_i is significantly different than 0, then we could conclude that X_t has significant impacts on predicting R_t.

Table 1.1: Granger Causality Analysis of Sentiment Information and Bitcoin Returns

Returns	lag	N_pos	N_neg (Prob > F=0.0308)	Bull	Agree	M_vol
	1	-0.0016	0.0139*	-1.3405*	-5.3718*	-0.8588
	2	-0.0040	-0.0140	0.8977	3.5167	-0.7598
	3	0.0032	0.0103	-0.3654	-1.5432	0.2634
	4	0.0003	0.0083	-1.3913	-5.5174*	0.8483
	5	-0.0016	-0.0106	0.3281	1.3996	-0.2866
	6	0.0079*	0.0222*	-0.2186	-1.1990	1.7238
	7	-0.0001	0.0029	-0.9939	-3.7421	0.4686

Note: ** for p-value < 0:05 and * for p-value < 0:1 which is 95% and 99% confidence interval respectively. And according to the F-statistic, only N_neg rejects the null hypothesis.

Table 1.1 presents the Granger Causality results of sentiment information and Bitcoin daily returns. According to the F-statistic, N_neg (total counts of negative tweets per day) rejects the null hypothesis. That is to say, N_neg makes a significant contribution in predicting future Bitcoin returns. However, the coefficient of N_neg (J=1/6) is negative, which means that the sentiment information of 1 or 6 days ago could predict positive returns at current time t. It is interesting to consider why negative tweets in the past could produce positive returns in the future. In order to explain that, I have also performed Granger Causality analysis of sentiment information and Bitcoin daily prices, volatilities, and trading volumes, further investigating how the market instruments change over time under the sentiment impacts.

Table 1.2: Granger Causality Analysis of Sentiment Information and Bitcoin Prices

Weighted price	Lag	N_pos	N_neg (Prob >F=0.0063)	Bull	Agree	M_vol
	1	-0.0105	0.0774*	-8.2536*	-32.8765*	-4.3082
	2	-0.0178	-0.0740	4.7827	19.1173	-2.7457
	3	0.0178	0.0525	-2.4782	-10.0863	1.4778
	4	0.0065	0.0537	-7.3763	-29.2524	5.8642
	5	-0.0083	-0.0591	2.0000	8.7205	-1.9495
	6	0.0507*	0.1359**	-1.5917	-7.9728	11.3895*
	7	0.0048	0.0148	-4.7953	-17.9093	3.3410

Note: ** for p-value < 0:05 and * for p-value < 0:1 which is 95% and 99% confidence interval respectively. And according to the F-statistic, only N_neg rejects the null hypothesis.

Table 1.3: Granger Causality Analysis of Sentiment Information and Bitcoin Trading Volatilities

Volatility	Lag	N_pos	N_neg (Prob > F = 0.0000)	Bull (Prob > F = 0.0238)	Agree (Prob > F = 0.0170)	M_vol (Prob > F = 0.0241)
	1	-0.00001	-0.00003*	0.00380*	0.01501*	-0.00400*
	2	-0.00001	-.00006**	0.00171	0.00813	-0.00402
	3	0.00000	0.00002	0.00055	0.00189	0.00077
	4	0.00000	-0.00003	0.00296	0.01139	-0.00146
	5	0.00001	0.00008***	-0.0043*	-0.01650*	0.00630*
	6	-0.00001	-0.00005*	0.00239	0.00892	-0.00403
	7	0.00000	-0.00001	0.00092	0.00428	-0.00141

Note: ** for p-value < 0:05 and * for p-value < 0:1 which is 95% and 99% confidence interval respectively. And according to the F-statistic, N_neg, Bull, Agree and M_vol reject the null hypothesis.

Tables 1.2, 1.3, and 1.4 present the Granger Causality results of sentiment information and three other Bitcoin market instruments (prices, volatilities, and trading volumes). Based on the F-statistics, N_neg has a significant contribution in predicting

Bitcoin future price and volatility. Meanwhile, bullishness, agreement, and message volume also make significant contributions to the prediction of volatility. Additionally, the results show that N_neg also has positive effects on future prices, which indicates that the negative tweets of 6 days ago result in an increment to the present price. Connected with the positive coefficient of N_neg on trading volumes, we can outline a scenario whereby negative tweets result in positive returns. If you receive negative tweets today, then you are more likely to purchase Bitcoin since you would expect the Bitcoin price to decrease (be relatively inexpensive). Six days later, the Bitcoin price goes up after many investors have bought in, leading to positive returns.

Table 1.4: Granger Causality Analysis of Sentiment Information and Bitcoin Trading Volumes

Trading volume	Lag	N_pos	N_neg	Bull	Agree	M_vol
	1	13.2422	65.5042	-7827.7440	-30565.8400	5281.7630
	2	-72.2000*	-72.9378	1113.9530	4123.7900	-16038.77*
	3	10.7307	84.1973	-11999.73*	-47638.12*	-257.0420
	4	-34.6974	-74.2753	-249.9330	410.9557	-6412.0160
	5	-14.4971	-56.7943	6454.8890	24707.0300	-6013.4070
	6	36.0934	119.0422*	-7132.1770	-26760.5600	9364.2530
	7	17.8018	15.5216	-160.7984	-1758.5700	10096.2800

Note: ** for p-value < 0:05 and * for p-value < 0:1 which is 95% and 99% confidence interval respectively. And according to the F-statistic, none sentiment variable rejects the null hypothesis.

4.2 ARIMA Forecasting Models

In section 4.1, I successfully demonstrated that sentiment information makes significant contributions in predicting future Bitcoin returns. The present section makes use of an expert system (ES) to show how to improve the forecasting model with the Twitter sentiments.

ES is a software that incorporates specialists' knowledge in a certain domain, and is designed to solve complex problems by reasoning about knowledge. Expert systems in data mining are widely applied, and incorporate a set of competing methods such as Exponential Smoothing, Auto Regressive Integrated Moving Average (ARIMA) models, and seasonal ARIMA models. These models are popular in financial modeling to predict the values of stocks, bonds, commodities, etc. (McCormick, 1969; Edward, 1990). In this section, I make use of an ARIMA model to investigate the sentiment information's predictive power.

An ARIMA model is, in theory and practice, the most general class of models for forecasting time series data, which is subsequently stationarized by a series of transformations. For a non-seasonal ARIMA (P,D,Q) model, P specifies the seasonal autoregressive order, D is the seasonal differencing order, and Q is the moving average order. Formally, it can be presented as following equation:

$$\left(1 - \sum_{i=1}^{P} \emptyset_i L^i\right)\Delta^d Y_t = \left(1 + \sum_{i=1}^{Q} \theta_i L^i\right)\varepsilon_t \qquad (1.4.3)$$

whereY_t is the financial time series, representing the Bitcoin price at time t in this study. In an expert system, it automatically selects the most significant predictors among all

others that are available when one runs the ARIMA model on time series Y_t, i.e., choosing the most fitted p, d, q order for certain time series.[22]

To analyze the performance of sentiment information in forecasting models, I apply the ARIMA forecasting model twice: first with sentiment information as independent variables, and a second time without them. This method provides us with a quantitative comparison of improvement in forecasting using sentiment information. Each time I apply the ARIMA model, I utilize a two-step procedure: (1) For the whole sample I have a time series for a total 242 days, out of which I use approximately 75%(i.e., the first 182 days) for training both models with and without the sentiment information. (2) Then I run the exact same, but trained, model to forecast over the remaining 25% sample, predicting the Bitcoin price of the rest of the sample. Then I compare the forecasting accuracy in the testing period for both models. In the comparison, the following four features are selection criteria for the forecasting model:

(1) RMSE (root-mean-square error), which is a frequently used measure of the difference between values predicted by a model or an estimator and the values actually observed.

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}} \tag{1.4.4}$$

where $\hat{y}_i$ is the predicted value of Bitcoin price y_i.

(2) Coefficient of determination (R-square), which is square of the value of the Pearson 'r' of fit values (from the ARIMA model) and actual observed values.

[22] A different procedure is followed if the goal is to decide the p, d, q order using econometric methods. The AIC/BIC score should be compared to choose the best fitted p, q order.

$$R^2 = 1 - \frac{\sum_i^n (y_i - \widehat{y_i})^2}{\sum_i^n (y_i - \bar{y_i})^2} \quad (1.4.5)$$

(3) MAPE (mean absolute percentage error), which is given by equation (1.4.6), where $\widehat{y_i}$ is the predicted value and y_i is the actual observed value.

$$MAPE = \frac{\sum_i^n |\frac{y_i - \widehat{y_i}}{y_i}|}{n} * 100 \quad (1.4.6)$$

(4) Direction accuracy, which is a measure of how accurately market or commodity up/down movement is predicted by the model, which is technically defined as the counts of M:

$$M = 1 \ if \ (\widehat{y_{i.t+1}} - y_{i,t}) * (y_{i,t+1} - y_{i,t}) > 0 \quad (1.4.7)$$

Table 1.5 presents the four selection criteria for both forecasting models. An entry of "Yes" in the predictors column means that this forecasting model includes the sentiment information. As we can see from Table 1.5, there is a significant reduction in RMSE (from 14.09 to 13.80) when the ARIMA forecasting model is used with predictors as events, which in my case are the sentiment features. Meanwhile, the R-squared increases from 0.9526 to 0.9551 if we use the predictor forecasting model. Moreover, there is a significant decrease in the value of MAPE, which is 0.087 in predictor model instead of 0.091 in the other model. Finally, the direction accuracy is also higher when using the sentiment information.

Table 1.5: ARIMA Model Fit Characteristics for Bitcoin Price

Bitcoin price	Predictors	Model Fit statistics			
		RMSE	R-squared	MAPE	Direction
	Yes	13.7952	0.9551	0.0874	23
	No	14.0946	0.9526	0.0913	21

Using sentiment features as part of the prediction process in the ARIMA model is a more robust approach than the traditional forecasting methods. As we can tell from the values of RMSE, R-squared, MAPE, and direction accuracy in Table 1.5 for both models, the proposed model with sentiment features had a superior performance over the one without. Since Expert System (ARIMA) is a customizable and scalable technique, our proposed model is bound to perform well when applied to a wide range of financial assets. In the next section, I apply this model with predictors in the real practical world to examine investment performance in the Bitcoin market.

5 Exercises

In section 4, I investigated the relationship patterns between Bitcoin returns and sentiment features using Granger Causality analysis, as well as the prediction power of the ARIMA forecasting models. In this section, I use the forecasting model developed in section 4.2 to address real-world issues. The first practical task makes use of the sentiment ARIMA model developed in section 4.2, aiming to predict the daily Bitcoin price using an instant updated information set. The second practical exercise provides an investment strategy that yields approximately 20% annual returns for investors.

5.1 Pseudo Real-Time Simulation

Pseudo real-time simulation is a new concept and means that each prediction involves a fixed number of computational steps for each input source. A pseudo real time simulation system that can be used for evaluating real-time Bitcoin instant prices using past information is presented in this section.

I implement the same ARIMA model in section 4.2 twice: first with sentiment information as independent variables, and a second time without them. For each implementation, instead of utilizing the two-step procedure, I adopt a day-by-day strategy. At first, I choose the first 60 days as the original training set. The predicted price of the 61th day is obtained after I run the ARIMA model on the first 60 observations. Next, I predict the 62nd day's Bitcoin price using the previous 61 observations. Similarly, the strategy keeps going on and on until the predicted price is attained for the 238th day (the last day of the whole sample). After I applied the simulation strategy twice on both models (with and without the sentiment information), the predicted prices are obtained separately. Figure 1.5.1 plots the observed Bitcoin daily weighted prices, predicted prices under the sentiment model, as well as the predicted prices of the non-sentiment model starting from day 61. The blue line represents the real observed Bitcoin price, and the red line represents the forecasting price under the predictor model. It is clear that the prices are more accurately predicted using the sentiment information from the graph. Meanwhile, green dots represent the predicted Bitcoin prices without the sentiment information. It is clear that quite a few green dots are "off the track" along the blue line,

which shows that the forecasting model without sentiment information lacked accuracy to some extent.

Figure 1.5.1: Real-time prediction for daily Bitcoin prices

If figure 1.5.1 is not enough to show the differences between two forecasting models, I have also computed the same selection criteria as in section 4.2, which are presented in Table 1.6. When comparing the values of RMSE, R-squared, MAPE, and direction accuracy, the forecasting model with sentiment information obviously has better performance than the model without predictors. Further, using the same predictor model, we observe that the real-time forecasting has smaller errors, higher direction accuracies, and better fitting statistics than forecasting using the 75% sample as the training set in

section 4.2. In particular, the direction accuracy of real- time prediction under of the sentiment model is 158 out of 178, whereas it is only 23 out of 58 in the previous section (Table 1.5). Real-time prediction is thus a better choice if when considering forecasting prices and investing in the Bitcoin market.

Table 1.6: Real-Time Prediction

Bitcoin price	Predictors	Model Fit statistics			
		RMSE	R-squared	MAPE	Direction
	Yes	12.6452	0.9630	0.0887	158
	No	16.9813	0.9349	0.0972	153

5.2 Investment Strategy

In the previous section, I have provided a real-time forecasting strategy of the Bitcoin price, which has even better performance than the ARIMA forecasting model presented in section 4.2. Therefore, using those predicted Bitcoin prices based on instant information, I propose a strategy for making intelligent sell/buy decisions. I propose a straightforward investment strategy under a simple assumption that we can make transactions at any given time (Mittal and Goel, 2013). The general steps of the strategy are as follows:[23]

(1) Pre-Calculation: Maintain a running average and standard deviation of actual adjusted Bitcoin price of previous X days.

[23] This investment strategy comes from the stock portfolio management in Mittal and Goel (2013)'s paper. And I applied the similar strategy on Bitcoin market. And I improved the strategy by optimization the x, y and Z parameter.

(2) Buy Decision: If the predicted price for the next day is Y standard deviations less than the mean, we buy Bitcoin; otherwise, we wait.

(3) Sell Decision: If the price is Z standard deviations more than the actual adjusted value at the buy time, we sell Bitcoin; otherwise, we hold.

Note that the above strategy involves three parameters: X, Y, and Z. Therefore, optimal parametrization is needed initially. Let X, Y, and Z belong to ranges [1, 30], [0, 2], and [0, 2], respectively. I ran this strategy on a data sample for September 2014 within the assigned ranges for X, Y, and Z, and the optimal solution is x=15, y=1, and z=1. Next, I ran this strategy on September 2015 as an example, supposing that we had $1 in total, which could buy 0.0002BTC on September 1. The strategy produced 12 transactions for this month and we received a total of $1.0465 at the end of this month.

The other key point is the transaction cost for trading Bitcoins. There is no transaction cost for trading Bitcoin itself. However, the online exchange platform requires a transaction fee, which varies from 0.2% to 0.25% per transaction, depending on different platforms. Moreover, most platforms also require a deposit fee equaling .05% of the entire transaction, with a minimum of $7.50. Because one can always get the deposit fee back if desired, I only consider the transaction fee as a transaction cost here. For my example, there were 12 transactions indicated for September 2015; thus the transaction cost is $0.03 for our case. Table 1.7 presents the monthly returns under this portfolio management strategy. To summarize, this strategy yields a 1.65% monthly return, which would produce an annual return of 19.8%.

Table 1.7: Monthly Returns of the Investment Strategy

	Monthly Mean	Predicted with Sentiment	Without sentiment	Cost	Net Returns
Returns	-1.41%	4.65%	4.37%	3%	1.65%

6 Discussion

6.1 Social Media Impacts on Other Financial Markets

Classical finance theory leaves no role for investor sentiment. Rather, this theory argues that competition among rational investors, who diversify to optimize the statistical properties of their portfolios, will lead to an equilibrium whereby prices equal the rationally discounted value of expected cash flows and where the cross-section of expected returns depends only on the cross-section of systematic risks. Even if some investors are irrational, classical theory argues that their demands are offset by arbitrageurs and thus have no significant impact on prices. However, in this chapter I present empirical evidence for the fact that social media sentiment has significant effects on Bitcoin returns. It is natural to question why social media would affect Bitcoin markets. It can be explained in the way that social information drives investors' attitudes that further form their expectations and investment decisions. Therefore, it makes sense to examine the impacts of social media on financial markets.

Several recent papers have investigated whether investor sentiment affects asset pricing. Broadly speaking, these studies can be categorized into one of two groups: studies that relate various sentiment proxies to returns at an individual-stock level (Hvidkjaer, 2008; Barberm, Odean, and Zhu 2009) and studies that relate proxies for

aggregate sentiment to broad market returns or the returns to a market sector thought to be prone to sentiment influence (Teo and Woo, 2004; Frazzini and Lamont, 2008). Particularly, according to A. Mittal et al. (2013) and Rao and Srivastava (2014), Twitter sentiments have already been used in predicting the DJIA and NASDAQ-100 indices. For instance, Rao and Srivastava (2014) reject the null hypothesis that the Twitter features (positive and negative) do not affect returns in the financial markets with a high level of confidence.

By contrast, we observe that only negative tweets have significant impacts on Bitcoin returns, while both positive and negative tweets affect the stock markets. Intuitively, investors usually pay more attention to negative information and are more sensitive to bad news when facing a relatively new asset. From another perspective, Bitcoin is indeed a new concept to all investors, and the most common way of becoming familiar with Bitcoin is through social media. Therefore, it is obvious that social media's impact on the Bitcoin market, which is at an early stage, acts differently than in other financial markets. It would be really interesting to explore relevant research on financial markets focused on behavioral economics, such as why people always care more about negative information when just starting to know about an asset.

6.2 Sentiment Robustness Test

In section 3, I adopted the NLTK sentiment analyzer to mine Twitter texts. In this section, I use the other sentiment analyzer, called Pattern, to conduct the robustness test. Pattern is the default sentiment analyzer in TextBlob, which outputs a float number from

-1 to 1 after the user inputs text contents. This number, also called a polarity score, is the quantitative assessment for sentiment in that sentence. The more positive the value, the closer it approaches to 1; the more negative the value, the closer it approaches to -1.

When implemented by Pattern, the sentiment analysis process is supported by a large lexicon that contains frequent adjectives with the sentiment score tagged manually on each word. The score indicates the assessment of sentiment for the word. When one sentence is passed to TextBlob, based on the lexicon, TextBlob retrieves the value of each adjective and calculates the average for them. The average score is just the final sentiment score for the text and the recorded accuracy for the movie reviews corpus was about 75% (De Smedt and Daelemans, 2012b). For example, if the text "this is a nice day!" is processed by the Pattern analyzer, the output sentiment is 0.75.[24]

Instead of the Naive Bayes analyzer, I have implemented the Pattern analyzer to obtain the sentiment information from Twitter data. I then apply these sentiment variables to the same Granger Causality analysis and ARIMA forecasting models. The results presented in Appendix A.2 are consistent with the results reported previously. Therefore, I conclude that the sentiment information contributes to predicting the Bitcoin market no matter which sentiment analysis tool is used.

6.3 Extensions

For the sake of investigating the efficiency of the Bitcoin market, this chapter has provided the theoretical background and related empirical studies on efficient market

[24] This example comes from Cao (2014). http://dare.uva.nl/cgi/arno/show.cgi?fid=544733

hypothesis tests. Analyzing the impact of social media sentiment on Bitcoin returns is a good way to prove the market's inefficiency. Through conducting experiments on the Bitcoin market of the relationship between its returns/prices/trading volumes and relevant Twitter-based sentiment information, the research further responds to the question of how to validate the impact of social media on Bitcoin returns. Social media information drives investors' attitudes that further form their expectations of future Bitcoin prices. The attitudes can be parsed from tweets, while the sentiment analysis tool TextBlob was harnessed for exposing the attitudes.

The revealed attitudes, namely the sentiment information, include positive/negative sentiment, bullishness, agreement, and message volume. Meanwhile, I have also collected the Bitcoin market instruments (price, volatility, trading volume, etc.) from bitcoincharts.com. Granger Causality analysis confirms relationships between Twitter-based variables and Bitcoin short-term market performances. Results show that negative dimensions of public mood have improved power for tracking movements of Bitcoin returns. I have also investigated various features such as how previous-week sentiment features affect the next day's price, trading volume, and volatility. Then, I have verified the strong performance of the ARIMA forecasting model in both theoretical and practical cases. Moreover, I discuss how this forecasting model brings the wisdom of the crowd to invest in the real Bitcoin market. Using this technique the investor can attain annual returns of approximately 20%. It is no surprise that this approach is far more robust and gives far better results than any previous work. In the near future, sentiment

analysis promises to be an effective strategy for investments not only in the Bitcoin market, but also in other financial markets.

To summarize, this chapter not only theoretically explains the principle behind the research, but also practically tests the research hypothesis. Furthermore, the use of big data and data mining methods increases its value as a quantitative financial economic study. Ultimately, this research provides statistically sufficient evidence for the existence of causality between social media sentiment and the Bitcoin market, and investors can make profits by taking advantage of sentiment forecasting models.

Finally, it is worth mentioning that many factors still have not yet been taken into consideration. First, the Twitter data may not really fully map genuine public sentiment; they only consider Twitter users and English-speaking users. We may need to consider worldwide sentiment, since Bitcoin is a popular global market. Second, it may be possible to obtain a higher correlation if the actual mood[25] is studied. It could be hypothesized that people's moods affect their investment decisions, and thus also the correlation. Moreover, we need to adopt a larger data set when computing the optimal strategy for investment decisions. All these areas remain for future research.

[25] Actual mood here refers the market composite real sentiment. It should contain information not only from social media, but also investors who don't use social media.

CHAPTER TWO

Examine the Episodes of Exuberance and Collapse in the Chinese Stock Market and the Second-Board Market

1 Introduction

According to the International Monetary Fund (IMF), in 2014 China's economy was the world's second largest in terms of nominal GDP and the world's largest by purchasing power parity.[26] It is the world's fastest-growing major economy, with growth rates averaging 10% over the past 30 years. And financial markets have always played a prominent role in China's economy. There are two stock exchanges (the Shanghai Stock Exchange and the Shenzhen Stock Exchange) operating independently in mainland China's stock market, which had a market value of $4.48 trillion as of November 2014, making it the second largest stock market in the world. The Shanghai Stock Exchange (SSE), based in Shanghai, is the world's third largest stock market by market capitalization—around US$5.5 trillion as May 2015. The current exchange was re-established on November 26, 1990 and began operation on December 19 of the same year.[27] From 2001 to 2005, there was a four-year market slump; SSE's market value shrank by half after reaching a peak in 2001. A ban on new IPOs was put in place in April 2005 to rectify the situation and allow more than US$200 billion of mostly state-owned equities to be converted to tradable shares. The SSE resumed full operation after the ban on IPOs was removed in May 2006. The world's second largest (US$21.9 billion)

[26] This information comes from Wikipedia: https://en.wikipedia.org/wiki/Economy_of_China
[27] Shanghai Stock Exchange introduction: https://en.wikipedia.org/wiki/Shanghai_Stock_Exchange

IPO, by the Industrial and Commercial Bank of China, was launched in both the Shanghai and Hong Kong stock markets. During 2007 and 2008, a "stock market frenzy" ensued in China's market, making China's stock exchange temporarily the world's second largest in terms of turnover. After reaching a peak of 6,124.044 points on October 16, 2007, the benchmark Shanghai Composite Index decreased in 2008, mainly due to the impact of the global economic crisis that started around 2008. Figure 2.1.1 (from Wikipedia) shows the course of the SSE Composite Index from 1991 to the beginning of 2009. We can observe unusual surges and declines in prices during certain sub-periods. What features caused those movements, whether bubbles existed, and whether the bubbles were rational or behavioral are among the most actively debated issues in financial economics.

Many researchers attributed the episodes to financial bubbles. Examples include Greenspan (1996), Thaler (1999), Shiller (2000), The Economist (2000), Cooper et al. (2001), Ritter and Welch (2002), Ofek and Richardson (2002), Lamont and Thaler (2003), and Cunado et al. (2005). Among those sources, the remark by US Federal Reserve Chairman Alan Greenspan (1996) on December 5, 1996, is the most citied, using the phrase "irrational exuberance" to characterize stock market arbitrageur behaviors. The concept has been influential in thinking about financial markets and herd investment behavior.

One purpose of this chapter is to empirically examine the SSE stock market performance in relation to market perceptions of exuberance ("bubbles"). In particular, it is of interest to determine whether the exuberance was supported by empirical evidence

in the data. To achieve this goal, I first define financial exuberance in the time series context in terms of explosive autoregressive behavior and then adopt some econometric methods based on previous studies (Phillips, Wu and Yu, 2009, 2011, 2013), as well as a mildly explosive regression asymptotic to assess the empirical evidence of exuberant behavior in the SSE stock market. Moreover, these time series models can identify the dates of booms and collapses of the stick market index. Henceforth, the key issues include identifying the origination, termination, and extent of the explosive behavior. In the end, I successfully identify explosive periods of price exuberance in the SSE stock market.

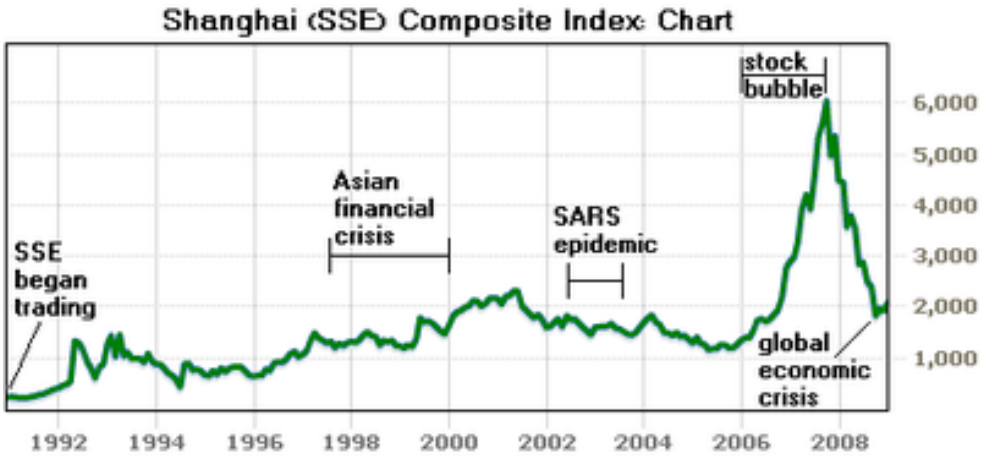

Figure 2.1.1: The Shanghai (SSE) composite index: 1991 to start of 2009[28]

Among the potential explanations for explosive behavior in economic variables, the most prominent are perhaps models with rational bubbles. Accordingly, I related the analysis of explosive behavior to the rational bubble literature, where it is well known

[28] This figure comes from the Wikipedia: https://en.wikipedia.org/wiki/Shanghai_Stock_Exchange

that standard econometric tests encounter difficulties in identifying rational asset bubbles (Flood and Garber, 1980; Flood and Hodrick, 1986; and Evans, 1991). In order to resolve those problems, I adopt four econometric tests (Phillips, 2013) to locate "exploding" subsamples of data and detect periods of exuberance. The econometric approach utilizes some new techniques that permit the construction of valid asymptotic confidence intervals for explosive autoregressive processes and tests of explosive characteristics in time series data.[29] In addition, this approach can detect the presence of exuberance in the data and date-stamp the origination and collapse of periods of exuberance.

I have applied this econometric approach to the SSE index over the full sample period from 04/04/2005 to 11/04/2013 and some sub-periods. Those time series models confirm the existence of multiple episodes of exuberance and successfully date the origin and conclusion of each explosive episode. The statistical evidence from these models indicates that explosiveness started in 2005 and the explosive environment continued until 2006. There was another sub-period of exuberance of the stock market around 2007. In addition to the SSE stock market, this chapter also investigates the Growth Enterprise Market (GEM), which is a stock market set up by the Hong Kong Stock Exchange (HKSE) for growth companies that do not fulfill the requirements of profitability or track record for inclusion in the HKSE. In the 1970s, the Growth Enterprise Market was designed to accommodate vigorous scientific and technological innovations and the boom of start-ups around the world. After three decades of development and especially after the

[29] This chapter is an empirical application of the time series techniques from P. Philips (2009. 2011, 2013). I adopted Philips' methods to analyze the explosive behaviors of Chinese stock market. This chapter contains many equations and definitions from Philips's papers.

international financial crisis, the GEM is again showing great vitality and appeal to a great many investors. And there was a suspicious surge of the GEM index during 2013, with many scholars arguing that there are bubbles in this secondary board market. In order to analyze abnormal behavior in the GEM stock market, I have applied the same methodology (described above for SSE) to test explosive episodes of GEM stock prices from 05/02/2011 to 11/04/2013.

The remainder of this chapter is organized as follows. Section 2.1 defines market exuberance, discusses model specification issues, and relates exuberance to the earlier literature on rational bubbles. Section 2.2 discusses some econometric issues and the econometric methods used to test the exuberance behavior of stock prices. Section 3 describes the data used in this chapter. The empirical results are reported in Section 4. Section 5 concludes and provides some interesting related questions for further research.

2 Methods

2.1 Definition of Stock Market Bubbles

"Irrational exuberance" could be interpreted as a typically cryptic warning that the market might be overvalued and at risk of a financial bubble. Theoretical studies on rational bubbles in the stock market include Blanchard (1979), Blanchard and Watson (1982), Shiller (1984), Tirole (1982, 1985), Evans (1989), Evans and Honkapohja (1992), and Olivier (2000); and empirical studies include Shiller (1981), West (1987, 1988), Campbell and Shiller (1987, 1989), Diba and Grossman (1988), Froot and Obstfeld (1991), Wu (1997), Flood and Hodrick (1990) and Gurkaynak (2008), investigating

econometric methodologies and testing for financial bubbles. It is well known in the rational bubble literature that bubbles, if they are present, should manifest explosive characteristics in prices. Intuitively, much of the literature has defined exuberance in terms of explosive behavior propagated by a process of the form $x_t = \mu_t + \delta x_{t-1} + \varepsilon_{x,t}$,where for certain sub-periods of the data δ>1.

The concept of rational bubbles can be illustrated using stock pricing models. The most common model is the Discount Dividend Pricing Model (DDP),

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f}\right)^i E_t(D_{t+i} + U_{t+i}) \quad (2.2.1)$$

where P_t is the after-dividend price of the asset, D_t is the payoff received from the asset, r_f is the risk-free interest rate, and U_t represents the unobservable fundamentals. The DDP model is not the only model to accommodate bubble phenomena (Cochrane, 2005; Cooper. 2008); another similar model used frequently in pricing models is called Free Cash Flow to Equity Model:

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f}\right)^i E_t(F_{t+i} + U_{t+i}) \quad (2.2.2)$$

where F_t is the free cash flow of the company.

Using DDP as an example, we can see that analysis of financial bubbles starts from the standard asset pricing equation:

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f}\right)^i E_t(D_{t+i} + U_{t+i}) + B_t \quad (2.2.3)$$

where B_t is the bubble component. If bubbles are present, $B_t \neq 0$. Diba and Grossman (1988) argued that, if a bubble presents, then B_t has an explosive property characterized by:

$$E_t(B_{t+1}) = (1 + r_f)B_t \quad (2.2.4)$$

If no bubbles are present, the degree of nonstationary of the asset price is controlled by D_t and U_t. For example: If D_t is an I(1) process and U_t is either an I(0) or an I(1) process, then the asset price is at most an I(1) process. So if U_t or D_t is at most I(1), empirical evidence of explosive behavior in asset prices may be used to conclude the existence of bubbles. Campbell and Shiller (1989) took a log-linear approximation of equation (2.3.3), for which we obtaint:

$$E_t(b_{t+1}) = (1 + \exp(\overline{d - p})b_t) \quad (2.2.5)$$

where $\overline{d - p}$ is the average log dividend-price ratio. Lee and Phillips (2011) also provided a detailed analysis of the accuracy of this log linear approximation under various conditions. Henceforth, I focus on the price/dividend ratio to test for stock market bubbles. Explosive or mildly explosive behavior in the $\overline{d - p}$ ratio is a primary indicator of market exuberance during the inflationary phase of a bubble and this time series manifestation may be subjected to econometric testing. According to Cochrane (1992) and Ang and Bekaert (2006), the price-dividend ratio is a function of the discount factor and the dividend growth rate in the absence of bubbles.

However in Chinese stock market, there are little dividend data for most stocks. Also there are only quarterly data for some companies, and dividends are even negative for some sectors. Femald and Rogers (2002) considered dividends as κ (some constant) times earnings for the Chinese stock market. Therefore, instead of the price/dividend ratio, I use the price/earnings ratio to test for exuberance behavior in this chapter.

2.2 Econometric Tests for Multiple Bubbles

Based on much of the previous literature, the basic idea of testing for the explosive behavior of time series x_t is applying the Augmented Dickey Fuller (ADF) test for a unit root against the alternative explosive root (the right tailed):

$$x_t = \mu_t + \delta x_{t-1} + \sum_{j=1}^{J} \varphi_j \Delta x_{t-j} + \varepsilon_{x,t}\,, \varepsilon_{x,t} \sim NID(0, \sigma_x^2) \quad (2.2.6)$$

In terms of lag choice, I use significance tests to determine the lag order J (Campbell and Perron, 1991). According to Diba and Grossman (1987, 1988), the identification of explosive characteristics in the data is equivalent to the detection of a stock bubble if the discount rate is time invariant. Using standard unit root tests applied to the real US Standard and Poor's Composite Stock Price Index over the period 1871-1986, Diba and Grossman (1988) tested levels and differences of stock prices for non-stationarity, finding support in the data for non-stationarity in levels but stationarity in differences. Since differences of an explosive process still manifest explosive characteristics, these findings appear to reject the presence of a market bubble in the data. Although the results were less definitive, further tests by Diba and Grossman (1988) provided confirmation of co-integration between stock prices and dividends over the same period, supporting the conclusion that prices did not diverge from long-run fundamentals and thereby giving additional evidence against bubble behavior. Evans (1991) criticized this approach, showing that a time series simulated from a nonlinear model that produces periodically collapsing bubbles manifests more complex bubble characteristics that are typically not detectable by standard unit root and co-integration tests. He concluded that standard unit root and co-integration tests are inappropriate tools for detecting bubble behavior because

they cannot effectively distinguish between a stationary process and a periodically collapsing bubble model. Patterns of periodically collapsing bubbles in the data look more like data generated from a unit-root or stationary auto-regression than a potentially explosive process. Recursive tests of the type undertaken in our paper are not subject to the same criticism and, as demonstrated in our analysis and simulations reported below, are capable of distinguishing periodically collapsing bubbles from pure unit root processes.

Phillips, Wu and Yu (2011; PWY hereafter) proposed a recursive method that can detect exuberance in asset price series during an inflationary phase. The approach is anticipative as an early warning alert system, so that it meets the needs of central bank surveillance teams and regulators, thereby addressing one of the key concerns articulated by Cooper (2008). The method is especially effective when there is a single bubble episode in the sample data, as in the 1990s NASDAQ episode analyzed in the PWY paper and in the 2000s U.S. house price bubble analyzed in Phillips and Yu (2011). Similar to the Sup Augmented Dickey Fuller (SADF) test, Homm and Breitung (2012) found that the PWY test was the most powerful detecting multiple bubbles.

In this chapter, I adopted four tests based on different variations of a right-tailed unit root test to test for the existence of a bubble and the dates when the bubble started and burst. The first two tests are the ADF (right-tailed) test and Rolling ADF (RADF) test. The other two tests come from Phillips and colleagues: the SADF test from Phillips et al. (2011, PWY) and the Generalized Sup ADF (GSADF) test from Phillips et al. (2013; hereafter PSY). I illustrate each briefly below.

The first test is just the simple right-tailed ADF test. More formally, the null hypothesis of this test is of a unit root, and the alternative is of a mildly explosive autoregressive coefficient:

$H_0: \delta = 1$
$H_1: \delta > 1$
where δ is the estimated first-order regression coefficient from:

$$x_t = \mu_t + \delta x_{t-1} + \sum_{j=1}^{J} \varphi_j \Delta x_{t-j} + \varepsilon_{x,t}\,, \varepsilon_{x,t} \sim NID(0, \sigma_x^2) \quad (2.2.6)$$

The RADF test is just a rolling version of the first test, in which the ADF statistic is calculated over a rolling window of a size specified by the user, and the RADF statistic is the maximal ADF statistic estimated among all possible windows. The SADF test is based on recursive calculations of the ADF statistics with an expanding window. The SADF statistic is defined as the supremum value of the ADF_{r_2} sequence for $r_2 \in [r_0, 1]$:

$$SADF(r_0) = \sup_{r_2 \in [r_0, 1]} \{\mathrm{ADF}_{r_2}\} \quad (2.2.7)$$

GSADF is a generalization of the SADF test that allows a more flexible estimation window, where the starting point, r1, is allowed to vary within the range: $[\,0, r_2 - r_0]$, and the GSADF statistic is defined as:

$$GSADF(r_0) = \sup_{r_2 \in [r_0, 1],\, r_1 \in [0, r_2 - r_0]} \{ADF_{r_2}^{r_1}\} \quad (2.2.8)$$

According to Cooper (2008), the SADF test is effective when there is a single bubble. And Phillips et al. (2013, PSY) also showed that the SADF test suffers from a loss of power in the presence of multiple periodically collapsing bubbles. However, the GSADF test surmounts this limitation and improves discriminatory power in detecting multiple bubbles. Phillips et al. (2013, PSY) attained critical values for SADF and GSADF test by numerical simulations, where the Wiener process is approximated by

partial sims of 2,000 independent N(0,1) variates and the number is 2,000. We concluded that if the minimum window size r_0 decreases, the critical values of the test statistic increases. And we could use the asymptotic critical values in practical work.

As discussed in the introduction, regulators and central banks concerned with practical policy implementation need to assess whether real-time data provide evidence of financial exuberance. The SADF and GSADF introduced by Phillips et al. (2011, 2013) could also be used as a date-stamping procedure, i.e., if the null hypothesis is rejected, we can estimate the start and the end points of a certain bubble. In particular, for the SADF test, the origination date of a bubble is calculated as the first chronological observation whose ADF statistic exceeds the critical value. And the estimated termination date of a bubble is the first chronological observation after $[start\ point] + logT$, whose ADF statistic goes below the critical value. Further, we imposed a condition that for a bubble to exist its duration must exceed a slowly varying quantity such as $logT$:[30]

$$\hat{r}_e = \inf_{r_2 \in [r_0,1]} \{ r_2 : \mathrm{ADF}_{r_2} > cv_{r_2}^{\beta_T} \} \quad (2.2.9)$$

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e + \log(T)/T,1]} \{ r_2 : \mathrm{ADF}_{r_2} < cv_{r_2}^{\beta_T} \} \quad (2.2.10)$$

where $cv_{r_2}^{\beta_T}$ is the $100\ (1-\beta_T)\%$ critical value of the ADF statistic based on $[Tr_2]$ observations. Similarly for GSADF test, we have:

$$\hat{r}_e = \inf_{r_2 \in [r_0,1]} \{ r_2 : \mathrm{BSADF}_{r_2} > cv_{r_2}^{\beta_{r_2}} \} \quad (2.2.11)$$

[30] All the imposed conditions and equations come from Phillips et. al. (2013, PSY)

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e + \log(T)/T, 1]} \{ r_2 : \mathrm{BSADF}_{r_2} < cv_{r_2}^{\beta_{r_2}} \} \quad (2.2.12)$$

where $\mathrm{BSADF}_{r_2}(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} \{ ADF_{r_1}^{r_2} \}$.

Similarly, we could date stamping if there are two or more bubbles (assume that the first bubble lasts longer). Phillips et al. (2013, PSY) provides more details regarding the date-stamping strategies as well as the detailed proof of the limit theory of those tests. In summary, if there are multiple bubbles present, the SADF test consistently estimates the first bubble and detects the second bubble with a delay. The sequential SADF procedure is consistent, even when the first bubble is shorter than the second bubble. Please check their paper for details.

3 Data

Instead of the SSE index, I have selected the Hushen 300 Index to represent the Chinese stock market, since the Hushen 300 Index includes information from the Shenzhen Stock Exchange, which is the second largest stock market in China. The Hushen 300 Index is a stock price index jointly issued by the Shanghai and Shenzhen stock exchanges on April 8, 2005 to reflect the A share market as a whole. It has a comprehensive market representation, and it can veritably represent stock price fluctuations in the Chinese stock market. Data for the Index are collected by the Windin Company. A second index used in my analysis is the Growth Enterprise Market, also known as the Second Board Market of the Hong Kong Exchange. It consists mainly of entrepreneurial ventures, particularly small and medium-sized high-tech enterprises and other establishments that need financing and development.

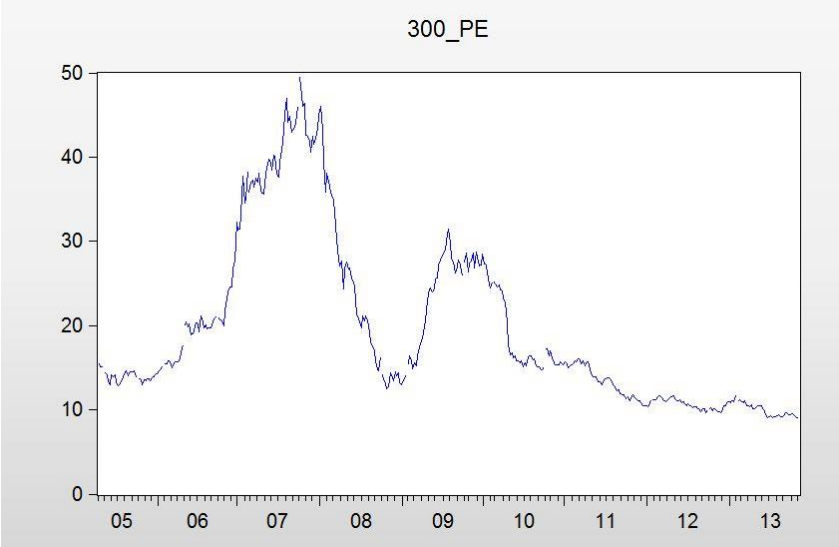

Figure 2.3.1: Time series plot of price/earnings ratio for Husen 300 Index from 04/04/2005 to 11/04/2013, weekly observations

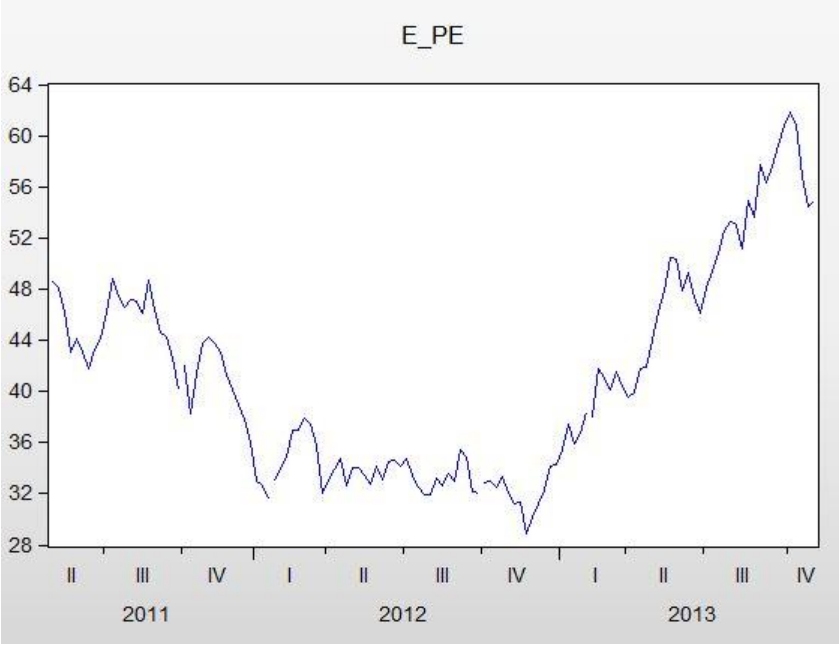

Figure 2.3.2: Time series plot of price/earnings ratio for GEM Index from 05/02/2011 to 11/04/2013, weekly observations

More specifically, the Growth Enterprise Market (GEM) is an alternative stock market operated by the Hong Kong Exchange. It seeks to provide fund- raising opportunities for growth companies of all sizes from all industries, i.e., enterprises that have good growth potential. The rules, requirements, and facilities of GEM are designed to satisfy the needs and standards desired by professional and informed investors. GEM does not require growth companies to have achieved a record of profitability as a condition for listing. The removal of this entry barrier enables growth enterprises to capitalize themselves on the growth opportunities of the region by raising expansion capital under well-established market and regulatory frameworks. In other words, GEM offers investors the alternative of investing in "high growth, high risk" businesses. Different than stocks, GEM also accepts the listing of equity warrants and debt securities of GEM issuers. GEM stocks utilize the same trading, clearing, and settlement systems as stocks listed on the Main Board of the Hong Kong Exchange. The GEM provides a favorable environment for medium-sized and small enterprises to gain more funds, and effectively supplements the Main Board market, occupying an important position in the capital market. Therefore, generally speaking, the Growth Enterprise Market is a stock market with low-entry, high-risk ventures and strict supervision, as well as a cradle for breeding scientific and growing enterprises. This is the main reason for the existence of bubbles in the GEM market.

We are interested in time series in terms of the price/earnings ratio. Therefore I collected weekly data for the Hushen 300 PE Ratio Index, i.e., price/earnings ratio from 04/04/2005 to 11/04/2013 for the whole market, for a total of 433 observations. For the

Second Board market, the GEM Index (average price index) was issued from May 31, 2010, and I also collected the daily GEM PE Ratio Index as well as weekly data, for a total of 617 and 128 observations, respectively. In the end, I applied four tests on the weekly GEM PE Ratio Index data from 05/02/2011 to 11/04/2013. Figures 2.3.1 and 2.3.2 present the research samples of the price/earnings ratio for the Chinese stock market and the GEM market.

4 Results

In order to test the exuberance behaviors and identify the exact dates when explosive behaviors started and ended, I have applied four different tests on the Hushen 300 price/earnings ratio and GEM price/earnings ratio separately. Figure 2.4.1 displays the results of the classical right-tailed ADF test on the Hushen 300 price/earnings ratio. Results from Figure 2.4.1 fail to reject the null hypothesis since the t-statistic is less than the critical value. In other words, there was no bubble in the Chinese stock market from 2005 to 2013. However, results from the other three econometric tests do not agree. Figure 2.4.2 shows the results from the rolling right-tailed ADF test on the Hushen 300 price/earnings ratio. The results show six exuberance behaviors during this time period, with episodes of exuberances occurring in 2005, 2006, 2008, 2009, 2010, and 2012.

Right Tailed ADF Tests
Sample : 04/04/2005 11/04/2013
Included observations: 433
Null Hypothesis: 300_PE has a unit root
Date: 11/20/13 Time: 19:49

		t-Statistic
ADF		-0.813069
Test critical values:	99% level	0.781336
	95% level	-0.070853
	90% level	-0.482001

*Right-tailed test

Figure 2.4.1: Right-tailed ADF test on Hushen 300 price/earnings ratio

Right Tailed ADF Tests
Sample : 04/04/2005 11/04/2013
Included observations: 433
Null Hypothesis: 300_PE has a unit root
Date: 11/20/13 Time: 19:53

		t-Statistic
max RADF		2.775598
Test critical values:	99% level	0.647936
	95% level	-0.039277
	90% level	-0.397694

*Right-tailed test

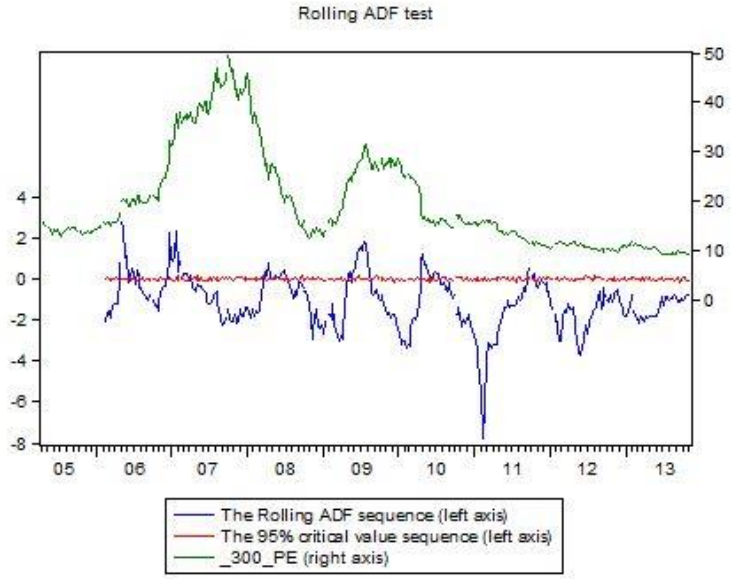

Figure 2.4.2: Rolling right-tailed ADF test on Hushen 300 price/earnings ratio (1000 times)

Right Tailed ADF Tests
Sample : 04/04/2005 11/04/2013
Included observations: 433
Null Hypothesis: 300_PE has a unit root
Date: 11/20/13 Time: 19:56

		t-Statistic
SADF		4.070472
Test critical values:	99% level	2.089376
	95% level	1.428670
	90% level	1.190447

*Right-tailed test

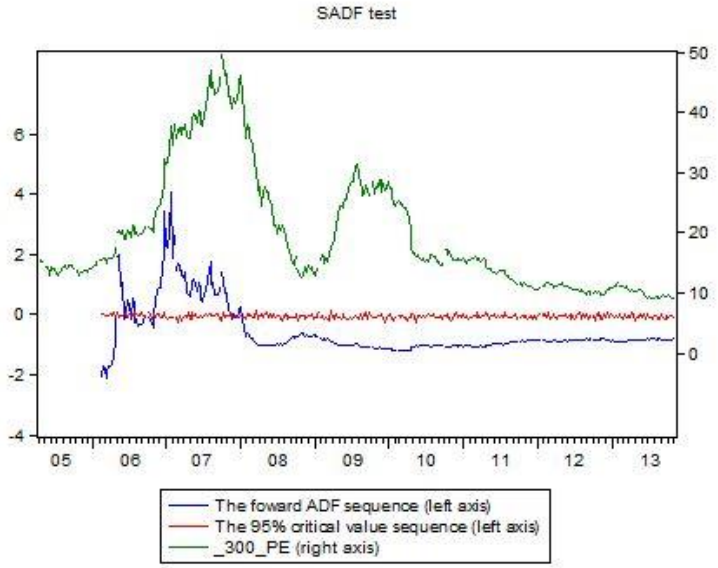

Figure 2.4.3: SADF test on Hushen 300 price/earnings ratio (1000 times)

However, according to Figures 2.4.3 and figure 2.4.4, the results from SADF and GSADF only detect two episodes of exuberance in the Chinese stock market between 04/042005 and 11/04/2013. One happened at the beginning of 2006, and the other lasted around a year, from the beginning of 2007 to the end of 2007. Based on the results from all four extended ADF tests, we can conclude that there were at least two periods of

exuberance in the Chinese stock market between 2006 and 2007. Clearly the traditional right-tailed ADF test fails to find multiple bubbles if they occurred during relatively short time periods. Moreover, among the other three extended right-tailed ADF tests, SADF and GSADF test results are more robust when compared with RADF. In sum, through these time series models, we have sufficient empirical evidence to validate exuberances in the Chinese stock market from 2005 to 2013, especially during 2006 to 2007.

Right Tailed ADF Tests
Sample : 04/04/2005 11/04/2013
Included observations: 433
Null Hypothesis: 300_PE has a unit root
Date: 11/25/13 Time: 11:17

		t-Statistic
GSADF		4.070472
Test critical values:	99% level	2.640947
	95% level	2.164189
	90% level	1.919450

*Right-tailed test

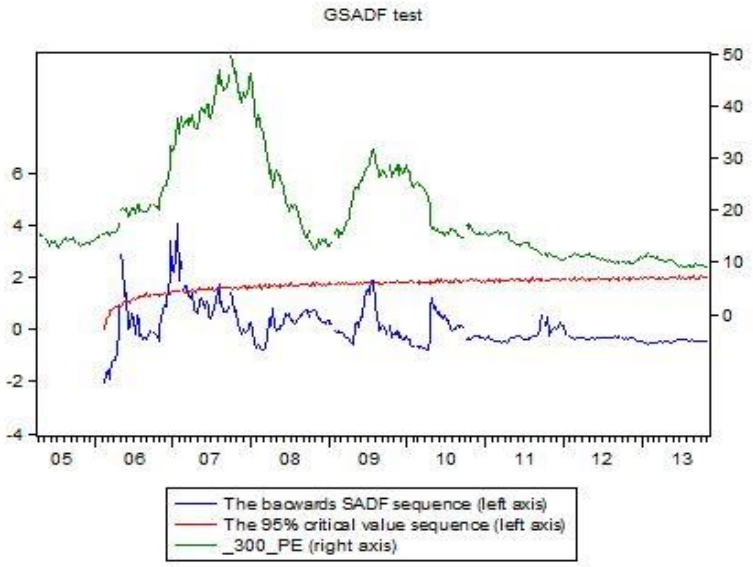

Figure 2.4.4: GSADF test on Hushen 300 price/earnings ratio (1000 times)

Right Tailed ADF Tests
Sample : 05/02/2011 11/04/2013
Included observations: 128
Null Hypothesis: E_PE has a unit root
Date: 11/25/13 Time: 19:11

		t-Statistic
ADF		-0.684004
Test critical values:	99% level	0.591543
	95% level	0.035788
	90% level	-0.326270

*Right-tailed test

Figure 2.4.5: Right tailed ADF test on GEM price/earnings ratio

Right Tailed ADF Tests
Sample : 05/02/2011 11/04/2013
Included observations: 128
Null Hypothesis: E_PE has a unit root
Date: 11/25/13 Time: 19:12

		t-Statistic
max RADF		1.221728
Test critical values:	99% level	0.704723
	95% level	-0.011929
	90% level	-0.384918

*Right-tailed test

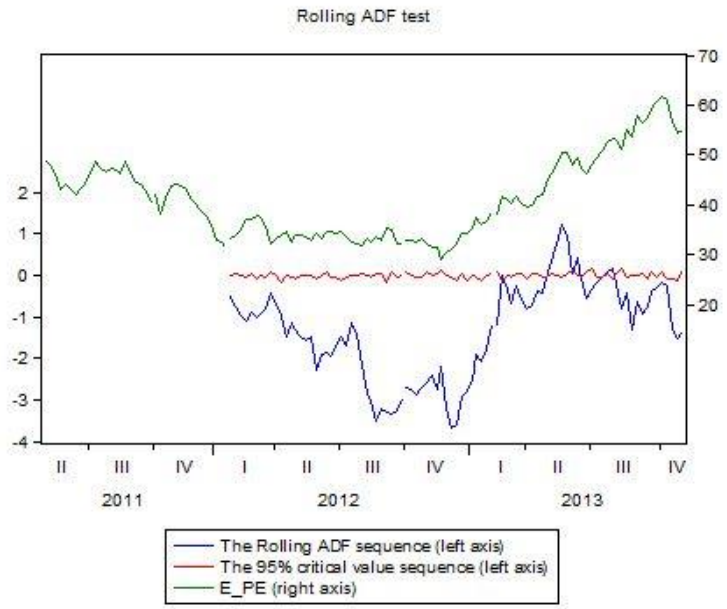

Figure 2.4.6: Rolling right tailed ADF test on GEM price/earnings ratio (1000 times)

Similarly, I applied the same methodology to the GEM stock market between 05/02/2011 and 11/04/2013. Figure 2.4.5 presents the results of the regular right-tailed ADF test on GEM price/earnings ratios in our sample. Not surprisingly, the right-tailed ADF's results do not reject the null hypothesis, i.e., this traditional test fails to detect bubbles again. However, the results from RADF and SADF confirm the existence of a single-bubble GEM stock market between 05/2011 and 11/2013. We can tell from Figures 2.4.6 and 2.4.7 that this single bubble occurred in 2013.

Right Tailed ADF Tests
Sample : 05/02/2011 11/04/2013
Included observations: 128
Null Hypothesis: E_PE has a unit root
Date: 11/25/13 Time: 19:15

		t-Statistic
SADF		0.276579
Test critical values:	99% level	1.675481
	95% level	1.217817
	90% level	0.892325

*Right-tailed test

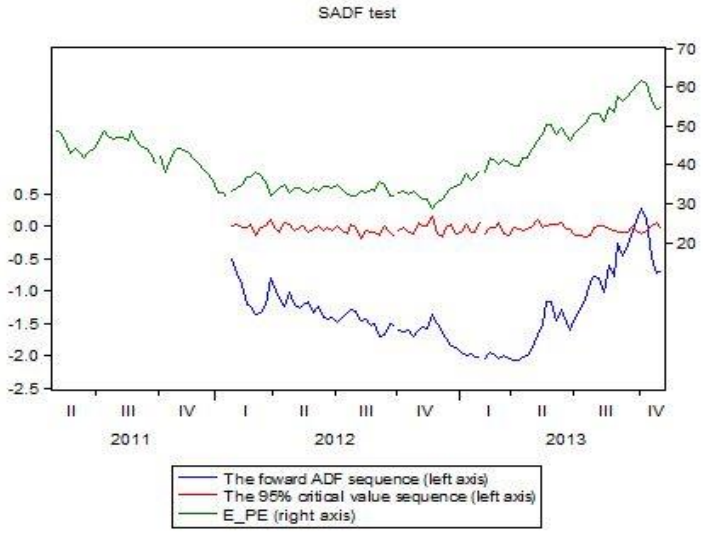

Figure 2.4.7: SADF test on GEM price/earnings ratio (1000 times)

Right Tailed ADF Tests
Sample : 05/02/2011 11/04/2013
Included observations: 128
Null Hypothesis: E_PE has a unit root
Date: 11/25/13 Time: 16:11

		t-Statistic
GSADF		1.549038
Test critical values:	99% level	2.254067
	95% level	1.623397
	90% level	1.317723

*Right-tailed test

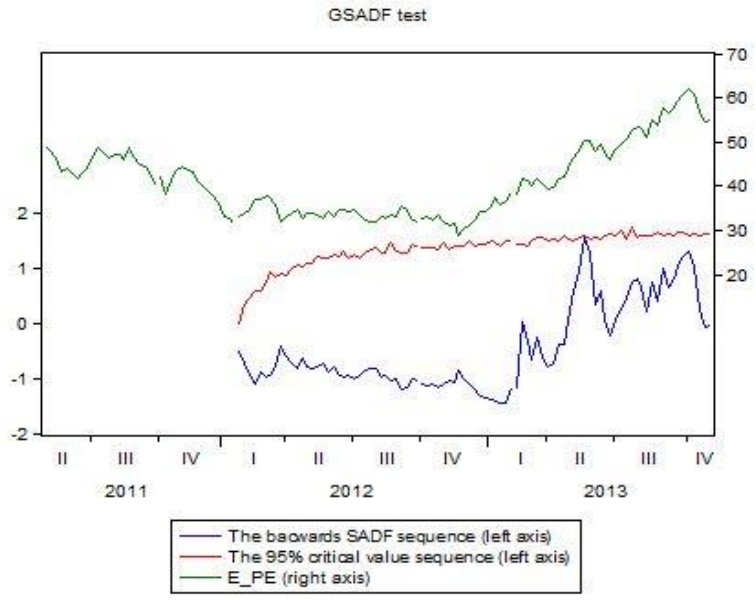

Figure 2.4.8: GSADF test on GEM price/earnings ratio (1000 times), from 2011

Figure 2.4.8 presents the results of the GSADF test on the GEM price/earnings ratio. It is noteworthy that the GSADF test also fails to reject the null hypothesis. The reason that GSADF did not detect any bubble in this context might because the sample size is too small. The sample size is not statistically sufficient if we need to roll over twice. Therefore I expanded the sample size to 205 observations, using data from 10/26/2009 to 11/04/2013. Figure 2.4.9 presents the results of GSADF on the new sample

size. However, the results again are not able to reject the null hypothesis, i.e., there was no exuberance in the GEM stock market between 10/2009 and 11/2013. Combined with the results in Figure 2.4.5, we could conclude that there is not sufficient empirical evidence to claim exuberance in the GEM markets from 2009 based on those four extended ADF tests. In future research, we could try to improve the empirical test by adding samples to the current pool or by using newly developed econometric models.

Right Tailed ADF Tests
Sample : 10/26/2009 11/04/2013
Included observations: 205
Null Hypothesis: E_PE has a unit root
Date: 11/25/13 Time: 11:39

		t-Statistic
GSADF		1.549038
Test critical values:	99% level	2.391258
	95% level	1.902664
	90% level	1.579244

*Right-tailed test

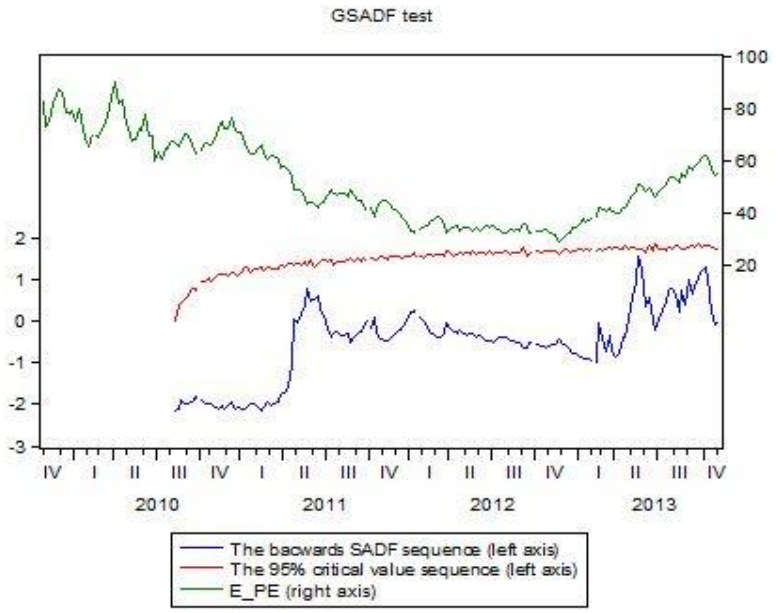

Figure 2.4.9: GSADF test on GEM price/earnings ratio (1000 times), from 2009

5 Discussion

For the sake of investigating the explosive behaviors of the historical Chinese stock market index and the GEM stock market, this chapter defines financial exuberance in the time series context in terms of explosive autoregressive behavior and adopts four advanced time series models to examine if any exuberance existed in the Chinese and GEM stock markets. I selected the Hushen 300 Index, a stock price index jointly issued by the Shanghai and Shenzhen stock exchanges on April 8, 2005, to represent the China A share market as a whole. I also analyzed an index of the GEM stock market, called the Second Board Market, which is associated with the Main-Board Market of a different class of securities markets. It is a relatively new and high-risk market, and great fluctuations of its price movements have drawn lots of concern from many scholars.

In section 2, I introduce an econometric methodology recently developed by P. Philips et. al. (2013) to assess the empirical evidence of exuberant behaviors in those two stock markets based on forward recursive regression tests and a mildly explosive regression asymptotic. Much prominent research has been focused on testing exuberant behavior in financial markets. The most common method involves using the right-tailed ADF test to examine price/earnings time series. This traditional method works well if there was only a single episode of exuberance during the sample time period. However this test always fails to detect multiple episodes of exuberance if the sample time period is too short. The regular right-tailed ADF test might just consider the price/earnings ratio time series to be stationary if this time series had several episodes of exuberance occurring very close to each other in time. Philips et. al. (2013) proposed a series

recursive right-tailed ADF test, such as the RADF, SADG and GSADF tests, to address the above issue. With these three advanced econometric techniques, we are able to identify the existence of multiple exuberant behaviors over a short time series.

In this chapter, I adopted the aforementioned three extended right-tailed ADF tests to examine the exuberance of a long weekly Chinese stock market from 2005 to 2013 and the GEM stock market from 2011 to 2013. I successfully detected explosive behaviors (bubbles) and identified the dates of their rising and bursting in the Chinese A share market. However, there was insufficient evidence to show the existence of any exuberance when I applied the same methodology to the GEM stock market.

To summarize, this chapter not only theoretically defines exuberant behaviors in the Chinese stock market, but also practically tests whether any exuberance has previously occurred there. Furthermore, making use of three advanced time series models, I am able to provide empirical evidence for the exuberance hypothesis for the Chinese stock market, as well as practical applications of those newly developed models. Ultimately, this chapter also compares the performances of those three tests during empirical studies. We can conclude that the results of the three tests are quite robust if the sample size if large enough. In particular, the GSADF test does not perform well if the sample size is small.

Finally, it is worth mentioning that there are still many factors that have not yet been taken into consideration. First, the data I use in this chapter could be expanded to include current data. And we could try running the tests on monthly instead of weekly data to avoid the volatility issue. Second, in addition to the extended right-tailed ADF

tests used here, there are many other tests for multiple explosive behaviors in a time series that could be employed. For example, the CUSUM monitoring procedure (Homm and Breitung, 2012) could also be used to detect the exuberances, as well as date-stamping the start and collapse of each exuberant behavior. We could adopt all other econometric tests to compare those tests' performances in terms of empirical results. All these areas remain for future research.

CHAPTER THREE

New Thoughts on "Power Couples": Does the Co-location Problem Still Exist

1 Introduction

This chapter investigates trends of power couples' concentration in large metropolitan areas of the United States between 1940 and 2010. "Power couple" refers to a couple of which husband and wife were both at least college educated[31]. From 1940 to 1990, power couples were increasingly and disproportionately located in large metropolitan areas. In 1940, 37 percent of power couples were located in metropolitan areas of at least 2 million population. By 1970 this number was 40 percent, by 1980 44 percent, and by 1990 49 percent. Meanwhile, there had been little change in the proportion of couples in which at least one spouse did not have a college education in large cities. However, between 1990 and 2010 the percentage of power couples in large metropolitan areas remained quite flat.

Costa and Kahn (2000) argued that college-educated couples were increasingly located in large metropolitan areas primarily due to the growth of dual- career households and the resulting severity of the co-location problem.[32] Their main argument stated that, as more households become dual-career households, more of them would face a co-location problem. All dual-career households are more likely to be joint decision-makers,

[31] College educated refers people has at least a college degree. There are different definition in different census year, you could go to section 4 for details.
[32] The co-location problem refers to the preference of a couple to stay in the same location during their job searches.

and they face the difficulty of finding two jobs commensurate with the skills of each spouse within a reasonable commute from home.

They found that co-location is the most likely explanation for the observed trend in location choice. Their paper documented trends in location choice since 1940 between large, mid-size, and small cities by household type. They also examined whether the long-term trends are most consistent with a worsening of the co-location problem or with the growing urbanization of the college educated.

However after 1990, the statistics had moved in a different direction. By 2000, the percentage of power couples located in large cities dropped to 46, and increased only slightly to 48 by 2010. In contrast, more and more power couples tended to live in mid-size metropolitan areas and small cities. By 2010, nearly 31 percent of power couples resided in mid-size metropolitan areas. Nevertheless, more households were becoming dual-career households. By 2010, the wife worked in 79.43 percent of all power couples. As noted previously, the partners in these dual-career households are more likely to be joint decision-makers; this increases the challenge of finding two jobs commensurate with the skills of each partner within a reasonable commuting distance/time from their home. It is this co-location problem that should lead to the greater concentration of power couples in large cities, where the supply and variety of jobs is greatest.

We may wonder why relatively fewer college-educated couples chose to stay in large metropolitan areas after 1990. Is it because the co-location effect faded away after the 1990s? Or is it because the mid-size metropolitan areas also began to provide sufficient opportunities for dual-career households after the technology boom? Or is it

because dual-career couples assigned co-location a lower priority when conducting a job search after 1990, since it is presently much easier for long-distance couples to commute and communicate?

As "power couple" become a more popular phenomenon, the concept of "power wife,"[33] a group with its own specific characteristics, has attracted much attention and debate from economists recently. Prime-aged college-educated women in 1940 generally became school teachers upon graduation, subsequently leaving the labor force because of marriage bars (Costa and Kahn, 2000). Goldin (1997) describes their experience as "first jobs then family." In contrast, the experience of their 1970s counterparts was "first family, then jobs." The majority majored in such fields as education and nursing regardless of their majors. And very few men were interested in those majors. They left the labor force when their first child was born, and only reentered when all children were in school. By 1990, and to a lesser extent by 1980, college-educated women aspired to "career, then family" or "career and family" (Goldin 1997). Their college majors were similar to men's, and in terms of labor supply parameters they began to resemble men as well, with small wage and income elasticity (Goldin 1990).

Therefore, in this case, we could expect that the co-location problem might not be an important factor for a power woman or power wife when they make career decisions. In this chapter, I focus on women whose age is from 23 to 37. Since in some sense these power women care about their career so much more than did women in earlier decades, they might not mind being separated by a long distance from some family members after

[33] Power wife refers to the wife in the household who has at least a college degree.

1990. Thus, the movement of power couples toward large metropolitan areas might only be in response to urbanization power and large labor market power. In other words, the co-location effect becomes less important.

This chapter documents trends in locational choice from 1940 to 2010 among large, mid-size, and small cities by household types. I adopt the concept "coincidental couple" from Costa and Kahn (2000) to rule out all the other reasons for why relatively fewer power couples tended to live in large metropolitan areas after 1990. I test whether the primary reason for urbanization of power couples is the co-location problem or not, and further analyze why the urbanization trend gradually disappeared after 1990. In the end, I successfully prove that the co-location effect faded away after 1990. My findings show that power couples were no longer increasingly located in large cities because the co-location effect became less important. The other contribution of this chapter are new definition and calculation methods for the "coincidental couple," which refers to two individuals (one male and one female) coincidentally living in a large metropolitan area. This concept is the crucial part of the empirical method, which is discussed in detail in section 3. Section 2 presents the general trends in locational choice from 1940 to 2010 between large, mid-size, and small cities by household types. Section 3 illustrates the econometric models used to estimate the probability of residence in a certain city size category, and the triple difference in difference model to estimate the co-location effect. Section 4 discusses the data collection procedure and variables used in empirical models. Following in section 5 are the results and an explanation of trends based on empirical

evidence. This chapter presents conclusions in section 6 and some relevant future research questions.

2 Trends

I categorize all households as one of five types: "power" couples in which both spouses have a college education, "part-power" couples in which only one spouse has a college education, "low-power" couples in which neither spouse has a college education, and single households of the college educated and the non-college educated. I define two individuals to be a couple if their marital status is married and both spouses are present. Singles can be never married, divorced, or widowed. Both couples and singles may be in multifamily households. This research only focuses on couples in which the husband was 25 to 39 years of age and the wife 23 to 37. This age restriction allows us to examine couples in the early stages of their careers and also allows us to create a comparable group of singles (Costa and Kahn, 2000). I impose the same age restrictions on singles.

According to Table 3.1, the proportion of couples in which both husband and wife have at least a college education is increasing continuously. From 1940 to 1990, the percentage of college-educated couples grew from 2.25 to 14.68; however this trend became even more salient after 1990. By 2000, 19.18 percent of all couples were those in which both husband and wife were at least college graduates, and by 2010, 29.24% were.

Table 3.1: Percentage of Marriages by Couple Type

	1940	1960	1970	1980	1990	2000	2010(5)
Low-power	91.14	82.8	76.7	67.62	67.15	61.87	48.64
Part-power	6.62	12.15	14.63	18.44	18.18	18.95	22.13
Power	2.25	5.04	8.67	13.94	14.68	19.18	29.24

Note: A power couple is defined as one in which both husband and wife are college graduates, a part-power couple as one in which only one spouse is a college graduate, and a low-power couple as one in which neither spouse is a college graduate. All numbers are estimated from the census samples [IPUMS] and are for households in which the husband was age 25 to 39 and the wife 23 to 37.

If we examine Table 3.2, we can see that wives' rising labor force participation rates have transformed couples into true dual-career households and increased the share of couples with a co-location problem. The labor force participation rate of power-couple wives rose from 19 to 76 percent between 1940 and 1990, whereas the increase for low-power wives was from 16 to 69.57 percent. By 2010, the labor participation rate of power-couple wives had risen to 79.43 percent. As we can see, the growth of the labor participation rate gradually slows after 1990. There was even a small decrease of the power- wife working percentage in 2000. In addition, there is another interesting point; the percentage of full-time working wives among all working wives had been quite steady from 1940 to 1990, no matter the wives' level of education level. However, there was a big increase of full-time working wives after 1990, which indicated that more and more wives chose to work more hours after 1990. This might have occurred because of the households' need for greater financial support, or because women now would like to fulfill their career achievements rather than being housewives or part-time workers. By 2010, 83.80 percent of working wives in power couples chose to work full time, which

also suggested the concept mentioned earlier in section 1: college-educated women now aspire to "career and family" (Goldin, 1997).

Next, let us move to the most important trend: locational choice. We classify trends in household-location choices conditional on marital status and on the educational levels of both spouses. That is, for every year, we estimate:

Prob (lives in big city | household type).

I classify the suburbs of central cities as part of the labor market of the central city and define three city-size categories: large metropolitan areas[34] (those with populations of at least 2 million), mid-size metropolitan areas (those with populations of between 2 million and 250,000), and small and nonmetropolitan areas (metropolitan areas with populations of less than 250,000 and nonmetropolitan areas). I do not use the 1950 census because education is known only for the sample line person, and I do not use 1960 census because metropolitan area is not identified

I document trends from 1940 to 2010, and I emphasize the 1970 to 2010 trends in Table 3.3. The table illustrates trends in location choice among large and mid-size metropolitan areas and small localities. We notice that the probability of a power couple residing in a large metropolitan area increases continuously from 1970 to 1990. However, after 1990 the probability of the locational choice of large city became quite steady for every household type. However, there was a big jump for power couples located in mid-size metropolitan areas from 2000 to 2010. By 2010, 31 percent of the power couples chose to live in mid-size cities; in contrast only 20 percent did so in 2000.

[34] A detailed definition of metropolitan sizes is presented in section 4.

Table 3.2: Employment and Fertility Trends by Education of Couple

	1940	1960	1970	1980	1990	2000	2010(5)[35]
Wife works(%)							
Low-power	16.09	28.35	38.77	56.47	69.57	67.38	74.84
Part-power	17.21	23.12	35.81	59.92	73.02	74.17	80.83
Power	18.73	29.68	45.00	67.10	76.06	73.73	79.43
Have child(%)							
Low-power	76.16	90.46	90.62	86.33	85.13	86.01	85.80
Part-power	62.91	87.08	82.85	75.06	73.99	72.50	72.75
Power	59.97	81.67	72.94	64.76	64.32	63.00	64.76
Wife works and works full time (%)							
Low-power	69.68	63.32	60.96	62.54	64.97	73.93	78.48
Part-power	72.16	63.75	58.38	64.06	67.87	76.60	81.99
Power	66.57	61.28	55.67	67.32	69.90	78.70	83.80

Note: A full-time job is defined as 35 hours or more per week. A traditionally female occupation is defined as one in which women were overrepresented relative to men in 1970; that is, one in which more than 50 percent of all employees age 18 to 64 were women in 1970. All couples are restricted to those in which the husband was 25 to 39 years of age and the wife 23 to 37. All numbers are estimated from the integrated public use census samples [IPUMS]. A power couple is defined as one in which both husband and wife are college graduates, a part-power couple as one in which only one spouse is a college graduate, and a low-power couple as one in which neither spouse is a college graduate.

[35] 2010(5) here means that for 2010 I use five years of census data. Please refer to section 4 for details.

Table 3.3: Probability of Locational Choice by Household Type

	1940	1970	1980	1990	2000	2010(5)[36]
Conditional on power couple						
Large metropolitan area	0.37	0.41	0.44	0.49	0.45	0.48
Mid-size metropolitan area	0.27	0.30	0.31	0.28	0.19	0.31
Small and non-metropolitan area	0.37	0.29	0.25	0.23	0.37	0.20
Conditional on part-power couple						
Large metropolitan area	0.36	0.38	0.40	0.39	0.36	0.36
Mid-size metropolitan area	0.27	0.31	0.32	0.30	0.18	0.33
Small and non-metropolitan area	0.37	0.31	0.28	0.31	0.47	0.32
Conditional on low-power couple						
Large metropolitan area	0.26	0.31	0.33	0.30	0.29	0.30
Mid-size metropolitan area	0.23	0.29	0.30	0.28	0.16	0.30
Small and non-metropolitan area	0.50	0.41	0.37	0.42	0.55	0.40
Conditional on single, power man						
Large metropolitan area	0.44	0.53	0.54	0.53	0.52	0.54
Mid-size metropolitan area	0.28	0.26	0.28	0.28	0.19	0.30
Small and non-metropolitan area	0.28	0.21	0.18	0.19	0.29	0.17
Conditional on single, power women						
Large metropolitan area	0.37	0.51	0.52	0.51	0.50	0.52
Mid-size metropolitan area	0.25	0.28	0.30	0.30	0.20	0.32
Small and non-metropolitan area	0.38	0.21	0.18	0.18	0.30	0.17
Conditional on single, low-power man						
Large metropolitan area	0.32	0.43	0.44	0.36	0.32	0.32
Mid-size metropolitan area	0.24	0.27	0.30	0.29	0.18	0.30
Small and non-metropolitan area	0.45	0.30	0.27	0.35	0.50	0.38
Conditional on single, low-power woman						
Large metropolitan area	0.33	0.43	0.45	0.38	0.34	0.35
Mid-size metropolitan area	0.30	0.29	0.31	0.31	0.20	0.33
Small and non-metropolitan area	0.38	0.28	0.24	0.32	0.46	0.31

Note: A full-time job is defined as 35 hours or more per week. A traditionally female occupation is defined as one in which women were overrepresented relative to men in 1970; that is, one in which more than 50 percent of all employees age 18 to 64 were women in 1970. All couples are restricted to those in which the husband was 25 to 39 years of age and the wife 23 to 37. All numbers are estimated from the integrated public use census samples [[PUMS]. A power couple is defined as one in which both husband and wife are college graduates, a part-power couple as one in which only one spouse is a college graduate, and a low-power couple as one in which neither spouse is a college graduate

[36] 2010(5) here means that for 2010 I use five years of census data. Please refer to section 4 for details.

3 Methods

This section adopts the empirical method in Costa and Kahn's paper (2000) to identify the co-location problem between 1940 and 1990 and estimate the co-location effect after 1990. We begin by considering the benefits of residing in a large metropolitan area for different couples within a short period.

First, we discuss the benefits of urbanization by couple type. For power couples, they would prefer locating in large cities if large cities offer higher returns to education and if urban amenities are normal goods. However, these features of large cities would also appeal to all other college-educated people, even though he/she is not married (or is married to a non-college educated spouse). Meanwhile large cities usually could also provide amenities such as marriage markets; therefore, singles will be more attracted to locating in large cities than even married power couples.

We found that most of the increasing urbanization of the college-educated population is explained by increasing returns to city size and not because of urban amenities are normal goods (Costa and Kahn, 2000). We cannot directly test whether large cities attract more college-educated singles because they offer better marriage markets for the college educated. But we can provide some evidence of the marriage propensities of the college educated relative to the non-college educated. More specifically, I adopt the concept of “coincidental couple,” introduced by Costa and Kahn (2000), to illustrate some empirical evidence on the marriage propensities of power singles compared to low-power singles. Details of the coincidental couple are provided later in this section.

Second, we should focus on the co-location problem. It is much easier for dual-career power couples to find jobs in a large city than in other city-size categories. The diversified labor markets of larger cities also insure against health and unemployment shocks to households regardless of their educational levels. Power couples can always take jobs for which they are overqualified, and therefore the co-location problem may be less severe for these couples than for others. On the other hand, the power couples might be more career oriented than low-power couples, so that the co-location problem might be more severe for them in some cases. According to Topel and Ward (1992), large cities offer more potential job matches, so the probability of drawing a good initial match is higher. The probability of drawing a good subsequent match is also higher, and this increased job mobility will lead to greater lifetime wage growth.

As Costa and Kahn (2000) did in their paper, I have tested the extent to which the increase in the proportion of power couples in large cities is caused by co-location by comparing power couples with other couple types and singles. These comparisons are based on "coincidental couples," which refer to two individuals (one male and one female) coincidentally living in a large metropolitan area. According to Costa and Kahn (2000), we could estimate the probability of a single power man living in city size $s(p_s^{M,P})$ and the probability of a single power woman living in difference city sizes $(p_s^{F,P})$, and then take the minimum of these probabilities $(\min(p_s^{M,P}, p_s^{F,P}))$. Finally, we estimate the probability that a "coincidental couple" will be living in a given city size $(\min(p_s^{M,P}, p_s^{F,P}))/\Sigma_s \min(p_s^{M,P}, p_s^{F,P})$. For example, suppose that there are 100 single power men and women each, and that 40 of the men are in large cities and 60 are in small

cities and that 60 of the women are in large cities and 40 are in small cities. At most 80 "coincidental marriages" could form—40 in large cities and 40 in small cities. The probability of a coincidental couple being in a large city is therefore 40/(40+40)=0.5.

However, there are two reasons that will make this estimate of coincidental couples unreliable. First, singles in large cities presumably have a higher probability to form couples because large cities have better marriage markets, and therefore more couples are formed in large cities even with the same numbers of males and females in large cities and small cities. Second, power singles will not necessarily marry power singles, which makes the calculation of coincidental power couples meaningless. This is also true for other types of coincidental couples. More explicitly, we can say that 80 marriages could form, but those 80 marriages are not necessarily power couples, since a power man might not marry a power woman.

Therefore in order to estimate the co-location effect of the location choice, I propose a more accurate way to estimate the probability of coincidental couples given metropolitan size. First I assign a weight $w_{s,p}$ for each minimum probability ($\min(p_s^{M,P}, p_s^{F,P})$) by city size. And for each year and each city type, I define this weight as: (total number of this couple type/min(total number of this single male type, total number of this single female type). More formally,

$$w_{s,p} = \min\{1, N(couple_{s,p})/\min[\, N(single\ male_{s,p}), N(single\ female_{s,p})]\} \quad (3.3.1)$$

And then we estimate the probability that a "coincidental couple" will be living in a given city size as:

$$p = (w_{s,p} * \min(p_s^{M,P}, p_s^{F,P}))/\Sigma_s w_{s,p} * \min(p_s^{M,P}, p_s^{F,P}) \quad (3.3.2)$$

The following is a table from Costa and Kahn (2000), which provides a schematic illustration of how to identify the co-location problem of power and low-power couples and the differential co-location problem of power relative to low-power couples.

Table 3.4 Change in Benefits of Living a Large City by Couple Type

Couple type	Benefits of living in a large city
1 Power couple	co-location power
	urbanization power
2 Coincidental power couple	urbanization power
	singles' amenities
3 Double difference(1-2)	co-location power, singles' amenities
4 Lower-power couple	co-location low-power
	urbanization low-power
5 Coincidental low-power couple	urbanization low-power
	singles' amenities
6 Double difference (4-5)	co-location low-power, singles' amenities
7 Triple difference (3-6)	co-location power relative low-power

Note: A power couple is defined as one in which both husband and wife are college graduates, and a low-power couple as one in which neither spouse is a college graduate. A coincidental power couple consists of two single college-educated individuals (one male and one female) coincidentally living in the same city size. A coincidental low-power couple consists of two single non-college educated individuals (one male and one female) coincidentally living in the same city size.

If the value of singles' amenities has not changed and if urban amenities are of the same value to singles as they are to married couples, then the double difference (the increase in power couple concentration minus the increase in coincidental couples' concentration) measures the differential effect of colocation. However, any differences in the valuation of urban amenities of married couples and singles will not be differenced out. Therefore, an alternative explanation for the increase in power couple relative to

power single concentration in large metropolitan areas is the increased amenity value of metropolitan areas to couples.

A similar situation is observed for the low-power couple case. Again, differences in the valuation of urban amenities to married couples and to singles will not be differenced out. The triple difference yields the differential effect of co-location for power couples relative to low-power couples, assuming either that there is no difference in the valuation of urban amenities between married couples and singles or, if there are, that these differences should be quite similar for the college educated as well as for the non-college educated. Therefore, the triple difference estimates the differential effect of co-location for the power couple compared to low-power couple. And I also compare power/low-power couples in which the wife works with coincidental couples in which the woman works, and then I calculate the differences. Larger double and triple differences for couples in which the wife works for all couples would suggest that co-location, not differences in amenity values, determines the concentration of power couples in large metropolitan areas.

I first demonstrate that the co-location problem is the main factor causing the urbanization of power couples from 1940 to 1990, by adopting this “coincidental couple” concept. Similarly, using the same triple difference model, I can also validate that the primary reason for power couples no longer further urbanizing after 1990 is that the co-location effect faded away. This requires an examination of differential trends in locational choice by couple type. I present these trends in the following section, standardized by race and age. For couples, I estimated:

Prob (lives in city size s and wife works | household type);

Prob (lives in city size s and wife does not work | household type).

For singles, I estimated:

Prob (lives in city size s and works | household type);

Prob (lives in city size s and does not work | household type).

I estimate these probabilities using a multinomial logit choice model. First I collect information on each city size category a couple chose to live in and whether the wife works or does not work, to create six groups. In other words, for each group, we are interested in the wife's labor force participation status and city size. I then estimate a multinomial logit of the choice of the wife's labor force participation and city size as a function of the husband's age and race and the educational level of the husband and wife (dummies for less than high school graduate, high school graduate, college graduate, and post-college graduate, with less than high school graduate as the omitted dummy). Therefore, I define $P_{s,ww}$ as the probability of being in one of three city sizes s and wife's labor participation choice ww, i.e., $P_{s,ww}$ equals 1 if wife works and lives in city size s. More formally, I estimated:

$$log\left(\frac{P_{s,ww}}{P_{s=0,ww=1}}\right) = \beta' X \tag{3.3.3}$$

for s=1, 2 and ww=0,1 and s=0, ww=0. I use s=0, ww=1 as the base group.

Finally, I predict the choice of location and wife's labor force participation, $P_{s,ww}$, for a white household in which the husband is 35 years of age and the wife is 33, conditional on being a power, part-power, or low-power couple. For single individuals I estimate similar multinomial logit specifications (separately for men and women), except

that labor force status is own labor force status and I control only for own characteristics. I also predict locational choice for single, white men age 35 and single, white women age 33 conditional on being a power or a low-power individual and estimate locational choice for coincidental couples.[37]

4 Data

In this chapter, I use the 1940 and 1970-2010 censuses of population and housing (1940-1990 1% sample, 2000 and 2010 5yrs are ACS sample), which can be downloaded from IPUMS (www.ipums.org). For each person I obtain the following variables: marital status, age, sex, race, education, labor force status, occupation, hourly wage, number of children, and metropolitan area. The couples are all married and living in the same household and the singles are either never married, divorced, or widowed. I deleted the observations of couples for which spouse is indicated as absent (The share of this category is relatively small, and thus will not affect the results.) In this chapter, I still treat the cohabitation couples as couples, since cohabitation case had only happened in 1990.

Based on the raw data described above, I construct the following variables:

(1) Education level

The definition of education differs across census years. For 1940, 1970, and 1980, I use the highest grade of school or year of college completed. Education is overstated in 1940

[37] The base group is small for nonmetropolitan area and wife out of the labor force. Our results are robust to the choice of base group.

(Goldin, 1997).[38]. For 1990, 2000, and the 2010 five-year sample, the education variable gives the respondent's highest grade of school completed through the 11th grade, but classifies high school graduates according to their highest diploma or degree earned. I therefore define the categories of less than high school as grade 11 or less; high school as grade 12 in 1940, 1970, and 1980, and as 12th grade, high school diploma, or GED in 1990; college as four or more years in 1940, 1970, and 1980 and as bachelors or graduate degree in 1990, 2000, and 2010 five years. I classify those who did not complete college (less than 4 years in 1940, 1970, and 1980 and some college but no degree or occupational/academic associate degree in 1990, 2000, and 2010 five years) together with the high school graduates.

(2) Metropolitan area size classifications

The boundaries of metropolitan areas have grown throughout the census years, and new metropolitan areas have emerged. Therefore, it is very important to define metropolitan area size classifications that allow for comparability across all census years.

I introduce three city size categories: large metropolitan areas (those with populations of at least 2 million), mid-size metropolitan areas (those with populations of between 2 million and 250,000), and small and non-metropolitan areas (metropolitan areas with a population of less than 250,000 and non-metropolitan areas). The 1940 census identified metropolitan areas if the population in these areas was at least 100,000 in 1980, and the 1980 and 1990 censuses identified metropolitan areas with populations of at least 100,000 in the census year. The 1970 census identified metropolitan areas with

[38] College graduation rates are overstated in the 1940 census, in part because individuals who entered into the preparatory department within a college were enumerated as having gone to college (Goldin, 1997).

populations of at least 250,000 in 1970. My definition of small and nonmetropolitan areas is therefore consistent across time.

Table 3.5 lists the consistent definition of large metropolitan size across most census years. I use the same classification for the 2010 five-year sample as well. For 1940, the large size metropolitan areas are: Chicago, New York, Boston, Cleveland, Detroit, Los Angeles, Philadelphia, Pittsburgh, St. Louis, and San Francisco.

Table 3.5: Large MSAs (population over 2M)

	1970	1980	1990	2000
New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	18,071,522	17,412,203	17,953,372	19,451,757
Los Angeles-Anaheim-Riverside, CA	9,980,859	11,497,549	14,531,529	16,036,587
Chicago-Gary-Lake County, IL-IN-WI	7,778,948	7,937,290	8,065,633	8,783,199
San Francisco-Oakland-San Jose, CA	4,754,366	5,367,900	6,253,311	6,873,645
Philadelphia-Wilmington-Trenton, PA-NJ-DE-MD	5,749,093	5,680,509	5,899,345	5,661,399
Detroit-Ann Arbor, MI	4,788,369	4,752,764	4,655,236	5,031,963
Washington, DC-MD-VA-WV	3,040,307	3,250,921	3,923,574	4,739,999
Dalls-Fort Worth, TX	2,351,568	2,930,568	3,885,415	4,909,523
Boston-Lawrence-Salem-Lowell-Brocton, MA-NH-ME-CT	3,709,642	3,662,888	3,783,817	4,440,881
Houston-Galveston-Brazoria, TX	2,169,128	3,099,942	3,711,043	4,493,741
Miami-Fort Lauderdale, FL	1,887,892	2,643,766	3,192,582	3,711,102
Atlanta, GA	1,684,200	2,138,136	2,833,511	3,857,097
Cleveland-Akron-Lorain, OH	2,999,811	2,834,062	2,759,823	2,910,616
Seattle-Tacoma, WA	1,836,949	2,093,285	2,559,164	3,023,741
San Diego, CA	1,357,854	1,861,846	2,498,016	2,820,844
Minneapolis-St. Paul, MH-WI	1,981,951	2,137,133	2,464,124	2,872,109
St. Louis, MO-IL	2,429,376	2,376,968	2,444,099	2,569,029
Baltimore, MD	2,089,438	2,199,497	2,382,172	2,491,254
Pittsburgh-Beaver Valley, PA	2,556,029	2,423,311	2,242,798	2,331,336
Phoenix, AZ	971,228	1,509,175	2,121,101	3,013,696
Tampa-St. Petersburg-Clearwater, FL	1,105,553	1,613,600	2,067,959	2,278,169
Denver-Boulder, CO	1,238,273	1,618,461	1,848,319	2,252,103

Note: Source is from: Compton and Pollak (2006). MSAs are defined as ”large” if their population is greater than 2 million in 1990, The analysis here uses the MSA definitions (i.e., county components) from the 1990 definitions.

5 Results

The tables in this section present the predicted results of the regressions. Table 3.6 shows the predicted probabilities, conditional on being a power, part-power, or low-power couple, of locational choice across different city-size categories and the wife's labor force participation status for a white couple in which the husband was 35 years old and the wife 33. Since 1940, power couples have been leaving nonmetropolitan areas and moving to the largest metropolitan areas. In 1940, 34 percent of such couples were in nonmetropolitan areas, whereas by 1970 the figure had dropped to 29 percent and by 1990 to 23 percent. However, it increased to 38 percent in 2000 and fell to 22 percent by 2010.

In 1940, 39 percent of power couples resided in the largest metropolitan areas, and in 1970 40 percent did, whereas in 1990 48 percent did. However, this trend did not continue, as by 2010, only 46 percent were found in the large metropolitan areas.

Table 3.7 presents predictions of locational choice from multinomial logit models for singles conditional on their education. Note that between 1940 and 1970 the probability of singles residing in a large metropolitan area increased substantially, rising from 43 to 50 percent among power women. Between 1970 and 2010 changes were much more modest. The probability of low-power women being in a large city decreased somewhat, with a decline from 39 to 35 percent among coincidental couples between 1970 and 1990, and it keeps declining after 1990; by 2000 the probability is only 33 percent.

Table 3.6: Predicted Probabilities of Locational Choice and Wife's Labor Force Participation (LFP) Status Conditional on Household Type

	1940	1970	1980	1990	2000	2010(5)
Conditional on power						
Large metropolitan area, LFP=1	0.086	0.152	0.284	0.353	0.317	0.363
Large metropolitan area, LFP=0	0.306	0.247	0.157	0.126	0.121	0.100
Mid-size metropolitan area, LFP=1	0.042	0.125	0.200	0.215	0.136	0.249
Mid-size metropolitan area, LFP=0	0.225	0.182	0.110	0.072	0.047	0.067
Small, nonmetropolitan area, LFP=1	0.055	0.143	0.175	0.185	0.292	0.182
Small, nonmetropolitan area, LFP=0	0.286	0.151	0.074	0.049	0.087	0.039
Conditional on part-power						
Large metropolitan area, LFP=1	0.065	0.095	0.212	0.257	0.231	0.273
Large metropolitan area, LFP=0	0.311	0.281	0.186	0.125	0.105	0.084
Mid-size metropolitan area, LFP=1	0.043	0.088	0.172	0.208	0.127	0.246
Mid-size metropolitan area, LFP=0	0.235	0.218	0.142	0.087	0.048	0.073
Small, nonmetropolitan area, LFP=1	0.046	0.103	0.166	0.236	0.366	0.261
Small, nonmetropolitan area, LFP=0	0.300	0.215	0.122	0.087	0.123	0.063
Conditional on low-power						
Large metropolitan area, LFP=1	0.048	0.099	0.171	0.191	0.168	0.201
Large metropolitan area, LFP=0	0.242	0.202	0.202	0.093	0.092	0.074
Mid-size metropolitan area, LFP=1	0.044	0.106	0.126	0.195	0.110	0.223
Mid-size metropolitan area, LFP=0	0.196	0.180	0.168	0.080	0.047	0.073
Small, nonmetropolitan area, LFP=1	0.057	0.163	0.155	0.309	0.417	0.331
Small, nonmetropolitan area, LFP=0	0.413	0.250	0.178	0.132	0.166	0.098

Note: All predictions are from a multinomial logit model in which the outcome variables were city size and labor force participation rate. The independent variables were husband's age, age squared, and race; wife's age and age squared; and dummy variables for educational levels (less than high school, high school, college, college plus) of the husband and the wife. The predictions are for a white couple in which the husband was 35 years old and the wife 33. Robust standard errors ranged from 0.001 to 0.010. Conditional on couple type, the rows should sum to one. All probabilities were estimated from the samples [IPUMS]

Table 3.7: Predicted Probabilities of Locational Choice, Unmarried Men and Women, Conditional on Education

	1940	1970	1980	1990	2000	2010(5)
Single, power man						
Large metropolitan area	0.484	0.539	0.546	0.501	0.474	0.492
Mid-size metropolitan area	0.264	0.255	0.277	0.286	0.190	0.295
Small and non-metropolitan area	0.252	0.206	0.177	0.213	0.336	0.213
Single, power woman						
Large metropolitan area	0.429	0.499	0.542	0.499	0.471	0.492
Mid-size metropolitan area	0.269	0.297	0.281	0.300	0.190	0.311
Small and non-metropolitan area	0.302	0.204	0.167	0.201	0.339	0.197
Coincidental power couple						
Large metropolitan area	0.454	0.521	0.550	0.506	0.472	0.500
Mid-size metropolitan area	0.279	0.266	0.281	0.290	0.191	0.300
Small and non-metropolitan area	0.267	0.213	0.169	0.204	0.337	0.200
Single, low-power man						
Large metropolitan area	0.334	0.400	0.432	0.351	0.303	0.319
Mid-size metropolitan area	0.229	0.261	0.287	0.282	0.178	0.293
Small and non-metropolitan area	0.437	0.309	0.271	0.367	0.519	0.388
Single, low-power woman						
Large metropolitan area	0.346	0.387	0.450	0.361	0.328	0.345
Mid-size metropolitan area	0.281	0.293	0.305	0.301	0.187	0.325
Small and non-metropolitan area	0.373	0.320	0.245	0.338	0.485	0.330
Coincidental low-power couple						
Large metropolitan area	0.357	0.404	0.448	0.362	0.314	0.339
Mid-size metropolitan area	0.245	0.273	0.298	0.290	0.184	0.311
Small and non-metropolitan area	0.398	0.323	0.254	0.348	0.502	0.350

Note: All predictions are from a multinomial logit model and are for white 35 years old white men and white 33 years old white women. Within each year for each group the predicted probabilities for singles and coincidental couples should sum to one. The independent variables were age, age squared, race, and educational level. The outcome variables were city size and labor force participation. With the exception of 1940, relatively few individuals were out of the labor force. The results that are presented are the predicted probabilities summed over city size. Robust standard errors ranged from 0.001 to 0.010. All probabilities were estimated from the samples [IPUMS].

Table 3.8 presents the predicted probabilities, conditional on being a power or low-power single women, of locational choice across city sizes and labor force participation status for a white woman at 33 years old. It also presents predictions of locational choice from multinomial logit models for coincidental couples where the wife works. If we compare the locational choices of coincidental couples in Table 3.7 with

those in Table 3.8, the probabilities of coincidental power couples residing in each city size category are almost identical to the probabilities of coincidental power couples in which the wife works being in the respondent city size category in 2010. By contrast, those probabilities were quite different in 1940. And we also notice that the probability of working power women residing in the large metropolitan area did not change too much from 1940 to 2010.

Table 3.8: Predicted Probabilities of Locational Choice for Working Women Only

	1970	1990	2000	2010(5)
Single, power woman				
Large metropolitan area, LFP=1	0.474	0.483	0.443	0.472
Large metropolitan area, LFP=0	0.025	0.015	0.029	0.020
Mid-size metropolitan area, LFP=1	0.277	0.290	0.178	0.295
Mid-size metropolitan area, LFP=0	0.019	0.010	0.012	0.015
Small and non-metropolitan area, LFP=1	0.191	0.191	0.316	0.184
Small and non-metropolitan area, LFP=0	0.013	0.011	0.022	0.014
Single, low-power woman				
Large metropolitan area, LFP=1	0.294	0.301	0.262	0.290
Large metropolitan area, LFP=0	0.094	0.060	0.066	0.055
Mid-size metropolitan area, LFP=1	0.228	0.250	0.152	0.270
Mid-size metropolitan area, LFP=0	0.065	0.051	0.034	0.055
Small and non-metropolitan area, LFP=1	0.230	0.272	0.389	0.264
Small and non-metropolitan area, LFP=0	0.090	0.066	0.096	0.066
Coincidental power, woman works				
Large metropolitan area	0.515	0.503	0.473	0.496
Mid-size metropolitan area	0.277	0.298	0.190	0.310
Small and non-metropolitan area	0.208	0.199	0.337	0.194
Coincidental low-power, woman works				
Large metropolitan area	0.391	0.366	0.326	0.352
Mid-size metropolitan area	0.303	0.304	0.189	0.328
Small and non-metropolitan area	0.306	0.330	0.485	0.320

Note: All predictions are from a multinomial logit model and are for 35 year old white men and 33 year old white women. Within each year for each group the predicted probabilities for singles and coincidental couples should sum to one. The independent variables were age, age squared, race, and educational level. The outcome variables were city size and labor force participation. With the exception of 1940, relatively few individuals were out of the labor force. The results that are presented are the predicted probabilities summed over city size. Robust standard errors ranged from 0.001 to 0.010. All probabilities were estimated from the samples [IPUMS].

Table 3.9 summarizes standardized trends in location choice by couple types, including coincidental couples for the period 1970-1990. Between 1970 and 1990 the probability of power couples being in a large metropolitan area rose by 0.080, whereas that of part-power, low-power, and coincidental power couples being in a large city rose by 0.006, -0.017, and -0.015, respectively. Among power couples in which the wife works, the probability of being in a large city rose by 0.201, whereas among power couples in which the wife does not work, this probability fell by 0.121.

Table 3.9: Trends in Propensity to Live in Given City Size, 1970-1990, by Couple Type (based on predicted probabilities)

	City size		
	Large	Mid-size	Small
Differences, 1990-1970			
Power couples (Δ^{P})	0.080**	-0.020**	-0.060**
Part-power couples (Δ^{pp})	0.006**	-0.011**	0.005**
Low-power couples (Δ^{LP})	-0.017***	-0.011**	0.028**
Coincidental power couples (Δ^{CP})	-0.015**	0.024**	-0.009**
Coincidental low-power couples (Δ^{CLP})	-0.420**	0.017**	0.025**
Power, wife works ($\Delta^{P,W}$)	0.201**	0.09**	0.042**
Part-power, wife works ($\Delta^{PP,W}$)	0.162**	0.120**	0.133**
Low-power, wife works ($\Delta^{LP,W}$)	0.092**	0.089**	0.146**
Power, wife does not works ($\Delta^{P,NW}$)	-0.121**	-0.110**	-0.102**
Part-power, wife does not works ($\Delta^{PP,NW}$)	-0.156**	-0.131**	-0.128**
Low-power, wife does not works ($\Delta^{LP,NW}$)	-0.109**	-0.10**	-0.118**
Coincidental power, wife works ($\Delta^{CP,W}$)	-0.012**	0.021**	-0.009**
Coincidental low-power, wife works ($\Delta^{CLP,W}$)	-0.025**	0.001**	0.024**

Note: Differences are in probability units. Probabilities are calculated from Tables 3.6, 3.7, and 3.8, except for coincidental couples in which the woman works. These were calculated by using the multinomial logit predictors for working women only. *=p<0.05, **=p<0.01, ***=p<0.001

And Table 3.10 summarizes standardized trends in location choice by couple types, including coincidental couples for the period 1990-2010. Between 1990 and 2010 the probability of power couples being in a large metropolitan area fell by 0.016, whereas

that of part-power, low-power, and coincidental power couples being in a large city changed by -0.025, -0.009, and -0.006, respectively. Among power couples in which the wife works, the probability of being in a large city rose by 0.010, whereas among power couples in which the wife does not work, this probability fell by 0.026.

Table 3.10: Trends in Propensity to Live in Given City Size, 1990-2010, by Couple Type (based on predicted probabilities)

	City size		
	Large	Mid-size	Small
Differences, 2010_ 5yrs-1990			
Power couples (Δ^{P})	-0.016**	0.029**	-0.013**
Part-power couples (Δ^{pp})	-0.025***	0.024**	0.001**
Low-power couples (Δ^{LP})	-0.009**	0.021**	-0.012**
Coincidental power couples (Δ^{CP})	-0.006**	0.010**	-0.004**
Coincidental low-power couples (Δ^{CLP})	-0.023**	0.021**	-0.002**
Power, wife works ($\Delta^{P,W}$)	0.010**	0.034**	-0.003**
Part-power, wife works ($\Delta^{PP,W}$)	0.016**	0.038**	0.025**
Low-power, wife works ($\Delta^{LP,W}$)	0.010**	0.028**	0.022**
Power, wife does not works ($\Delta^{P,NW}$)	-0.026**	-0.005**	-0.010**
Part-power, wife does not works ($\Delta^{PP,NW}$)	-0.041**	-0.014**	-0.024**
Low-power, wife does not works ($\Delta^{LP,NW}$)	-0.019**	-0.007**	-0.034**
Coincidental power, wife works ($\Delta^{CP,W}$)	-0.007**	0.012***	-0.005**
Coincidental low-power, wife works ($\Delta^{CLP,W}$)	-0.014**	0.024**	-0.010**

Note: Differences are in probability units. Probabilities are calculated from Tables 3.6, 3.7, and 3.8, except for coincidental couples in which the woman works. These were calculated by using the multinomial logit predictors for working women only. *=p<0.05, **=p<0.01, ***=p<0.001

In table 3.11, the triple difference shows that 0.099 of the increase in power couple concentration between 1970 and 1990 is accounted for by the unique co-location problems of the college educated relative to the non-college educated. The triple difference estimate of 0.106 suggests that this part of the increase in power couple concentration is accounted for by the co-location problem of the college educated relative to the non-college educated.

In table 3.12, the triple difference of -0.024 shows that the effect of the unique co-location problems of the college educated relative to the non-college educated no longer exists. The triple difference estimate -0.007 suggests that the greater concentration of power couples in large metropolitan areas is not because of the co-locational problem.

Table 3.11: Differential Trends in Propensity to Live in Given City Size, 1970-1990, by Couple Type (based on predicted probabilities)

	City size		
	Large	Mid-size	Small
Double Differences,1990-1970			
$(\Delta^{P}-\Delta^{PP})$	0.074**	-0.009**	-0.065**
$(\Delta^{P}-\Delta^{LP})$	0.097**	-0.009**	-0.088**
$(\Delta^{P}-\Delta^{CP})$	0.095**	-0.044**	-0.051**
$(\Delta^{LP}-\Delta^{CLP})$	-0.002**	-0.035**	0.037**
$(\Delta^{P,W}-\Delta^{CP,W})$	0.223**	0.012**	0.051**
$(\Delta^{LP,W}-\Delta^{CLP,W})$	0.117**	0.088**	0.122**
Triple Differences,1990-1970			
$(\Delta^{P}-\Delta^{CP})- (\Delta^{LP}-\Delta^{CLP})$	0.099**	0.026**	-0.125**
$(\Delta^{P,W}-\Delta^{CP,W})- (\Delta^{LP,W}-\Delta^{CLP,W})$	0.106**	-0.100**	-0.071**

Note: ΔP, ΔPP, ΔLP, and ΔCP represent the change from 1970 to 1990 of the probability of being in a given-sized metropolitan area for power, part-power, low-power, and coincidental power couples, respectively. ΔP,W, ΔLP,W, ΔCP,W, and ΔCLP,W represent the probability of being in a given-sized metropolitan area for power and low-power couples in which the wife works and the probability for coincidental power and low-power couples in the woman works, respectively. Probabilities are calculated from Tables 3.6 and 3.7.
*=p<0.05, **=p<0.01, ***=p<0.001

Table 3.12: Differential Trends in Propensity to Live in Given City Size, 1990-2010, by Couple Type (based on predicted probabilities)

	City size		
	Large	Mid-size	Small
Double Differences,2010_ 5yrs-1990			
$(\Delta^{P}-\Delta^{PP})$	0.009**	0.005**	-0.014**
$(\Delta^{P}-\Delta^{LP})$	-0.007**	0.008**	-0.001**
$(\Delta^{P}-\Delta^{CP})$	-0.010**	-0.019**	-0.009**
$(\Delta^{LP}-\Delta^{CLP})$	0.014**	0.000**	-0.010**
$(\Delta^{P,W}-\Delta^{CP,W})$	0.017**	0.022**	0.002**
$(\Delta^{LP,W}-\Delta^{CLP,W})$	0.024**	0.004**	0.032**
Triple Differences,2010_ 5yrs-1990			
$(\Delta^{P}-\Delta^{CP})-(\Delta^{LP}-\Delta^{CLP})$	-0.024**	0.019**	0.001**
$(\Delta^{P,W}-\Delta^{CP,W})-(\Delta^{LP,W}-\Delta^{CLP,W})$	-0.007**	0.018**	-0.030**

Note: ΔP, ΔPP, ΔLP, and ΔCP represent the change from 1990 to 2010 of the probability of being in a given- sized metropolitan area for power, part-power, low-power, and coincidental power couples, respectively. ΔP,W, ΔLP,W, ΔCP,W, and ΔCLP,W represent the probability of being in a given-sized metropolitan area for power and low power couples in which the wife works and the probability for coincidental power and low-power couples in which the wife works, respectively. Probabilities are calculated from Tables 3.6 and 3.7.
*=p<0.05, **=p<0.01, ***=p<0.001

6 Discussion

This chapter analyzes the trends of power couples' concentration in large metropolitan areas from 1940 to 2010. The rising trend lasted from 1940 to 1990. After 1990, the concentration reaches a plateau. Using similar methods as those in Costa and Kahn's (2000) paper, I detect similar trends of rising concentration for the period 1940-1990. I then use the same "triple difference-in-difference" model to identify the differential co-location problem of power relative to low-power couples after 1990. For power couples and low power-couples, I compare each group with corresponding coincidental power/low-power couples. Then the triple difference attained by differencing the two "double differences" received above shows the co-location effect for power couples relative to low-power couples.

For the definition and the measurement of coincidental couples, I have proposed a new method to define and estimate the "coincidental couple," and use multinomial logit models to predict the probability of certain types of couples residing in certain size categories of metropolitan areas. The results of a triple difference-in-difference model confirm that the relative co-location effect faded away after 1990. I re-confirmed that from 1940 to 1990 the increasing urbanization of power couples could be explained mainly by the co-location problem, but that after 1990 the trend became relatively steady because the co-location effect faded away.

.

APPENDIX

A.1 Bitcoin Data

This section presents the correlation matrix for three different data samples.

Table A.1: Correlation Matrix for One Single Market in US

	Weighted Price _btce	Return	Volatility
Positive	0.1629	-0.1437	-0.0229
Carried Positive	0.1467	-0.0130	0.0639
Negative	0.0626	-0.1119	-0.0651
Carried Negative	0.0709	0.1182	-0.0574
Bullishness	-0.0303	0.0124	0.0908
Carried Bullishness	-0.0181	-0.1333	0.1381
Agreement	-0.0213	0.0162	0.0915
Carried Agreement	-0.0127	-0.1347	0.1382
Message Volume	0.1412	-0.1196	-0.0417
Carried Message Volume	0.1295	-0.0086	0.0396

Table A.2: Correlation Matrix for the Global Market

	Weighted Price_ALL	Return_all	Volatility_all
Positive	0.1573	-0.1409	0.0302
Carried Positive	0.1454	-0.0087	-0.0072
Negative	0.0546	-0.1361	0.0080
Carried Negative	0.0620	0.0887	-0.2026
Bullishness	-0.0251	0.0534	0.0333
Carried Bullishness	-0.0066	-0.0812	0.2266
Agreement	-0.0160	0.0568	0.0314
Carried Agreement	-0.0011	-0.0819	0.2308
Message Volume	0.1338	-0.1368	0.0306
Carried Message Volume	0.1235	-0.0239	-0.0461

Table A.3: Correlation Matrix for the Composite US Market

	Weighted Price_us	Return_us	Volatility_us
Positive	0.1608	-0.1644	0.0087
Carried Positive	0.1499	-0.0104	-0.0137
Negative	0.0502	-0.1482	-0.0177
Carried Negatice	0.0706	0.1329	-0.1735
Bullishness	-0.0162	0.0400	0.0566
Carried Bullishness	-0.0115	-0.1308	0.1850
Agreement	-0.0070	0.0436	0.0558
Carried Agrrement	-0.0063	-0.1321	0.1887
Message Volume	0.1354	-0.1521	0.0036
Carried Message Volume	0.1300	-0.0173	-0.0490

According to the above correlation matrices, we can tell Bitcoin instruments trends behave quite similar in all three markets: one single US market, global market and the composite U.S. market. Therefore, I chose the composite US market as the sample data for research in Chapter 1.

A.2 Sentiment Robustness Test

This section presents the results of models in first chapter using the Pattern analyzer as the sentiment analysis tool. The idea and the method of sentiment analysis using Pattern analyzer is as same as using Naive Bayes Classifier analyzer. I processed the same sample using Pattern analyzer, and I got the sentiment score for each day. Next, I implemented the same Granger Causality test on the same sample. Table A.4 presents the results from the Granger Causality test. According to the results, we got the same conclusion as we got from our study in Chapter 1 using Naive Bayes Classifier analyzer.

Therefore, I could conclude that this sentiment analysis of the Bitcoin market is quite Robust no matter which tool we use.

Table A.4: Granger Causality Analysis of Sentiment Information and Bitcoin Returns (Pattern Analyzer)

Returns	lag	N_pos	N_neg (Prob>F=0.0123)	Bullishness	Agreement	Mes_vol
	1	-0.0022	-0.0078	0.9071	3.8645	-1.4804
	2	-0.0021	-0.0175	0.5347	1.9218	-0.5420
	3	0.0075	0.0070	0.4631	1.5962	0.7365
	4	0.0090	0.0012	0.1185	0.3390	0.8991
	5	-0.0089	0.0021	-0.5797	-2.0256	-0.3430
	6	0.0176*	0.0213*	-0.4179	-2.0190	1.5292
	7	0.0029	-0.0013	0.5250	1.7557	0.4314

Note: ** for p-value < 0:05 and * for p-value < 0:1 which is 95% and 99% confidence interval respectively. And according to the F-statistic, only N_neg rejects the null hypothesis.

A.3 Scripts

A.3.1 Scripts for Preparing Data (Shell Scripts)

```
## extract the tar file:

$7z = "C:\Program Files\7-Zip\7z.exe"

$path = "F:\Bitcoin Project\raw data\1411\*\*"

$files = Get-ChildItem $path *.bz2 -Recurse

foreach($file in $files) {

   $fn = $file.fullName

   $fp = $file.DirectoryName

   #Write-Host $fn

   #Write-Host $fp
```

```
    & "$7z" e $fn -o"$fp"
}
```

A.3.2 Scripts for Filtering Relevant Tweets

```
import os
import glob
import json

os.chdir("C:\\Users\\Sherry Li\\Desktop\\Bitcoin Project\\filtered data\\140228")
path="C:\\Users\\Sherry Li\\Desktop\\1402\\28\\*\\*.json"

bitcoin_keywords_set=set(["bitcoin","bitcoins","bitcoins'","#bitcoin","bitcoin","bit
coins","bitcoin's","bitmines","bitmine","bitmining","mtgox","mt.gox","bitstamp","bitcoin
.org","blockchain","bitpay","cryptocurrency","btcvert.com","btcvert","coindesk","bitcoin
d","coinbase","coinbase's","blockchain.info","bitcoinexpo","bitcoin-
qt","bitinstant","bitspark","cointellect",
"bitpesa","coinfire","bitshare","bitshares","dogecoin","litecoin","bitfinex","bitx","bitcure
x","bittrex","c-cex","btce","btc-e",
"campbx","coinjar","cyptonit","cryptsy","fybse","btcinstant","virtex"])
from datetime import datetime
start_time = datetime.now()
```

```
for f in glob.iglob(path):
    path, filename= os.path.split(f)
    #drive, tail=os.path.splitdrive(f)
    filename=os.path.splitext(filename)[0]
    output_file="%s_filtered.json" %filename
    tweets_data = []
    with open(f,encoding="utf-8") as f_input:
        for line in f_input:
            tweets_data.append(json.loads(line))


    with open(output_file, "w+") as outfile:
        for tweet_data in tweets_data:
            if tweet_data.get("user"):
                if tweet_data["user"]["lang"]=="en":
                    tweet_words = tweet_data["text"].split()
                    tweet_words = [x.lower() for x in tweet_words]
                    tweet_words_set=set(tweet_words)
                    intersection = set.intersection(tweet_words_set, bitcoin_keywords_set)
                    if intersection:
                            json.dump(tweet_data, outfile)
                            outfile.write('\n')
```

```
end_time = datetime.now()
print('Duration: {}'.format(end_time - start_time))
```

A.3.3 Scripts for Calculating Sentiment Scores

```
from datetime import datetime
start_time = datetime.now()
import os
import glob
import json

import csv
#import nltk
from textblob import TextBlob
from textblob.sentiments import NaiveBayesAnalyzer

os.chdir("C:\\Users\\Sherry Li\\Desktop\\final data")

list1=['140201','140202','140203','140204','140205','140206','140207','140208','140209','1
40210','140211','140212','140213','140214','140215','140216','140217','140218','140219','
140220','140221','140222','140223','140224','140225','140226','140227','140228']
```

list2=['140301','140302','140303','140304','140305','140306','140307','140308','140309','1
40310','140311','140312','140313','140314','140315','140316','140317','140318','140319','
140320',
'140321','140322','140323','140324','140325','140326','140327','140328','140329','140330'
,'140331']

list3=['140401','140402','140403','140404','140405','140406','140407','140408','140409','1
40410','140411','140412','140413','140414','140415','140416','140417','140418','140419','
140420','140421','140422','140423','140424','140425','140426','140427','140428','140429',
'140430']

list4=['140501','140502','140503','140504','140505','140506','140507','140508','140509','1
40510','140511','140512','140513','140514','140515','140516','140517','140518','140519','
140520','140521','140522','140523','140524','140525','140526','140527','140528','140529',
'140530','140531']

list5=['140601','140602','140603','140604','140605','140606','140607','140608','140609','1
40610','140611','140612','140613','140614','140615','140616','140617','140618','140619','
140620','140621','140622','140623','140624','140625','140626','140627','140628','140629',
'140630']

```
list6=['140701','140702','140703','140704','140705','140706','140707','140708','140709','1
40710','140711','140712','140713','140714','140715','140716','140717','140718','140719','
140720','140721','140722','140723','140724','140725','140726','140727','140728','140729',
'140730','140731']
list9=['140711','140712','140713','140714','140715']

list7=['140801','140802','140803','140804','140805','140806','140807','140808','140809','1
40810','140811','140812','140813','140814','140815','140816','140817','140818','140819','
140820','140821','140822','140823','140824','140825','140826','140827','140828','140829',
'140830','140831']
list8=['140901','140902','140903','140904','140905','140906','140907','140908','140909','1
40910','140911','140912','140913','140914','140915','140916','140917','140918','140919','
140920','140921','140922','140923','140924','140925','140926','140927','140928','140929',
'140930']

for m in list9:
    path="C:\\Users\\Sherry Li\\Desktop\\Bitcoin Project\\filtered data\\%s\\*.json" %m

    for f in glob.iglob(path):
        head, filename = os.path.split(f)
        filename=os.path.splitext(filename)[0]
        output_file="%s.csv" %filename
```

```
tweets_data=[]
with open(f,encoding="utf-8") as f_input:
    for line in f_input:
        tweets_data.append(json.loads(line))

with open(output_file,'w', newline='') as f_outfile:
    writer=csv.writer(f_outfile, delimiter=',')
    for tweet_data in tweets_data:
        tweet=tweet_data['text']
        tweettb= TextBlob(tweet)
        tweettb=tweettb.replace("\n"," ")
        tweettb=tweettb.replace("\r"," ")
        sentimenttb=tweettb.sentiment.polarity

        tweetnba=TextBlob(tweet,analyzer=NaiveBayesAnalyzer())
        #tweetnba=tweetnba.replace("\n"," ")
        #tweetnba=tweetnba.replace("\r"," ")
        sentimentnba=tweetnba.sentiment.p_pos

        some_values=[(filename, sentimenttb, sentimentnba)]
        writer.writerows(some_values)
```

```
end_time = datetime.now()

print('Duration: {}'.format(end_time - start_time))
```

BIBLIOGRAPHY

Akerlof and Shiller (2009). Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism.

Ang, Andrew, and Geert Bekaert (2006): "Stock return predictability: Is it there?," Forthcoming, Review of Financial Studies.

Ajzen, I. (1985). From intentions to actions: A Theory of Planned Behavior (pp. 11-39). Springer: Berlin Heidelberg.

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211.

Antweiler, W., Copeland, B. R., & Taylor, M. S. (2001). Is free trade good for the environment?. The American Economic Review, 91(4), 877.

Arthur, W. B., Holland, J. H., LeBaron, B., Palmer, R. G., & Tayler, P. (1996). Asset pricing under endogenous expectations in an artificial stock market. Available at SSRN 2252.

Asur, S., & Huberman, B. (2010, August). Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on IEEE (Vol. 1, pp. 492-499).

Ajzen, I., (1985). From Intentions to Actions: A Theory of Planned Behavior. Springer.

Ajzen, I., (1991). The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, 50(2):179–211.

Baker, Malcolm and Wurgler, Jeffrey, (2006). Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, Vol, LXI, NO.4 August, 2006

Barber, B.M., T. Odean, and N. Zhu (2009a), "Do Retail Trades Move Markets?" Reviewof Financial Studies 22:151-186.Barber, B.M., T. Odean, and N. Zhu (2009b), "Systematic Noise," Journal of FinancialMarkets, 12:547-469.

Blanchard, Olivier J., (1979), Speculative bubbles, crashes and rational expectations, Economics Letters 3, 387-389.

Blanchard, Olivier J., and Mark W. Watson, (1982), Bubbles, rational expectations and financial markets, in Paul Wachtel (ed.) Crisis in the Economic and Financial Structure, Lexington, Mass: Lexington Books, 295-315.

Bird, S. (2006, July). NLTK: the natural language toolkit. In Proceedings of the COLING/ACL on Interactive presentation sessions (pp. 69-72). Stroudsburg, PA: Association for Computational Linguistics.

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. Cambridge, MA: O'Reilly Media, Inc..

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

Cai, F., Wang, M., and Dewen, W. (2012). The China Population and Labor Yearbook, volume 3.Brill.

Campbell, J.Y., and Perron, P., (1991), Pitfalls and opportunities: what macroeconomists should know about unit roots. NBER Macroeconomics Annual, 6, 141-201.

Campbell, John Y., and Robert Shiller, (1987), Cointegration and tests of present value models, Journal of Political Economy 95, 1062-1087.

Campbell, J.Y., and Shiller R.J., (1989), The dividend-price ratio and expectations of future dividends and discount factors. The Review of Financial Studies, 1, 195-228.

Cochrane, J. H., (1992). Explaining the Variance of Price-Dividend Ratios. The Review of Financial Studies, 5(2), 243-280.

Cochrane, J. H., (2005). Asset Pricing, Princeton: Princeton University Press.

Cooper, M.J., O. Dimitrov, and P.R. Rau, (2001), A rose.com by any other name, Journal of Finance 56, 2371-2388.

Cooper, G., (2008), The Origin of Financial Crises: Central Banks, Credit Bubbles and the Efficient Market Fallacy, Vintage Books, New York.

Costa, Dora L. and Matthew E. Kahn. (2000) "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990." Quarterly Journal of Economics, Vol. 115, No. 4, (November 2000), 1287-1315

Cunado, J., Gil-Alana, L.A., and Perez de Gracia, (2005), A test for rational bubbles in the NASDAQ stock index: a fractionally integrated approach, Journal of Banking and Finance 29, 2633-2654.

Tom De Smedt and Walter Daelemans. (2012b). “Vreselijk mooi!” (“Terribly Beautiful!”): A Subjectivity Lexicon for Dutch Adjectives. In Proceedings of the Eigh International Conference on Language Resources and Evaluation (LREC’12), pages 3568–3572, Istanbul, Turkey.Diba, Behzad, and Herschel Grossman, (1987), On the inception of rational bubbles, Quarterly Journal of Economics 87, 697-700.

Diba, Behzad, and Herschel Grossman, (1988), Explosive rational bubbles in stock prices, American Economic Review 78, 520-530.

George Edward Pelham Box and Gwilym Jenkins. (1990) Time Series Analysis, Forecastingand Control. Holden-Day, Incorporated,

Evans, George W., (1989), The fragility of sunspots and bubbles, Journal of Monetary Economics 23, 297-317.

Evans, George W., (1991), Pitfalls in testing for explosive bubbles in asset prices, American Economic Review 81, 922-930.

Evans, George W., and Seppo Honkapohja, (1992), On the robustness of bubbles in linear RE models, International Economic Review 33, 1-14.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work*. The Journal of Finance, 25(2), 383-417.

Femald and Rogers. (2002). Puzzles in the Chinese stock market, Rev. Econ. St. LXXXIV :3:8:2002 pp. 416-432

Flood, Robert, and Peter M. Garber, (1980), Market fundamentals versus price-level bubbles: the first tests, Journal of Political Economy 88, 745-770.

Flood, Robert, and Robert Hodrick, (1986), Asset price volatility, bubbles, and process switching, Journal of Finance 41, 831-842.

Flood, Robert, and Robert Hodrick, (1990), On testing for speculative bubbles, Journal of Economic Perspective 4, 85-101.

Fox, Z. (2013). “Top 10 most popular languages on Twitter.” Mashable. Dec. 17, 2013. Retrieved May 2, 2015.

Frazzini, Andrea, and Owen A. Lamont, (2008), Dumb money: Mutual Fund flows and the crosssection of stock returns, Journal of Financial Economics 88, 299-322. Froot, Kenneth, and Maurice Obstfeld, 1991, Intrinsic bubbles: the case of stock prices, American Economic Review 81, 1189-1214.

M. B. Garman and M. J. Klass, (1980). “On the estimation of security price volatilities from historical data,” The Journal of Business, vol. 53, no. 1, pp. 67–78, 1980.

Gao, G. (2014). The impacts of information on stock price assessed by social media sentiment. (Doctoral dissertation, Universiteit van Amsterdam).

Garth P. McCormick. (1969), Communications to the editorexponential forecasting: Some new variations. Management Science, 15(5):311-320, 1969.

George Edward Pelham Box and Gwilym Jenkins. (1990), Time Series Analysis, Forecasting and Control. Holden-Day, Incorporated, 1990.

Gilbert, E., & Karahalios, K. (2009, April). Predicting tie strength with social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 211-220). New York: ACM.

Goldin, Claudia. (1990). Understanding the Gender Gap: An Economic History of American Women. New York-Oxford: Oxford University Press.

Goldin, Claudia. (1997). "Career and Family: College Women Look to the Past." In F. Blau and R. Ehrenberg, Eds, Gender and Family Issues in the Workplace. New York: Russell Sage

Greenspan, Alan, (1996), Minutes of the Federal Open Market Committee, Available from www.federal.reserve.gov/transcript/1996/19960924meeting.pdf.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. The American Economic Review, 70(3), 393-408.

Gurkaynak, R. S., (2008), Econometric tests of asset price bubbles: taking stock. Journal of Economic Surveys, 22, 166-186.

Homm, U., and Breitung, J., (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. Journal of Financial Econometrics, 10(1), 198-231.

Hvidkjaer, S., (2008). Small Trades and the Cross-Section of Stock Returns. Review of Financial Studies, 21(3), 1123-1151.

Jermain C. Kaminski. (2014). Nowcasting the Bitcoin Market with Twitter Signals, MIT Media Lab

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons, 53(1), 59-68.

Kroll J.A., Davey I.C. and Felten E.W. (2013) "The economics of Bitcoin mining, or Bitcoin in the presence of adversaries", Mimeo.

Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In Proceedings of the 19th international conference on World wide web, pages 591–600. ACM.

Lamont, O.A., and R.H. Thaler, (2003). Can the market add and subtract? Mispricing in tech stock carve-outs, Journal of Political Economy 111, 227- 268.

Lane, V., & Jacobson, R. (1995). Stock market reactions to brand extension announcements: The effects of brand attitude and familiarity. The Journal of Marketing, 59 (1), 63-77.

Lamont, O.A., and R.H. Thaler, (2003). Can the market add and subtract? Mispricing in tech stock carve-outs, Journal of Political Economy 111, 227- 268.

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. Review of Financial Studies, 19(4), 1499-1529.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining Text Data (pp. 415-463). New York: Springer US.

Lo, A. W., & MacKinlay, A. C. (2011). A non-random walk down Wall Street. Princeton, NJ: Princeton University Press.

Garth P. McCormick. (1969). Communications to the editorexponential forecasting: Some new variations. Management Science, 15(5):311{320, 1969.

Mittal, A. Goel, (2013). "Stock Prediction Using Twitter Sentiment Analysis" in proceeding of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2013

Muth, J. F. (1961). Rational expectations and the theory of price movements. Econometrica: Journal of the Econometric Society, 29(3), 315-335.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. International AAAI Conference on Web and Social Media (ICWWAM), 11(122-129), 1-2.

Ofek, Eli, and Matthew Richardson, (2003), DotCom mania: the rise and fall of internet stock prices, Journal of Finance 58, 1113-1137.

Olivier, Jacques, (2000), Growth enhancing bubbles, International Economic Review 41, 133-151.

Pang, B., & Lee, L. (2005, June). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (pp. 115-124). Stroudsburg, PA: Association for Computational Linguistics.

Parkinson M.,(1980), "The Extreme Value Method for Estimating the Variance of the Rate of Return", Journal of Business, 1980, Volume 53 (No. 1), 61-65.

Phillips, Peter C. B., and Jun Yu, (2009), Limit theory for dating the origination and collapse of mildly explosive periods in time series data, unpublished manuscript

Quintaro, P. (2015) "Twitter MAU were 302M for Q1, up 18% YoY." Benzinga. April 28, 2015. Retrieved May 2, 2015.

Ritter, J.R., Welch, I., (2002). A review of IPO activity, pricing, and allocations. Journal of Finance 57, 1795-1828.

Rick, Scott, and George Loewenstein. (2008). "Intangibility in Intertemporal Choice." Philosophical

Transactions of the Royal Society B: Biological Sciences, 363(1511): 3813–24.Rao, T., & Srivastava, S. (2014). Twitter sentiment analysis: How to hedge your bets in the stock markets. In State of the Art Applications of Social Network Analysis (pp. 227-247). Heidelburg: Springer International Publishing.

Sewell, M. (2011). History of the efficient market hypothesis. RN, 11(04), 04.

Shiller, Robert, (1981), Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review 71, 421-436.

Shiller, Robert, (1984), Stock prices and social dynamics, Brookings Papers on Economic Activity 2, 457-498.

Shiller, Robert, (2000), Irrational Exuberance, Princeton, NJ: Princeton, University Press.

Souleles, N. S. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys. Journal of Money, Credit and Banking, 36(1), 39-72.

Teo, M., Woo, S., (2004). "Style Effects in the Cross-Section of Stock Returns," Journal ofFinancial Economics, vol.7, 367–398.

Thaler, R.H., (1999), The end of behavioral ¯nance, Financial Analysts Journal, 12- 17.

The Economist, (2000), September 21, Bubble.com.

Tirole, Jean, (1982), On the possibility of speculation under rational expectations, Econometrica 50, 1163-1181.

Tirole, Jean, (1985), Asset bubbles and overlapping generations, Econometrica 53, 1499-1528.

Topel, Robert H. and Michael P. Ward. (1992). "Job Mobility and the Careers of Young Men." Quarterly Journal of Economics. 107(2, May): 439-479.

Vicki Lane and Robert Jacobson, (1995). "Stock Market Reactions to Brand Extension Announcements: The Effects of Brand Attitude and Familiarity," Journal of Marketing, January 1995, 63-77

West, Kenneth, (1988), Bubbles, fads, and stock price volatility tests: a partial evaluation, Journal of Finance 43, 639-655.

Wu, Yangru, (1997), Rational bubbles in the stock market: accounting for the U.S. stock-price volatility, Economic Inquiry 35, 309-319.

Yae In Baek & Jin Seo Cho & Peter C.B. Phillips, (2013). "Testing Linearity Using Power Transforms of Regressors," Cowles Foundation Discussion Papers 1917, Cowles Foundation for Research in Economics, Yale University.

CURRICULUM VITAE

MENGMENG LI

EDUCATION

Ph.D., Economics, Boston University, Boston, MA, May 2016

M.S., Mathematical Finance, Questrom School of Business, Boston University, Sep 2009

B.A., Dual-degree in Mathematics and Economics, Wuhan University, June 2008

FIELDS OF INTEREST

Financial Economics, Computational Economics, Econometrics, and Labor Economics

WORK EXPERIENCE

Consultant, Center for Multicultural Mental Health Research, Cambridge, August 2014-May 2015

Analyst, Boston Merchant Financial, Ltd., Boston, November 2009- April 2010

Intern Analyst, Trudeau & Trudeau Associates, Inc., Boston, June 2009- August 2009

Analyst, Dongxing Securities, Wuhan, China, July 2007- June 2008

Assistant Analyst, Bank of China, Wuhan, China, June 2006 - February 2007

TEACHING EXPERIENCE

Teaching Assistant, Intermediate Microeconomics, Department of Economics, Boston University, Spring 2012

Teaching Assistant, Sports Economics, Department of Economics, Boston University, Spring 2012

Teaching Assistant, Sports Economics, Department of Economics, Boston University, Fall 2011

Teaching Assistant, Economic Statistics, Department of Economics, Boston University, Fall 2011

ACADEMIC WORK EXPERIENCE

Department Research Assistant for Professor Albert Ma, Department of Economics, Boston University, 2012-2015

Research Assistant for Professor Yanbo Wang, Questrom School of Business, Boston University, Summer 2013

WORKING PAPERS

"Big Data in Testing the Efficient Market Hypothesis of the Bitcoin Market," September 2015.

"Examine the Episodes of Exuberance and Collapse in the Chinese Stock Market and the Second-Board Market," January 2014.

"New thoughts on 'Power Couples': Does the Co-location Problem Still Exist," October 2012.

"How Medicare Advantage Has Impacted Mental Health Service Utilization," (with Benjamin L. Cook, Daniel E. Jimenez, and Darcie DeAngelo), 2015

"Understanding Provider Prescribing Behaviors after the Black Box Warning for Youth Antidepressant Use," (with Benjamin L. Cook, Darcie DeAngelo, and Alan Zaslavsky), 2015

FELLOWSHIPS AND AWARDS

Research Fellowship, Boston University, 2012-2015

Teaching Fellowship, Boston University, 2011-2012

Beta Gamma Sigma Honor Society, Questrom School of Business, Boston University, 2009

First Prize in Mathematics, Study Competition of Institute for Advanced Study, Wuhan University, 2007

Freshman Fellowship, Wuhan University, 2005

LANGUAGES

English, Chinese

COMPUTER SKILLS

STATA, SAS, R, Python, MATLAB, C/C++, Mathematica, Microsoft Office

DATABASE

API, IPUMS, Bloomberg, DataStream, Compustat, CRSP