2018

Characteristics of Stocks and Individual Investor Herd Behavior: A Causal-Comparative Study

Tze Sun Wong
Walden University

Walden University

College of Management and Technology

This is to certify that the doctoral dissertation by

Tze Sun Wong

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Mohammad Sharifzadeh, Committee Chairperson, Management Faculty
Dr. Javier Fadul, Committee Member, Management Faculty
Dr. Sunil Hazari, University Reviewer, Management Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2018

Abstract

Characteristics of Stocks and Individual Investor Herd Behavior: A Causal-Comparative Study

by

Tze Sun Wong

MA, City University of Hong Kong, 1997

BS, Chinese University of Hong Kong, 1993

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management – Finance

Walden University

November, 2018

Abstract

Some individual investors follow institutional investors in trading, a phenomenon called herding, that leads to excess market volatility and mispriced stocks. Individual investors who herded suffered from inferior investment performances and monetary losses, and the impact is broader in an individual investor dominant market such as Taiwan. Behavioral finance is the theoretical base of herd behavior. The purpose of this causal-comparative study was to examine individual investor herd behavior as related to characteristics of stocks in the Taiwan stock market. The research questions addressed what differences in individual investor herd behavior, if any, existed by market capitalization, price-to-book (P/B) ratio, and industry affiliation. The target population was the individual investors who traded in Taiwan Stock Exchange (TWSE) between January and December 2016. Participants were a purposive sampling of the target population with the exclusions of individual investors who traded illiquid stocks or exchange sanctioned stocks only. Data were collected through a subscription of TWSE data. The extent of individual herding estimated with Lakonishok, Shleifer, and Vishny's measure was 0.04. The 3 characteristics of stocks were separately and as a whole related to individual herding. The findings confirmed more serious sell-herding than buy-herding. The result from the logistic regression extended the knowledge of more serious herding in low P/B ratio stock with other variables controlled and different extents of herding by industry affiliation. The findings may improve individual investor financial literacy that may result in the positive social change of the alleviation of both herding and inferior investment performance.

Characteristics of Stocks and Individual Investor Herd Behavior: A Causal-Comparative Study

by

Tze Sun Wong

MA, City University of Hong Kong, 1997

BS, Chinese University of Hong Kong, 1993

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Management– Finance

Walden University

November, 2018

Dedication

I dedicate this scholarly work as a symbol of my pursuit of knowledge that is worthy of emulation to my children, Kennice Wong and Alvin Wong. I also dedicate this scholarly work as my wholehearted gratitude to my wife, Winnie Chow, for her understanding and support on my Ph.D. journey. Without my beloved family always on my mind, I would not have the strongest motivation to accomplish this.

Acknowledgments

My deepest thankfulness goes to the Almighty God for the gifts of intellectual capability in creating new knowledge and persevering mind in completing my Ph.D. journey. I am equally thankful to my dissertation committee: Dr. Mohammad Sharifzadeh, the mentor and chair, Dr. Javier Fadul, the committee member, and Dr. Sunil Hazari, the university research reviewer, for their advice and encouragement. Without them, I could never have advanced this far. I also thank Dr. David Fogarty and Dr. David Zhiwei Ma for leading by example in commerce as scholars-practitioners that have inspired me to follow in their footsteps.

Table of Contents

List of Tables

Chapter 1: Introduction to the Study

Introduction

Some individual investors follow institutional investors in trading, a phenomenon called herding. Herding leads to excess market volatility by one to four times the normal volatility and mispriced stocks (Andrikopoulos, Albin Hoefer, & Kallinterakis, 2014). Individual investors who herded suffered from inferior investment performances and monetary losses (Ahmed, 2014; Chung & Wang, 2016). In fact, more educated investors herded less (Nguyen & Schuessler, 2013). Nevertheless, there was a lack of knowledge about individual herding by characteristics of stocks in Taiwan where individual investors dominated. The lack of such knowledge impeded individual investors in Taiwan from improving their financial literacy.

In this chapter, I will introduce herd behavior found in previous research through experiments and empirical data from stock exchanges and the effects of herd behavior on institutional investors and individual investors. Concerning the negative effects of herd behavior on individual investors from previous research, I will present the problem statement and the purpose statement. The research questions and hypotheses section will include my hypotheses of the relations of the three independent variables of market capitalization, price-to-book (P/B) ratio, and industry affiliation of stock as well as the dependent variable of herd behavior indicator. I will then elaborate on the emergence and development of herding theory in the theoretical base section. The chapter will also include a description of the research design, research method, abbreviated data collection technique, and abbreviated data analysis plan as well as a discussion of the definitions of key terms, assumptions, limitations, and scope and delimitations. Lastly, I will provide the potential significance of this study and summarize the chapter.

Background

When an investor observes other investors' actions in a stock market and then considers their actions as such useful information that the investor adjusts his or her own information and trades on the same side of the market, this behavior is herding. Herd behavior is a branch of behavioral finance and commonly interpreted as the notion of the informational cascade (Bikhchandani, Hirshleifer, & Welch, 1992). If the proposition of informational cascade holds in reality, the herd behavior impacts the direction of the market. Previous researchers have found evidence of informational cascade.

Previous researchers have detected the herd behavior of investors in experiments and empirical settings with different measures. In experiments, investors trade in a stock market simulator. Any behavior difference between the investors who receive the information of others' trades and the investors who receive fictitious information of others' herding is an effect of the informational cascade. Andersson, Hedesström, and Gärling (2014) found that even fictitious trade information influenced investors in their study. As herd behavior happens in a stock market simulator, herd behavior very likely happens in a real stock market. In a real stock market, a stock exchange releases trade information to the public promptly. Based on empirical market data, Yao, Ma, and He (2014) found investors in China, an emerging market, herded. Litimi, BenSaida, and Bouraoui (2016) found investors in the United States, a developed market, too herded. The two studies are an evidence establishing herding in reality.

Stocks are an investment tool that some people park their wealth in for the purpose of retirement. The performance of such investment may affect the peoples' retirement lives, so it was imperative to understand the effects of herding on investors. Herd behavior brings price volatility to markets. Messis and Zapranis (2014a) and Blasco, Corredor, and Ferreruela (2012) found, respectively in Greece stock market and Ibex-35 index of Spain stock market, that herding led to higher volatility. Amid higher volatility, most stocks are mispriced, either overpriced or underpriced. Investors hardly trade at a fair price. Buying an overpriced (underpriced) stock usually results in a bigger loss (profit) when volatility subsides. Institutional investors are more knowledgeable about the economy, industry cycle, and financials of list companies than individual investors, but both institutional and individual investors herd. Foreign institutional herding caused short-term volatility in a tranquil period in Taiwan (Chen, Yang, & Lin, 2012) and India (Garg & Mitra, 2015). Despite a volatile market, institutional investors traded mispriced stocks well and produced superior investment performance (Ahmed, 2014). Given that there are only two types of investors, it was urgent to find out the effects of herding on individual investors.

Individual herding has led to inferior investment performances and monetary losses in South Korea, Qatar, and Taiwan (Ahmed, 2014; Chung & Wang, 2016; Lin, Tsai, & Lung, 2013). The losses found in these studies indicate that individual investors traded mispriced stocks on the wrong side. Individual investors commonly hold some stocks as a source of fund to prepare or support their retirement lives, and an asset loss due to herding with other investors is undesirable. Therefore, it was pressing to understand more about individual herding; however, individual herding has been less studied relative to institutional herding. Out of the 80 studies in the literature review by Spyrou (2013), there were only *four* about individual herding. A gap of knowledge about individual herding is present. With the presence of such a gap, individual investors are unable to improve their financial literacy in the aspect of herding (Bucher-Koenen & Ziegelmeyer, 2013).

Problem Statement

Herding leads to excess market volatility by one to four times the normal volatility (Andrikopoulos et al., 2014). Excess market volatility causes mispricing and mispriced stocks that make the market less efficient (Huang, Lin, & Yang, 2015). Only investors who are competent in dealing with mispriced stocks may beat the market. Institutional investors yielded superior investment performance (Ahmed, 2014). The general problem was individual investors who herded suffered from inferior investment performances and monetary losses (Chung & Wang, 2016).

The negative effect of herding – inferior investment performances and monetary losses – is likely broader in an individual investor dominant market. Taiwan is an example of such a market. Individual investors in Taiwan constituted 69% of the trading value (Lin et al., 2013) and dominated over institutional investors. There was previous research about institutional herding while institutional investors are knowledgeable; however, there was little research on individual herding while individual investors are less financially literate. The specific problem was a lack of knowledge about individual herding as related to characteristics of stocks. This lack of research about individual herding impeded individual investors in Taiwan from improving their financial literacy, which may alleviate herding and its effects.

Purpose of the Study

The purpose of this quantitative research with a causal-comparative design was to examine individual investor herd behavior as related to characteristics of stocks in the Taiwan stock market. The characteristics of stocks included (a) market capitalization, (b) P/B ratio, and (c) industry affiliation. I adopted Lakonishok, Shleifer, and Vishny’s (1992) measure (LSV measure) to estimate the extent of individual investor herd behavior. I collected buy and sell transaction data of all stocks listed in the Taiwan Stock Exchange (TWSE) between January and December 2016. I used logistic regression to examine what differences in individual investor herd behavior, if any, existed by characteristics of stocks. The findings of this study contributed to the understanding of herd behavior by industry affiliation advocated by Litimi et al. (2016) and the knowledge of behavioral finance. I will provide organizations that promote financial education to Taiwan individual investors with the findings of this study and suggest that they incorporate the findings that may lessen individual investor herd behavior. The potential positive social changes are the alleviation of both herding and inferior investment performances of individual investors

Research Questions and Hypotheses

The overarching research question of this study was what differences in individual investor herd behavior, if any, existed by the following characteristics of stocks: market capitalization, P/B ratio, and industry affiliation. I developed the following three research questions and corresponding hypotheses to guide this study:

RQ1: What differences in individual investor herd behavior, if any, exist by market capitalization of stock?

$H_0$1: There is no statistically significant difference in individual investor herd behavior by market capitalization.

H_a1: There is a statistically significant difference in individual investor herd behavior by market capitalization.

The independent variable in RQ1 was market capitalization, which is the product by multiplying current stock price with the number of shares outstanding of a company (see Nasdaq, n.d.). To assess herd behavior, I adopted the LSV measure, which in principle, is the difference between the actual proportion and the expected proportion of individual investor who net buys a stock (Lakonishok et al., 1992). Based on a *t*-test result of the LSV measure, I indicated positive or no herd behavior in another variable, the herd behavior indicator, which was the dependent variable.

RQ2: What differences in individual investor herd behavior, if any, exist by P/B ratio of stock?

$H_0$2: There is no statistically significant difference in individual investor herd behavior by P/B ratio.

H_a2: There is a statistically significant difference in individual investor herd behavior by P/B ratio.

The independent variable in RQ2 was P/B ratio, which is a financial ratio dividing a company's current market price by its book value. High P/B and low P/B stocks are characterized as growth stocks and value stocks respectively. The dependent variable was the herd behavior indicator of positive or no herd behavior.

RQ3: What differences in individual investor herd behavior, if any, exist by industry affiliation of stock?

$H_0$3: There is no statistically significant difference in individual investor herd behavior by industry affiliation.

H_a3: There is a statistically significant difference in individual investor herd behavior by industry affiliation.

The independent variable in RQ3 was an industry affiliation. I adopted the 28 industry affiliations defined by TWSE (2018a). The dependent variable was the herd behavior indicator of positive or no herd behavior.

I tested these hypotheses with a logistic regression as specified in equation 1.

$$\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (1)$$

where

X_1 is the market capitalization of a stock,

X_2 is the P/B ratio of a stock,

X_3 is the industry affiliation of a stock, and

logit(p) is the odds ratio of the positive herd behavior in logarithm form, ln[p/(1-p)], and p is the actual proportion of positive herd behavior.

The herd behavior indicator was my *t*-test result on the LSV measure. The LSV measure represents the difference between an actual proportion and an expected proportion of *r*-type investor who net buys stock *i* at time *t* (Lakonishok et al., 1992). When statistically greater than zero, it meant herd behavior on either purchase or sale side of the market, and the herd behavior indicator was positive. When not statistically greater than zero, it meant no herd behavior on both sides of the market, and the herd behavior indicator was no.

Theoretical Base

In classical finance, Fama (1970) postulated that market prices of securities in the strong form of the efficient market hypothesis (EMH) reflect all relevant information from the public to proprietary quickly (Cinar, 2014). Under or overpriced securities do not exist in a market in the strong form of EMH. It is impossible for investors to outperform the overall market by timing or identifying mispriced securities. Taking a higher investment risk is the only rational way to yield a higher investment return (Fama, 1970). In reality, both over and underpriced securities exist. Shiller (1981) defined the market price unjustified by the variation of future dividends as excess volatility. The presence of excess volatility is a sign of insufficiency of EMH assumptions; hence, a market price is a reflection of not only public and proprietary information of a company but also any excess volatility (Shiller, 1981).

To explain excess volatility, researchers like Kahneman and Tversky (1979) looked at behavioral and cognitive characteristics of individuals, and the field of behavioral finance has emerged from this combination of psychology and classical finance. Behavioral finance theory addresses investor irrationality in financial decisions. Herd behavior, also known as herding, is one behavioral finance concept tested and found in stock markets (Mahmood, Kouser, Abbas, & Saba, 2016).

Herding is the correlated behavior patterns among agents. Bikhchandani and Sharma (2001) distinguished between two types of herding: spurious and intentional. Spurious herding refers to agents' similar, but not coordinated, reactions toward latest news of an economy, industry, or a company. Intentional herding refers to agents' purposefully copied behaviors among one another. Spurious herding usually leads to an efficient outcome, whereas intentional herding may result in excess volatility and a systemic risk to markets (Spyrou, 2013). Devenow and Welch (1996) divided intentional herding into two views, nonrational and rational. The nonrational view is completely related to investor psychology, and agents neglect their own rational investment analysis and blindly follow one another to act. The rational view is related to investor cognition on other agents' behaviors and involves three scenarios: (a) the payoff externalities model, (b) the principal-agent model, and (c) information externalities also known as the informational cascade model (Devenow & Welch, 1996). The payoff externalities model is a scenario in which an agent benefits financially by following other agents' actions. The principal-agent model is a scenario of an agent's mimicking actions of other agents in the same team. The informational cascade model is a scenario in which an agent considers other agents' actions as useful information. Then, the agent rationalizes to change his or her private information and takes the same action. Across the three models, the common benefits of rational herding do not diminish and subside.

Nature of the Study

Research Design

The nature of this study was a quantitative approach with a causal-comparative design. Such a design is suitable for the research in which the researcher is unable to manipulate the variables (Brewer & Kuhn, 2010). The three independent variables of market capitalization, P/B ratio, and industry affiliation of stock were not manipulatable. The dependent variable, herd behavior in a market, is an incident that I had no way to introduce. As all variables of this study were beyond my control as the researcher, the causal-comparative design was suitable. Also, the causal-comparative design is applicable for inferring a relationship between an independent variable and a dependent variable when an incident is ex-post facto (Brewer & Kuhn, 2010). In the overarching research question, any individual herding is an incident that has already happened. Any difference in individual herding is an inferred effect of the corresponding characteristic of stock.

The causal-comparative design and correlational design are similar because both are used for examining relationships among variables with no manipulation of any variable. A feature – one group of samples, instead of two or more groups – differentiates correlational design. The feature of one group of samples made correlational design not applicable to this study in which a comparison among 28 industry affiliations was necessary. Experimental design and quasi-experimental design were not applicable to this study either. In experimental design or quasi-experimental design, a researcher manipulates the independent variable of a treatment group. The three independent variables of this study were not manipulatable. In conclusion, the causal-comparative design was the most suitable for this study.

Research Method

The unit of analysis in this study was an individual investor. Due to the dominance of individual investors and occurrences of individual herding in Taiwan market (Lin et al., 2013), I selected individual investors in Taiwan as the target population. As TWSE does not release the latest 12 months trade transaction data, December 2016 was the most recent month of trade transaction data available as of January 2018. I defined 1 year as the data span, and therefore, the sampling period was from January to December 2016.

Both Chung and Wang (2016) and Ahmed (2014) used purposive sampling, a nonprobability type of sampling. Following their example, I also used purposive sampling and included all TWSE trade transaction data of individual investors with the exclusions of illiquid stocks and exchange sanctioned stocks. The samples consisted of the majority of the population. The independent variables were market capitalization, P/B ratio, and industry affiliation of stock, while the dependent variable was the herd behavior indicator, which captured my *t*-test result of the LSV measure. I used the LSV measure to further assess individual herding by buying and selling separately. Last, I ran a logistic regression with the three independent variables against the dependent variable. I interpreted not only the logistic regression model results but the validity and reliability of the entire test.

Abbreviated Data Collection Technique

I used the statistical analysis software, SAS, in the Oracle Virtual Machine environment to process TWSE data, which were a secondary data. TWSE is an order-driven market where investors must submit orders into the TWSE system for call auction order matching before the market is open or for normal order matching when the market is open. As a result, TWSE data are official, complete, and reliable. The data sources were (a) monthly trade transaction data files, (b) basic information of all stocks files, (c) monthly statistics files, and (d) a list of international securities identification numbers. A monthly trade transaction data file captures the trade transactions of all stocks in a month. A basic information of all stocks file consists of closing price and the number of shares listed for market capitalization estimation. A monthly statistics file consists of P/B ratio. The list of international securities identification number consists of the security code and industrial group. The data sources collectively were sufficient for this study.

Abbreviated Data Analysis Plan

I created a master data file by combining the three secondary data sources, (a) monthly trade transaction data files, (b) basic information of all stocks files, and (c) monthly statistics files. Following Wermers' (1999) practice, I excluded new issues and delisted stocks. Also following Barber, Odean, and Zhu's (2009) suggestion, I excluded illiquid stocks traded by fewer than 10 individual investors in a day. I identified all trade transactions of individual investors by referring to the column of investor type; however, TWSE data are anonymous. I followed Jame and Tong's (2014) practice to assume the uniqueness of individual investor of every transaction within a month. I aggregated separately all buying and selling transactions, then calculated the LSV measure for each stock-day in the master data file. The LSV measure is a value from -1.0 to 1.0 (Chang, Chen, & Jiang, 2012). I performed a *t* test on the LSV measure result at 1% significance level. When statistically greater than zero, it meant herd behavior on either purchase or sale side of the market, and the herd behavior indicator was positive. When not statistically greater than zero, it meant no herd behavior on both sides of the market, and the herd behavior indicator was no.

I produced the three independent variables in this study. First, I multiplied the closing price and the number of shares listed to get the market capitalization of stock by stock-month. I sorted all stocks by market capitalization in ascending order and by stock-month then divided them by quintile. The quintiles were low, mid-low, middle, mid-high and high market capitalization. Second, I used the P/B column in the master data file by stock-month. Last, I classified each stock in one of the 28 industrial affiliations by stock-month. I created the dependent variable, herd behavior indicator, which was dichotomous. I assigned the herd behavior indicator to 1 and 0 respectively for positive and no herd behavior tested out. I tested the hypotheses of this study by running a logistic regression in SAS. The goal was to find the best fitting model to describe the relationship between the herd behavior indicator and the three characteristics of stock, market capitalization, P/B ratio, and industry affiliation. I used Wald tests to examine independent variables at 5% significance level. Also, I used Nagelkerke R^2 to assess the goodness of fit of the logistic regression model as a whole.

Definitions of Terms

The following terms appeared in this study. Some have special meanings in the context of finance, so I specified definitions for them all:

Adverse herding: A representation of the mean reversion toward the long-term equilibrium β, risk-return relationship (Lakshman, Basu, & Vaidyanathan, 2013). Adverse herding must exist to resume asset prices back to fundamental values after herding.

Consensus heuristic: The abbreviation of "the consensus implies correctness" heuristic that is an influence resulted from the belief that "the majority is always right" (Andersson et al., 2014, p.227).

Efficient market hypothesis: Fama (1970) postulated that current security prices in a market of EMH strong form reflect all relevant information.

Excess volatility: Shiller (1981) defined excess volatility as the portion of market price unjustified by the variation of future dividends.

Financial education: This type of education includes informational programs, such as credit counseling, homeownership classes, and retirement seminars, that policymakers promote to raise the personal finance management capability of individuals (Meier & Sprenger, 2013).

Financial literacy: The ability to understand financial knowledge and an individual's financial condition and use skills to manage financial resources. Higher financial literacy is positively related to better financial outcomes (Meier & Sprenger, 2013).

Herd behavior: This behavior is correlated but unnecessarily coordinated investment behavior patterns among investors. Devenow and Welch (1996) pronounced difficulty in defining herd behavior in finance precisely. The terms, herd behavior and herding, are interchangeable.

Herding: The terms, herding and herd behavior, are interchangeable.

Individual herding: Chang et al. (2012) used individual herding to stand for individual investor herd behavior.

Individual investor: These investors are a highly heterogeneous group of natural persons some of who are highly skilled while others are naïve in investment (Fong, Gallagher, & Lee, 2014). They trade their assets through intermediaries of different types such as discount retail brokers and full-service brokers (Fong et al., 2014).

Industry affiliation: A group of companies that operate in the same segment of the economy. TWSE (2018a) defined 28 industrial groups or industry affiliations. Each company belongs to one that represents the company's business nature the most.

Institutional herding: Choi and Skiba (2015) used institutional herding to stand for institutional investor herd behavior.

Institutional investor: These investors act on behalf of other entities including investment banks, pension funds, mutual funds, hedge funds, and others who trade relatively a big volume of shares in one transaction (Chen et al., 2012).

Intentional herding: Investors' purposefully copied behaviors among one another (Spyrou, 2013).

Market capitalization: A company’s market value that is estimated by multiplying the current market price with the total number of shares outstanding of a company (see Nasdaq, n.d.).

Nonrational herding: The blind following of investors that is completely related to investor psychology (Devenow & Welch, 1996).

Price-to-book ratio: A financial ratio dividing a company’s current market price by its book value. High P/B and low P/B stocks are characterized as growth stocks and value stocks respectively (Garcia & Oliveira, 2018).

Rational herding: Devenow and Welch (1996) postulated rational herding as investor cognition on other investors’ behaviors and following.

Spurious herding: Investors’ similar but not coordinated reactions toward new fundamental-driven information (Spyrou, 2013).

Zero-cost strategy: Stemming from the momentum effect, zero-cost strategy creates a portfolio by buying top 10% and selling bottom 10% of stocks measured regarding monthly returns during the formation period (Chang et al., 2012).

Assumptions

In this study, the first assumption I held was the validity of the LSV measure in three aspects: construct, content, and empirical. The rational view of intentional herding postulates market participants’ revaluation and trades of stock after other market participants’ trades. Lakonishok et al. (1992) built the measure upon investors’ trade behaviors to estimate the proportions of buyer and seller and to then assess any disproportionate buyer or seller on one stock during a given period. The trades of any market participants are the most direct evidence of behavior in their measure. There is a logical tie between the LSV measure and the concept and assumptions of herding theory; therefore, the LSV measure is construct-valid (Wermers, 1999). Besides, the LSV measure covers an exhaustive list of domains: herding presence and extent. The LSV measure does not leave any relevant attribute out, so it is also content-valid (Lakonishok et al., 1992). I approached the empirical validity of the LSV measure with a predictive validity test to evaluate the relationship between the LSV measure and herd behavior. Lin et al. (2013) used a herding measure generated by a bootstrap run and found higher herding among institutional investors than individual investors in Taiwan. Chang et al. (2012) used another herd measure, the LSV measure, and had a similar finding. Notwithstanding that the two studies were of different sampling periods, the reconcilable results imply predictive validity, and therefore, the empirical validity of LSV measure. With the validity of all three aspects, construct, content, and empirical, I assumed the validity of the LSV measure.

Another assumption I held was a mitigation to the anonymity of trade transaction data. TWSE trade transaction data are anonymous. In the absence of a unique investor identifier, there is no way to link transactions traded by the same investor. Chang et al. (2012) and Lin et al. (2013) did not explain how to deal with the anonymity of their Taiwan data, and they estimated herding measures in a manner that each trade transaction belonged to a unique investor. It was an implicit assumption that one investor at most buys and sells each stock once in a day. I held the same implicit assumption of the maximum of one buy trade and one sell trade on each stock per investor in a day for this study.

Limitations

I evaluated limitations addressing two parts of validity: internal and external. For internal validity, I did not recruit individual investors as samples. I did not need to set up and assign individual investors to experiment and control groups either. There was no selection per se; hence, the selection effect was not relevant. I used purposive sampling to exclude illiquid stocks and exchange sanctioned stocks only. The samples consisted of a majority of the population; hence, regression artifact was least relevant. I referred to history as events including the measures of Taiwan Securities and Futures Bureau implemented in 2016. Individual investors might have changed their herd behavior unintentionally as a result of these events; hence, history was relevant. Maturation was referred to a natural process through which individual investors got more mature in 2016; hence, maturation was relevant. Experimental mortality referred to the drop-outs of individual investors in 2016. It was possible that some individual investors traded in early 2016 but not anymore in late 2016. Such individual investors were the drop-outs.

For external validity, reactive arrangement was irrelevant because there was no experimental setting in this study. All investor trades in the stock exchange were authentic actions. There was no need for investors to react to an experiment. Hence, reactive arrangement was irrelevant. The representativeness of the sample was high because the samples consisted of 94%, a majority of the population. Only individual investors who traded illiquid stocks or exchange sanctioned stocks only were the portion unrepresented.

I also evaluated limitations to the reliability of the results, referring to the consistency of the LSV measurements in this study. I estimated the LSV measures directly from TWSE data, a secondary data. The mathematical formula of the LSV measure was standard; hence, the limitations to the reliability of this study were negligible. I evaluated generalizability to a larger population, a different time, or a different market. In this study, the samples consisted of 94%, a majority of the population. The findings from such a majority were already general for the individual investors of Taiwan. The data span of 1 year was long enough to capture many variations. The variations were probably generalizable to certain months or years before and after the sampling period. The findings of more serious herding in either high market capitalization or high P/B ratio stocks were already general between Taiwan and China and should be generalizable to different markets. Overall, the limitations to the generalizability of this study were negligible.

I also evaluated two tiers of limitations to the use of secondary data: primary and secondary. All three primary limitations, a gap between research purposes, limited data accessibility, and information insufficiency, were not relevant. Code variation of the secondary limitation was not relevant either.

Lastly, I evaluated the limitations specific to TWSE data. Data not connectable to individual investor level were a limitation to computing the actual number of buyers and actual number of sellers. Nonnormality of the LSV estimates was another limitation from TWSE data. The distributions were not normal even after Box-Cox transformation.

There was a combined limitation of history, maturation, and experimental mortality to internal validity but least limitation to external validity. The overall validity was typical of causal-comparative design (Frankfort-Nachmias, Nachmias, & DeWaard, 2015). In tandem, there were negligible limitations to reliability, generalizability, and the use of secondary data. There were bigger limitations in TWSE data. I mitigated the bigger limitations with the adoptions of an implicit assumption and an alternative multivariate analysis method. The overall limitations of this study were typical and mitigable, and therefore, acceptable.

Scope and Delimitations

Taiwan is a suitable market for a study of individual herding. In contrast to a small portion of shares held by individual investors in a developed stock market, Taiwan individual investors constituted about half of total investors from January to June 2016 (TWSE, 2018b). Taiwan individual investors also constituted 69% of the total trading value in 2010 (Lin et al., 2013). Nevertheless, the previous research (Huang et al., 2015; Huang, Wu, & Lin, 2016; Lin et al., 2013) covered mostly institutional herding but barely individual herding. The scope of this study was the underresearched topic of individual herding in Taiwan. The individual investors who traded between January and December 2016 were the target population.

There were two key limitations of this study, data not connectable to the individual investor level and a combined limitation of history, maturation, and experimental mortality. Data not connectable to the individual investor level was a limitation to computing the actual number of buyers and actual number of sellers. Following the example of Chang et al. (2012) and Lin et al., I mitigated the limitation with an implicit assumption of each trade transaction belonging to a unique investor. The origin of a part of the combined limitation of history, maturation, and experimental mortality was the nonexperimental nature of this study. By nonexperimental nature, I mean there was no splitting of individual investors in groups and no intervention such as exposure to objective investment information. Therefore, it was possible to delimit the combined limitation of history, maturation, and experimental mortality.

Significance of the Study

Individual herding is mitigable to a certain extent. Bucher-Koenen and Ziegelmeyer (2013) found German investors with both lower financial literacy and lower cognitive ability were the most likely to sell all their assets and realize bigger losses during the U.S. mortgage crisis. Conversely, investors with higher financial literacy were more likely to hold on to assets even with losses in value and benefited more from the market recovery (Bucher-Koenen & Ziegelmeyer, 2013). Investors with higher financial literacy maneuvered herding with better investment performance. Financial education can improve the financial literacy of investors. Nguyen and Schuessler (2013) had a similar finding showing that more educated investors herded less. Nevertheless, financial education does not cover herding. Herd behavior is neither a topic nor a question in the financial literacy quiz of the U.S. national financial capability study (see FINRA Investor Education Foundation, 2017). Neither is herd behavior a part of the syllabus in the online financial education course offered by Taiwan Securities and Futures Bureau (2018a). The findings of this study filled in the gap of knowledge about individual investor herd behavior by characteristics of stocks in Taiwan and contribute to the overall knowledge of behavioral finance. I will provide the results of this study to organizations that promote financial education to Taiwan individual investors and suggest that they incorporate the findings to potentially lessen individual investor herd behavior. The potential positive social changes are the alleviation of both herding and the inferior investment performance of individual investors.

Summary

Some individual investors follow institutional investors in trading, a phenomenon called herding. Herding leads to excess market volatility by one to four times the normal volatility and mispriced stocks (Andrikopoulos et al., 2014). Individual investors who herded suffered from inferior investment performances and monetary losses (Ahmed, 2014; Chung & Wang, 2016), and the impact is broader in an individual investor dominant market such as Taiwan. In fact, more educated investors herded less (Nguyen & Schuessler, 2013). Nevertheless, herd behavior was not a part of the syllabus in the online financial education course offered by Taiwan Securities and Futures Bureau (2018a). A lack of knowledge about herd behavior impeded individual investors in Taiwan from improving their financial literacy. This study was necessary to fill in the gap of knowledge about individual investor herd behavior concerning characteristics of stocks in Taiwan and to contribute to behavioral finance knowledge. The potential positive social changes are the alleviation of both herding and the inferior investment performance of individual investors. In next chapter, I will discuss herd behavior from its theoretical base to empirical findings by other authors. I will also explain the conflicting findings among previous studies and the scarcity of literature about individual herding in Taiwan.

Chapter 2: Literature Review

Introduction

Among the peer-reviewed articles I located for this literature review, there were four types of herd behavior studies: theory, empirical evidence, experiment, and simulation. Bikhchandani and Sharma's (2001) study was an example of the theory type. Bikhchandani and Sharma postulated the difference between spurious herding and intentional herding and defined each. Gavriilidis, Kallinterakis, and Tsalavoutas' (2016) study was an example of the empirical evidence type, in which the authors hypothesized Ramadan and nonRamadan days as an independent variable and herd behavior as a dependent variable. Gavriilidis et al. used the actual trade data of seven countries with major populations of Muslims and tested the hypothesis empirically. Ivanov, Levin, and Peck's (2013) study was an example of the experiment type. Ivanov et al. studied herd behavior in an endogenous-timing game and generalized the systematic biases demonstrated by four players to the real world. Chen, Zheng, and Tan's (2013) study was an example of the simulation type, in which the authors constructed an agent-based model that took account of individual and collective behaviors of the investors in real markets. Rather than from statistical model fitting, Chen et al. performed simulations 100 times for an average of 10,000 agents. Among the four types, I chose to conduct this study using an empirical approach because findings from empirical research are more likely implementable in reality.

In this chapter, I will present my literature review with the following organization. In the first part, I will describe the emergence of behavioral finance, especially herd behavior. The second part will include a discussion of herding theory, herding measures, and causes and effects of herding, while the third part will contain a review of the unfavorable effects of individual herding. The confluence of the three parts will point to a need for further understanding of individual herding. In the fourth part, I will identify and discuss the research variables for this study.

Herding Theory

Deficiency of Efficient Market Hypothesis

The strong form of EMH does not exhibit in a market at all time. In classical finance, Fama (1970) postulated that market prices of securities in the strong form of the EMH reflect all relevant information from the public to proprietary quickly (Cinar, 2014). In the context of one stock, its market price reflects all relevant information related to its company instantaneously. The market price moves again when further information related to the company appears. The market price is thus called a fair price. Under or overpriced securities do not exist in a market in the strong form of EMH. It is impossible for investors to outperform the overall market by timing or identifying mispriced securities. Taking a higher investment risk is the only rational way to yield a higher investment return (Fama, 1970). In reality, both over and underpriced securities exist. Shiller (1981) defined the market price unjustified by the variation of future dividends as excess volatility. Excess volatility is empirically present in stock markets of the United States (Wang & Ma, 2014), Hong Kong (Lam & Qiao, 2015), and others. Trading at excess volatility is beyond the assumption of investor rationality in classical finance. The presence of excess volatility is also a sign of insufficiency of EMH assumptions. Hence, a market price is a reflection of not only public and proprietary information of a company but also any excess volatility. To explain excess volatility, researchers like Kahneman and Tversky (1979) looked at behavioral and cognitive characteristics of individuals. The field of behavioral finance has emerged from this combination of psychology and classical finance. Behavioral finance theory addresses investor irrationality in financial decisions. Herd behavior, also known as herding, is one behavioral finance concept tested and found in stock markets (Mahmood et al., 2016).

The Emergence of Herding Theory

Investors' observing the trades of each other is a common behavior that may explain investor irrationality in financial decisions to some extent. In the 1950s, psychologists introduced decision making as a research topic in an attempt to understand people's information processing and decision making under an assumption of bounded rationality (Simon, 1979). Daniel Kahneman and Amos Tversky's findings on judgment and decision making set the course of cognitive psychology (Royal Swedish Academy of Sciences, 2002). Investors do not decide as rationally as what the EMH and the capital asset pricing model assume (Shiller, 1981). Rather, investors are prone to cognitive biases where cognitive psychology is relevant. Kahneman and Tversky's (1979) research bridged between economics and psychology, two academic disciplines, and became the most cited paper in *Econometric*, the academic journal in economics (Noble Prize Organization, 2002).

Nowadays, the bridge takes two forms, theory and application, and these two forms together serve as the basis of behavioral finance (Royal Swedish Academy of

Sciences, 2002). Behavioral finance theory addresses investor irrationality in financial decisions and includes concepts such as anchoring, familiarity bias, loss aversion, herd behavior, overconfidence, prospect theory, and others (Park & Sohn, 2013). Researchers validated these concepts through experiments or with empirical data.

Herd behavior is one of the concepts validated in developed markets, implying that investors do follow each other to trade. Herd behavior may be the imitated actions induced by sociological factors among market participants in the midst of uncertainty that Keynes (1936) suggested. It was important to understand the extent of the sociological impact on a market. Bikhchandani et al. (1992) suggested that waves of investor sentiment could cause stock market price movements as if herding could shift a social equilibrium. The market price movements may lead to bubbles or crashes which can cost investors heftily. The Tulip Mania was an incident in which a commodity price tremendously deviated from its intrinsic value (Kindleberger, 2016). Herding might be one of the factors to the price deviation of the Tulip Mania. Amid the emergence of herding theory, researchers defined its concepts and assumptions.

Development of Herding Definition

At the early stage, herding was a mere idea of correlated behavior among market participants, particularly in a boom or bust. Devenow and Welch (1996) pronounced difficulty in defining herding in finance precisely, stating that the real trigger behind a market participant to trade on the same side as others' is hard to know. One possibility is that the market participant independently cognizes the information about a company in the same way others do. Another possibility is that the market participant has a “follow the leader” mentality (Bakar & Yi, 2016) and assumes the majority not to go wrong.

Eventually, Bikhchandani and Sharma (2001) distinguished between two types of herding: spurious and intentional. Spurious herding refers to market participants’ similar but not coordinated reactions toward new fundamental-driven information, such as a company’s latest earnings release. Intentional herding refers to market participants’ purposefully copied behaviors among one another. Spurious herding usually leads to an efficient outcome, whereas intentional herding may result in excess volatility and a systemic risk to a market (Spyrou, 2013).

Devenow and Welch (1996) further divided intentional herding into two views: nonrational and rational. The nonrational view is completely related to investor psychology with market participants neglecting their own rational investment analysis and blindly following one another to trade. The rational view is related to market participants’ cognition of others’ behaviors. Between the two types of investors, institutional and individual, institutional investors stay more closely with market news, and they usually herd in a spurious way (Chang et al., 2012). Most individual investors take time to digest news and market movement, and they usually herd rationally and intentionally (Chang et al., 2012). There are far more individual investors than institutional investors, so a further understanding of the rational view of individual herding was necessary. The rational view of intentional herding occurs in three models: (a) the payoff externalities model, (b) the principal-agent model, and (c) information externalities also known as the informational cascade model.

Payoff externalities model. A payoff externality is a positive consequence of an economic activity experienced by an unrelated third party (Devenow & Welch, 1996). There are three payoff externalities models: bank runs, liquidity in markets, and information acquisition (Devenow & Welch, 1996). In a bank run, a large number of customers join concerned customers to withdraw their deposits from the same bank. The concern can be a lack of confidence in a particular bank, the national banking system, or others. Not being the last customer who can be left empty-handed is a positive consequence. Market participants like joining to trade in stock exchanges with economies of scale, and high liquidity is a positive consequence of this (Devenow & Welch, 1996). Some investors choose to acquire information about the stock at later times than other investors do. The latecomers may benefit from the forerunners by saving efforts to research all stocks in the first place and selecting popular ones later. Such benefit is a positive consequence. All three payoff externalities models apply to both institutional and individual investors.

Principal-agent model / reputation. When an agent manages an investment portfolio on behalf of a principal, the agent is always experiencing a comparison between his or her performance and other agents by the principal. The principal usually assumes equal access to public information among all agents inside an institution. To avoid being seen as incompetent in investment that results in a loss, an agent may mimic other agents' trades (Ortiz, Sarto, & Vicente, 2013). As such, the agent is in the mainstream. If the investment performance turns out undesirable, the market but not the agent is to be liable. Even competent agents may hide as they prefer conformity (Lavin & Magner, 2014). The principal-agent model hides an agent in a herd to preserve the agent's reputation. Due to fewer principal-agent relationships among individual investors, the principal-agent model is more relevant to institutional investors.

Informational cascade model. Investors observe other investors' actions in a stock market. Some investors consider such actions as such useful information that they rationalize adjustments in their own private information and trade on the same side. The new trades become a continuation of the *useful information*. More investors observe, rationalize, and trade. Such notion of the informational cascade is commonly referred as herding (Bikhchandani et al., 1992). Institutional investors act promptly upon news released in a market and usually are the ones cascading information to individual investors through their trades. Hence, informational cascade model is more relevant to individual investors.

In summary, spurious herding, the payoff externalities model, and the principal-agent model of intentional herding are more relevant to institutional investors. The payoff externalities model and informational cascade model of intentional herding are more relevant to individual investors. Using these bases, I criticized the evidence of herd behavior in previous literature in this literature review.

Evidence of Herd Behavior

Herd Behavior Detected in Experiments

Researchers have attempted to validate herd behavior theory in reality in parallel to the development of the theory. An experimental approach is ideal for this validation because with an experimental approach, researchers can measure any direct effect of treatment. Duxbury (2015) stated that the ability of controls over environmental parameters in an experimental approach is key for studying herding theory. In an experiment, Andersson et al. (2014) introduced a financial incentive as a treatment to motivate one group of participants, but not other groups, to predict market prices more accurately. The financially-incentivized group did not adopt a consensus heuristic from the majority in their study; instead, the group processed information independently and predicted more accurately than other groups. On the contrary, the other groups were under the informational social influence, or informational cascade effect, in the context of behavioral finance (Andersson et al., 2014). They assumed the majority's decision was correct, so they imitated it. Nevertheless, the other groups did not always adopt a consensus heuristic, especially when the volatility of stock price was low, so financial incentive and stock price volatility were two factors with opposite effects. I considered the evidence of the informational cascade found in an experimental approach the most important of all.

The informational cascade model of intentional herding is not only a concept but a reality. Delfino, Marengo, and Ploner (2016) manipulated three pieces of social information in experiments to cause participants to imitate in investment, accordingly further substantiates the reality of herding. (a) A big deviation between a participant's investment choice and peers', (b) social information referred as the average of a group than a single peer, and (c) a short time window for investment decision are the three pieces of social information that each piece causes herding among participants. The experimental approach is straightforward and rigorous. The evidence from the experimental approach is solid. It was necessary to understand whether the evidence would be valid beyond the experiment and in a real market.

To assess the external validity to a real market setting, I focused on the three findings of Andersson et al.'s (2014) experiments and Delfino et al.'s (2016) experiments which logically should happen in a real market. First, investors likely predict more accurately in pursuit of bigger financial incentives. Second, when investors decide under time pressure or have their own estimates way different than the current market price, the investors likely accept the current market price as a market consensus. Last, with an access to market and company information, investors likely ignore the current market price and predict independently. The three findings were likely externally valid but not sufficient to a real market. First, the information cascaded in the experiments was far less than that in a real market in terms of both information sources and information amount. Second, the experiments resembled only informational cascade but not together with payoff externalities and principal-agent. In reality, all three models of the rational view of intentional herding happen concurrently. Last, the participants in the experiments did not trade genuine stocks. Their trade behavior might be different with genuine stocks and their own money. The low external validity and the unnoticeability of investor's private information (Duxbury, 2015) are inherent limitations from the experimental approach. The evidence from the experimental approach is necessary but not sufficient for understanding herd behavior in a real stock market. In a real stock market, there are many participants who value and trade stocks in different ways. To reflect the complexity of reality, a herding measure developed upon empirical data of a stock market becomes critical.

Herding Measures for Empirical Data

There are two families of herding measures for the empirical data of a stock market. It is needful first to understand how the empirical stock market data originate, then to understand each family of herding measures. Market participants place trade orders in their brokers' computer systems. The computer systems eventually send all trade orders to the computer system of a stock exchange. The computer system of the stock exchange processes and logs trade transactions. The trade transaction logs become empirical data. The trade transaction logs may have captured a herding incident that a researcher may discover with a herding measure. A researcher is unable to resemble any experimental data with the complexity at par with the empirical data. Because of the complexity of the empirical data, different researchers developed herding measures with different focuses. Spyrou (2013) classified herding measures in two families: (a) the closeness between an asset return and a market consensus, and (b) the difference between the buy and sell sides. In the first family, the three key herding measures are Christie and Huang's (1995) cross-sectional standard deviation (CSSD), Chang, Cheng, and Khorana's (2000) cross-sectional absolute deviation (CSAD), and Hwang and Salmon's (2004) CSSD of systematic risk. In the second family, two key herding measures are Lakonishok et al.'s (1992) measure and Sias' (2004) cross-sectional correlation of systematic risk. I discussed the characteristics of each family of herding measures in detail.

Closeness between an asset return and a market consensus. The idea of the first family herding measure is an increase in the closeness between an asset return and a market consensus amid herding in the market. Christie and Huang (1995) argued that, under extreme market movements, investors do not insist on their views but follow the market consensus to act. If so, returns of individual stocks should not deviate too much from the market return. Return dispersions should decrease. When return dispersions do not decrease, rational asset pricing presumably holds. Christie and Huang developed CSSD to capture the closeness between an asset return and a market consensus and regressed the CSSD against two dummy variables. One was of market return at an extreme low-tail and another was of market return at an extreme high-tail. Positive coefficients of the dummy variables indicate rational asset pricing whereas negative coefficients indicate investor herding. In a concept similar to the CSSD, Chang et al. (2000) developed CSAD by taking the average of the deviation of each stock relative to the return of the equally weighted market portfolio in absolute value. Chang et al. postulated a replacement of linear and increasing relation between a market return of its dispersion by a nonlinear or decreasing one amid herding. For each direction of the market, rising or falling, there is an equation regressing the CSAD against a linear term and a quadratic term of the market return. A negative coefficient of the quadratic term indicates an occurrence of herding in the specific rising or falling market. I considered the CSAD as an enhancement of the CSSD. In parallel, Hwang and Salmon (2004) enhanced the CSSD by considering the cross-sectional standard deviation of systematic risk, instead of the return, in a state-space framework. Hwang and Salmon’s herding measure is proportional to the deviations of the true beta from the market unit beta. The herding detection is an autoregressive AR(1) process (Solakoglu & Demir, 2014) that is different from the CSSD. After all, the three key measures in the first family are a comparison of an asset return against a market consensus. They are applicable for a study in all, not a part, of the market participants. When there is an interest in only a part of the market participants, for example, individual investors only, the three key measures are not applicable. Another family of herding measures, the difference between the buy and sell sides, are more suitable.

Difference between the buy and sell sides. The idea of the second family herding measure is a disproportionate stock buying or selling by specific investors amid herding in the market. Lakonishok et al. (1992) considered a tendency of market participants especially money managers disproportionately buying or selling an individual stock as herding. The LSV measure is the proportion of net buying market participants relative to all market participants who trade a particular stock in a given quarter minor an adjustment factor. The adjustment factor declines as the number of market participants trading the same stock rises. A steady LSV measure from a period to another indicates the absence of herding whereas a disproportionate LSV measure indicates the presence of herding. On a similar basis to the LSV measure, Sias (2004) argued the proportion of net buying market participants of stock this quarter covarying with the proportion of last quarter as a sign of herding. Sias derived a new measure by transforming the raw fraction of institutional buying into a standardized fraction of institutional buying. As such, a positive cross-sectional correlation between the standardized fractions of buying on one security between two quarters indicates an occurrence of herding. One prerequisite of both key measures in the second family is the availability of trade transaction data. Stock exchanges are usually a source of trade transaction data. An example is Comisión Nacional del Mercado de Valores (Spanish Securities Markets Commission) from where Gavriilidis, Kallinterakis, and Ferreira (2013) obtained the data from June 1995 to September 2008.

Applications of herding measures on empirical data. Depending on the characteristics of the empirical data, researchers used the more applicable family of herding measures. For example, Ahsan and Sarkar (2013) and Vo and Phan (2017) applied the CSSD and CSAD from the first family of herding measures on the return data from Dhaka Stock Exchange in Bangladesh and Ho Chi Minh City Stock Exchange in Vietnam respectively. Zhang and Zheng (2016) applied the LSV measure from the second family of herding measures on the top 10 security investment funds in China. The examples showed not only the applicability of the herding measures, but also the detectability of herding from stock exchange data which were empirical. Next, it was essential to understand the extent and duration of herd behavior in stock markets.

Herd Behavior Detected in Markets

The researchers of the following studies detected herd behavior from stock exchange data in five markets, namely China, Indonesia, Jordan, Turkey, and the United States. I discussed about the stock exchange tenures which were related to their maturity levels. Then, I focused on their findings on the extents and durations of herd behavior.

China. Shanghai Stock Exchange and Shenzhen Stock Exchange both handle A- and B-shares trading. There are four segments in total. A-shares are renminbi-denominated. In the past, only local investors could trade A-shares. Since 2003, foreign institutions can trade A-shares too through Qualified Foreign Institutional Investor. B-shares in Shanghai and Shenzhen are U.S. dollar-denominated and Hong Kong dollar-denominated respectively. In the past, only foreign investors could trade B-shares. Since February 2001, local investors can trade B-shares too. Yao et al. (2014) used the CSSD and CSAD to detect any herd behavior in the four segments between January 1999 and December 2008. All four segments exhibited herd behavior in the beginning of the decade (Yao et al., 2014). The extent of herd behavior diminished over time.

Indonesia. Ramli, Agoes, and Setyawan (2015) adopted the LSV measure to examine any herd behavior between two groups of investor – domestic and foreign – in Indonesia stock market. Continuous buy- and sell-herd behaviors were present between January 2009 and December 2011. Domestic investors exposed to information asymmetry always tended to follow foreign investors' trades, especially sell trades.

Jordan. Ramadan (2015) employed the CSAD in a time-series design to detect any herd behavior upon the 100 companies constituted to the Amman Stock Exchange (ASE) Index of Jordan between January 2000 and August 2014. Ramadan found a decrease of the CSAD related to an increase in market return. In other words, investors in ASE overly followed market performance and resulted in herding.

Turkey. Borsa Istanbul (BIST) entitles only companies fulfilling the National Market listing criteria. Second National Market (SNM) entitles small-to-medium sized companies not fulfilling the National Market listing criteria. Solakoglu and Demir (2014) examined any difference in investor trade behavior between BIST30 stocks and SNM stocks with Hwang and Salmon's (2004) AR(1) state-space model. The persistence parameter and the variance of the signal error were insignificant for BIST30 but significant for SNM; hence, herding was not present in BIST30 but SNM. Furthermore, Solakoglu and Demir found three SNM herding stages: (a) lack of confidence in the government from 2000 to 2004, (b) no herding from 2005 to 2008, and (c) adverse herding by conflicting signals from shocking events.

The United States. Litimi et al. (2016) enhanced the CSSD and CSAD with the introduction of three components, (a) potential herding triggers, (b) vector autoregression, and (c) Granger causality test. Litimi et al. then examined any effect of trading volume and herd behavior on conditional volatility of the average stock return by sector. Among all firms listed on New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ Stock Market (NASDAQ) between January 1985 and December 2013, herd behavior occurred in eight out of twelve sectors. Similarly, Roszkowski and Richie (2016) found herding in the United States stock markets among a sample of the 1,581 Jim Cramer's *Mad Money* buy and sell recommendations over 3.5 years. From abnormal returns with control for momentum, buy- (sell-) herding was more likely on the day following Cramer's buy (sell) recommendation. Roszkowski and Richie's and Litimi et al.'s findings of industry herding were in line with each other.

Across markets and by industry affiliation. Herd behavior happened across the five markets. China Shenzhen Stock Exchange and Jordan ASE have been operating for

26 and 18 years since respective establishments in 1991 and 1999. China Shanghai Stock Exchange and Indonesia Stock Exchange have been operating for 27 and 40 years since respective reopening in 1990 and 1977. Turkey SNM, formerly named Regional Markets and Wholesale Market, has been operating for 22 years since the launch in 1995. Relative to NYSE’s 224 years of establishment since 1792 and BIST’s 151 years of establishment since 1866, China, Indonesia, Jordan, and Turkey SNM stock exchanges are emerging. Except for Turkey BIST (Solakoglu & Demir, 2014), herd behavior happened in both emerging markets and developed markets. Hence, I did not consider the tenure of a stock exchange as a critical criterion for the market selection for this study. Herd behavior did not happen continuously but occasionally. For China, it was only the beginning from 1999 to 2008 (Yao et al., 2014). For Turkey, it was only the first 4 years from 2000 to 2008 (Solakoglu & Demir, 2014). To increase the likelihood of getting the stock exchange data with the occasional herd behavior, the sampling period for this study should be reasonable long such as 1 year. Litimi et al. (2016) and Roszkowski and Richie (2016) examined herd behavior at industry level which is a level more granular than an entire market. A finding at the industry level is useful for investors. The investors may manage their portfolios by rebalancing at the industry level. For this study, I pursued the industry level. As herd behavior was common, it was imperative to understand its causes and effects.

Causes of Herd Behavior

There are two types of factors, exogenous and endogenous, to herd behavior. Exogenous factors are outside the stock trading model or experiment whereas endogenous factors are within. I discussed investor cognitive psychology, signal strength, social events, stock characteristics, and trade characteristics from previous studies and identified the factors relevant to this study.

Exogenous Factors

Investor cognitive psychology. Investor cognitive biases are the postulation of behavioral finance. Investor cognitive psychology which may lead to the biases is the factor that researchers have studied. There are three factors of investor cognitive psychology including confidence, anxiety, and risk tolerance. Among sell-side analysts with lower confidence, the ones last in a group issuing earnings forecast herded more (Durand, Limkriangkrai, & Fung, 2014). Lin (2012) had a similar finding that investors with higher confidence mediated by higher risk tolerance herded less. In addition, investors with higher anxiety mediated by lower risk tolerance herded more. The three factors of investor cognitive psychology are measurable and predictive in an experimental setting. In reality, investor cognitive psychology is not mandatorily measured and reported. Moreover, investor cognitive psychology is unobservable in stock exchange transaction data which I used; hence, investor cognitive psychology factors were irrelevant to this study.

Signal strength. There is a lot of information such as interest rate rise, corporate earnings release published to a market from time to time. Luchtenberg and Seiler (2013), in an experimental setting, measured the effects of strong and weak information signals on participants’ herding likelihood in strategical default of underwater mortgages. Under a strong information signal, respondents acted and even provided peers with advice. It resulted in a higher herding likelihood. Under a weak information signal, respondents kept any decision with each of them. The herding likelihood was not noticeable. Signal strength can be defined clearly in an experimental setting, but not in reality. Messis and Zapranis (2014b) attempted to identify which macroeconomic variables caused changes in major indices. Messis and Zapranis found positive shocks of Industrial Production index and 10 years bond increasing investor herding magnitude in the DAX of Germany; on the contrary, the positive shocks of Industrial Production index were an indication of an overheated economy in the United States and led to a sell-off in the Standard and Poor’s (S&P) 500 index. The release of Industrial Production index or other macroeconomic variables is an event whose signal strength on each investor is hard to measure. Due to the lack of signal strength measurement, signal strength factor was irrelevant to this study.

Social events. Social events such as market mergers, the U.S. mortgage crisis, military takeover, and religious occasion caused herd behavior in different markets. The mergers among four EURONEXT equity markets (Belgium, France, the Netherlands, and Portugal) were one example (Andrikopoulos et al., 2014). Herd behavior was significant only in the Netherlands before the market mergers, but in Belgium, France, and the Netherlands after the market mergers. Andrikopoulos et al. attributed the intensified herding to the higher participation of foreign investors who flew more information at faster paces. Besides, the enhanced transparency of EURONEXT accounted for an extent in the short term. Despite a market merger was not a social event for this study, I considered its nature of impacting personal investment relevant. I took notes of the social events which might impact personal investment in the period of this study. For Egypt stock market, Guvercin (2016) attributed the intensified herd behavior between July 2002 and May 2014 to two social events, the U.S. mortgage crisis and the Egyptian military takeover. Poshakwale and Mandal (2014) also found the higher levels of herding to persist after the U.S. mortgage crisis from 2009 to 2011. The U.S. mortgage crisis was another social event imparting personal investment. The Egyptian military takeover was a social event related to politics whose nature I considered relevant for this study too. In contrast, Filip and Pochea (2014) concluded the U.S. mortgage crisis as a weak factor of herd behavior to five Central and East Europe stock markets between 2008 and 2010. Herd behavior happened in only certain sectors in Bulgaria, Czech Republic, Hungary, and Poland. The social event caused herding in different markets to different extents. Ramadan is a Muslim religious occasion which is a social event of another nature. Gavriilidis et al. (2016) associated the significant herding in the stock markets of Bangladesh, Egypt, Indonesia, Malaysia, Morocco, Pakistan, and Turkey with the positive mood of investors during Ramadan; however, Javaira and Hassan (2015) had a conflicting finding that Pakistan stock market did not exhibit herding. Anyway, social events related to personal investment or politics may cause herding. However, validations are in a post-mortem manner. I took note of the social events in these two aspects in the period of this study.

Endogenous Factor

Stock characteristics. Every stock represents a listed company that operates in one industry such as finance, telecommunication, and retail. Every industry has its business cycle. Food manufacturer is relatively stable whereas aviation is sensitive to the economy. Every industry also has its potential. For example, traditional retailers are threatened by online retailers whereas healthcare services are prosperous with an aging population. These variations collectively constitute an endogenous factor, industry affiliation. In the American Depository Receipt market, Demirer, Kutan, and Zhang (2014) found herding more prevalent at the industry level in (a) basic industries; (b) capital goods; (c) food and tobacco; and (d) textile and trade in the midst of large market downtowns. The extent of herding varies by industry affiliation. Apart from industry affiliation, each listed company is unique regarding the number of shares outstanding, earning, and book value. Even two companies are in the same industry, their earnings are likely different. The product of current stock price and the number of shares outstanding is market capitalization. The quotient of market value by book value is P/B ratio. The extent of herding varies by market capitalization and P/B ratio too. Each stock characteristic has an implicit meaning. For example, the meaning of high P/B ratio is growth. Investors may herd in different extents for a growth company. Investors usually have basic knowledge about some characteristics of stocks such as industry affiliation, market capitalization, and P/B ratio. Any findings related to the basic knowledge are more understandable and implementable in reality; therefore, I adopted the characteristics of stocks for this study.

Trade characteristics. During stock trading, there are data related to the trades including bid-ask spread, stock price, and transaction volume. A bid-ask spread is the difference between the highest price that a buyer bids and the lowest price that a seller asks. A stock price is a price at which an exchange of a stock between a buyer and a seller happens. The stock price changes over time because of subsequent exchanges of the stock at different prices. A transaction volume is the number of shares exchanged between buyers and sellers. These data collectively reflect trade characteristics of an investor. Litimi et al. (2016) found the transaction volume together with the average market return which was estimated from stock prices related to herd behavior in the U.S. stock markets. The market return and transaction volume in time series are common trade characteristics for the herding measures of the first family but not for the LSV measure which I used; so, the trade characteristics were not relevant to this study.

Factors Chosen for this Study

Among the three exogenous factors, investor cognitive psychology and signal strength are the two factors whose data are not publicly available. Even if I invited participants and collected their data, I would never be able to link the collected data with the anonymous data from a stock exchange. Due to such limitation, these two factors were irrelevant to this study. The third factor, social events related to personal investment and politics, may cause herding; so, it is relevant to this study. Between the two endogenous factors, stock characteristics were relevant to this study because investors exhibited different extents of herd behavior by stock characteristics. Trade characteristics were not relevant to this study because I used the LSV measure which is based on number of buyers and number of sellers. Apart from the causes, it was imperative to understand the effects of herding.

Effects of Herd Behavior

I discussed the effects of herding from a macro to micro level. First, I reviewed how herding in one market incurred herding in other markets. Then, I focused on one market and further on different types of investors. The conclusion of herding effects from previous studies was critical. Based on it, I could anticipate potential values and potential social changes of this study.

Effects on Other Markets

Yang, Hsu, Lai, and Lee (2015) used vector autoregression and found Dow Jones Industrial Average (DJIA) at a significant leading position of herding over four out of six East Asian stock market indices from 1995 to 2009. Continuing bear market of 2- to 4-day in DJIA was related to herding in Nikkei Stock Average, Hang Seng Index, and Korea Composite Stock Price Index whereas 2- to 3-day continuing bear market in DJIA was related to herding in Taiwan Capitalization Weighted Stock Index. Shenzhen Stock Exchange Composite Index and Shanghai Stock Exchange Composite Index were exceptions. I considered DJIA’s herding effects on other markets as a potential cause of herding, a social event related to personal investment.

Effects on Volatility of Local Market

Besides other markets, herd behavior impacts a local market. Blasco et al. (2012) found herd behavior causing volatility – an excessive fluctuation of stock price – in the constituent stocks of Spain Ibex-35 index from 1997 to 2003. Similarly, Messis and Zapranis (2014a) examined any effect of herding on market volatility in Athens Stock Exchange by creating three portfolios based on the beta. The high beta portfolio exhibited continuous herding since 1997 till mid-2004. The medium beta portfolio exhibited herding on and off in two periods, from early 1997 to late 1998 and from 2001 to 2003. The low beta portfolio exhibited adverse herding since 1995 till 2001 at a decreasing rate. Herding occurred at all beta levels to different durations. The high beta portfolio that consisted of value stocks exhibited more lasting herding than medium or low beta portfolio did. The finding conflicted with the convention wisdom about value stocks. Compared with growth stocks, value stocks are at less aggressive prices and generate steady dividends. Investors should hold on value stocks to growth stocks. Furthermore, Messis and Zapranis found a linear effect of herding on volatility. Amid herding, market volatility increased. Stock prices fluctuated. It was critical to understand the outcome of investors trading overpriced or underpriced stocks.

Investor Type

There are two major types of investors, institutional and individual, participating in a stock market. Institutional investors act on behalf of other entities including investment banks, pension funds, mutual funds, hedge funds, and others who invest in a relatively big volume of shares. Institutional investors are usually knowledgeable about fundamentals of an economy, asset valuations, and regulations set forth by the Securities and Exchange Commission. Noninstitutional or more commonly known as individual investors include all other market participants. Individual investors trade their wealth through an intermediary such as a broker, bank and so on. Individual investors, in general, are less knowledgeable than their institutional counterparts.

Effects of Institutional Herding

Institutional herding refers to the trades of institutional investors following each other into and out of the same securities. Its first effect is market volatility. Chen et al. (2012) found, in Taiwan market, foreign institutional investors chased to buy (sell) stocks with high (low) prior industrial returns in a tranquil period from 2002 to 2006. As such, foreign institutional herding drove industrial returns extremely up (down). Higher (low) returns were the results of higher (low) stock prices; hence, foreign institutional herding caused extreme stock prices which were market volatility. Garg and Mitra (2015) had a similar finding in India stock market. Based on an adjusted daily closing CNX Nifty 50 index from January 2003 to June 2014, Garg and Mitra established lead-lag relations between the series of LSV measure and conditional volatility. The foreign institutional investor feedback trading behavior led to herding which was higher in a bullish market than in a bearish market. The foreign institutional herding created short-term volatility and impeded price efficiency of India stock market. Institutional herding led one common effect, an increase in market volatility. This finding was consistent with that of overall market herding.

Another common effect was superior investment performance of institutional investor. Ahmed (2014) examined herd behavior by institutional and individual investors in the Qatar Exchange with daily net investment flow and daily closing price of the stock composite index. The net investment flow was an indicator of an investor group’s ability in responding more promptly to the market than another investor group’s. With the positive auto-correlation in lagged net investment flows, Ahmed concluded each investor group tended to herd together. The institutional investors technically pursued a positive feedback investment strategy that resulted in superior investment performance in 4 years. The institutional investors in the Qatar Exchange maneuvered and even profited from a volatile market. It was imperative to see how generalized the finding of institutional investor superior investment performance through institutional herding was. Chang et al. (2012) modified zero-cost strategy to estimate the profit by taking advantage of herding by investor type in Taiwan stock market. There were three portfolios of (a) all investors, (b) institutional investor, and (c) individual investor. Each comprised buy-herding stocks of an investor type as the top 10 percent and sell-herding stocks of the same investor type as the bottom 10 percent. The zero-cost strategy would have generated a smaller profit from the institutional investor portfolio than that from the individual investor one. The smaller profit meant that it was harder to exploit institutional herding than individual herding. Institutional investors weathered a volatile market well. Whether individual investors were at stake was a pressing concern.

Effects of Individual Herding

Institutional investors keep abreast of financial news, so act timely. Individual investors are rarely the first to move. With the response lag, individual herding likely ends up in a financial loss. Chung and Wang (2016) found that the increases in two factors of the lagged period, (a) net buy in the number of shares by individual investors and (b) return of the stock led to the increase in net buy in the number of shares by individual investors in the current period. So, individual investors in South Korea kept buying the shares of one stock at a price already increased after other individual investors had bought. Buying shares at an increased price was the first effect of individual herding. Chung and Wang also found a significantly negative relation between 1-week (4-week) lagged returns and 1-week (4-week) current returns. So, individual investors who herded to buy the stock at an increased price turned out holding the stock at a decreased price in 1 or 4 weeks later. A loss in investment was the second effect of individual herding. Ahmed (2014) had a similar finding that individual herding in Qatar resulted in poor investment performance. In an autoregressive model, Ahmed found for individual investors the estimated coefficients of almost all lagged market returns significantly negative. Individual investors herded to buy shares at price rises but ended up selling or holding the shares at losses. The effect of individual herding is opposite to that of institutional herding. The loss in investment is adverse to individual investors. Such adversity occurred in Taiwan too. Lin et al. (2013) examined the relation between herding and trading noise by investor type. In a crisis period, the more serious individual herding was, the higher was the subsequent trading noise. More and more individuals herded; however, institutional investors did not do the same. The subsequent trading noise of institutional herding faded away quickly. In conclusion, individual herding and its adversity occurred in South Korea, Qatar, and Taiwan, and may be generalizable to other markets. Individual investors were certainly at stake. It became essential to size the broadness of its potential impact.

Target Chosen for this Study

In a market, there are usually more individual investors than institutional investors, though the former altogether may not contribute a higher trading value than the latter. One way to size the broadness of the potential impact by individual herding is based on the contribution of total trading value by individual investors. If the contribution is considerable, there is an urgency for a further study. Relative to other developed markets, there was a higher individual investor contribution, 88% by transaction or 61% by value, in South Korea stock market (Chung & Wang, 2016). Taiwan was a market with an even higher individual investor contribution with 69% by stock trading value (Lin et al., 2013). Given the substantial contribution of total trading value by individual investors in Taiwan, the potential impact by individual herding was broad and urgent. I chose the individual investors in Taiwan as the target population of this study.

Problems of Individual Herding

General Problem

Institutional investors respond to new information timely and trade in a market. When institutional investors trade in the same side of the market, they herd. Some individual investors follow to trade in the same side; therefore, herd too. Herding leads to excess market volatility by one to four times the normal volatility (Andrikopoulos et al., 2014). Excess market volatility causes mispricing and mispriced stocks that make the market less efficient (Huang et al., 2015). Only investors who are competent in dealing with mispriced stocks may beat the market. Institutional investors yielded superior investment performance (Ahmed, 2014). The general problem was individual investors who herded suffered from inferior investment performances and monetary losses (Chung & Wang, 2016).

Individual herding is to a certain extent mitigable. Nguyen and Schuessler (2013) examined any difference in three behavioral finance concepts including herding, home equity bias, and anchoring among 890 German respondents by socio-demographic attributes. Respondents with higher education levels demonstrated lower inclinations to herding and the other two behavioral finance concepts. Bucher-Koenen and Ziegelmeyer (2013) had a similar finding that the German investors with higher financial literacy likely held on their assets at higher losses in value and benefited from the market recovery during the U.S. mortgage crisis. Conversely, the investors with lower financial literacy likely sold off their assets, realized losses, and missed the market recovery. Investors can improve financial literacy through financial education. Nevertheless, financial education did not cover herding. Herding was neither a topic nor a question in the financial literacy quiz of the U.S. national financial capability study (see FINRA Investor Education Foundation, 2017). There was a need to include herding in the syllabus of financial education. I saw an escalation of such need because online trading individual investors acted on new information promptly and herded more heavily than offline trading individual investors did (Choi, 2016). The growth in network connection makes online communication relentless and likely fuels individual herding. For example, a group of six conspired and circulated rumors about a nuclear reactor explosion in North Korea through instant messaging on January 6, 2012, then put the financial market in panic mode to make $54,314 profit (Shin & Yoon, 2012). Any mitigation of individual herding was a positive thing to do; however, the unavailability of individual herding knowledge and the growing popularity of online trading would aggravate individual herding. It was reasonable to carry out this study for Taiwan where the individual investor contribution of total trading value was high.

Specific Problem

The negative effect of herding – inferior investment performances and monetary losses – is likely broader in an individual investor dominant market. TWSE is ranked the 19^{th} in the globe with US$950 billion market capitalization as of March 31, 2017 (TWSE, n.d.). Unlike the dominance of institutional investors in the U.S. stock markets, individual investors in Taiwan constituted 69% of the trading value (Lin et al., 2013) and dominated over institutional investors. There was previous research about institutional herding while institutional investors are knowledgeable; however, there was little research on individual herding while individual investors are less financially literate. According to Lin et al.'s best knowledge, Chang et al.'s (2012) study was the only one individual herding study about Taiwan. I regarded Lin et al.'s study as the second. Two previous studies were too scarce. The specific problem was a lack of knowledge about individual herding as related to characteristics of stocks. This lack of research about individual herding impeded individual investors in Taiwan from improving their financial literacy, which may alleviate herding and its effects. Taiwan was ranked the 25th in the globe with 80% Internet user penetration (Internet Society, 2017). The high Internet user penetration enables swift flows of financial news, analyst opinions, and market rumors. With online security trading functionality, individual investors may herd more seriously.

Descriptions of Research Variables

Individual investors are less financially literate; therefore, it would be more effective for me to study comprehensible characteristics of stocks and communicate any finding through a financial education channel for individual investors. Yao et al. (2014) studied three comprehensible characteristics of stocks including market capitalization, P/B ratio, and industry affiliation in China markets. Market capitalization is a company's market value. The current market price and number of shares outstanding of a company are the two pieces of information easily accessible and for market capitalization estimation. P/B ratio is a financial ratio for differentiations between growth stock and value stock. The current market price and book value of a company are the two piece of information easily assessable for P/B ratio estimation. Industry describes a group of companies that operate in the same segment of the economy. Examples are banking, textile, and manufacturing. Individual investors should understand an industry and have come across market capitalization and P/B ratio in financial news. I selected these three stock characteristics as research variables for this study and describe each of them in detail below.

Market Capitalization

Market capitalization is a company's market value, a measure of corporate size. It is the product by multiplying current stock price with the number of shares outstanding of a company (see Nasdaq, n.d.). Both institutional and individual investors usually track large capitalization stocks. Relative to small capitalization stocks, trade amounts of large capitalization stocks are higher. Yao et al. (2014) found, in the Chinese market, the strongest herd behavior among (a) the smallest and (b) the largest capitalization stocks, but insignificant herd behavior among mid-tier companies. Zheng, Li, and Zhu (2015) had a similar finding that institutional investors in the Chinese market herded in smaller and larger capitalization stocks. I validated these findings with the studies of other markets. Arjoon and Bhatnagar (2017) found more prominent herding in small capitalization stocks listed on the Trinidad and Tobago Stock Exchange too; therefore, herding in small capitalization stocks was common. Nonetheless, Gong and Dai (2017) found differently that herd behavior was across all portfolios by market capitalization regardless under up or down market conditions in the Chinese market. The finding of herding in small capitalization stocks was not universally common. There was a need to further validate in another market like Taiwan.

Price-to-Book Ratio

P/B is a financial ratio dividing a company's current market price by its book value. High P/B and low P/B stocks are characterized as growth stocks and value stocks respectively. The companies of growth stocks reinvest most revenues in merger and acquisition, business expansion, product innovation, and so on whereas the companies of value stocks control expenses and steadily generate dividends. Book-to-market ratio (BTM) denotes the same idea in a reciprocal form of P/B ratio. Yao et al. (2014) sorted all stocks by BTM in ascending order and found herd behavior across the board except the quintile of the highest BTM or value stock. Hence, investors did not follow to trade value stocks amid herding. Gong and Dai (2017) found strong herd behavior in growth stocks and weak herd behavior in value stocks regardless under up or down market conditions in the Chinese market; therefore, herding in growth stocks was common. Low to no herding in value stocks was also common. In other words, investors followed to trade growth stocks but not value stocks. However, Hou, McKnight, and Weir (2014) found the higher the BTM is; the stronger is the herding. Investors followed to trade value stocks. Hou et al.'s findings conflicted with Yao et al.'s and Gong and Dai's. There was a need to examine the relation between P/B ratio and individual herding for Taiwan market.

Industry

Industry describes a group of companies that operate in the same segment of the economy. The term industry is interchangeable with the sector. In a top-down investment approach, investors organize stocks by industry. Investors then understand an industry with its characteristics. For example, the finance industry is highly competitive. The airline industry is much dependent on the economic cycle. Investors, especially individual ones, who are constrained with time resources to evaluate stocks one by one benefit from industry-based evaluation and investment. However, Jame and Tong (2014) found that individual investors in the United States chased an industry affiliation with great past returns to its price away from fundamentals. Institutional investors herded by industry affiliation too despite to a different extent (Lavin & Magner, 2014).

Usually, every stock exchange classifies a company in one of the preset industries. Yao et al. (2014) classified 1,314 companies listed on Shanghai Stock Exchange and Shenzhen Stock Exchange into 21 industries. Industry herding occurred in 15 industries including (a) food and beverage, (b) textile and apparel, (c) paper and print,

(d) petrochemical, (e) electronics, (f) metals, (g) pharmaceutical, (h) other manufacturing, (i) utilities, (j) construction, (k) transport, (l) wholesale and retail, (m) real estate, (n) social services, and (o) media. Industry herding did not occur in the remaining six industries including (a) agriculture, (b) mining, (c) machinery, (d) information technology, (e) financials, and (f) conglomerate. Hence, investors followed to trade certain industries only. Industry herding happened in other markets too. Litimi et al. (2016) followed the NASDAQ classification to group the 4,183 companies listed on NYSE, AMEX, and NASDAQ from 1985 to 2013 into 12 industries. Ouarda, El Bouri, and Bernard (2013) classified the 174 companies consistently listed in the Euro Stoxx 600 from 1998 to 2010 into 10 industries according to Bikhchandani and Sharma's (2001) suggestion of sufficient homogeneity within each industry affiliation. Although Litimi et al.'s and Ouarda et al.'s classifications were not identical, most industries were comparable. In the United States, industry herding occurred in public utilities and transportation only. In Europe, industry herding occurred in almost all industries. They included (a) industrials, (b) basic materials, (c) consumer services, (d) oil and gas, (e) finance, (f) health care, (g) public utilities, (h) technology, and (i) telecommunication. Consumer goods which were equivalent to consumer durables and consumer nondurables in the United States were an exception. Consumer goods were the common industry that investors did not follow to trade whereas public utilities were the common industry that investors followed to trade in both the U.S. markets and Europe market. Filip and Pochea (2014) examined industry herding in Central and East Europe stock markets with a focus on six industries including (a) banks, (b) constructions, (c) oil and gas, (d) other financials, (e) pharmaceuticals, and (f) hotels. Under low market volatility, herding occurred in banks, pharmaceuticals, and other financials in Bulgaria, Hungary, and Poland. Under high market volatility, herding occurred in oil and gas, banks, and hotels in Bulgaria and Czech Republic. Romanian market did not exhibit any industry herding. Construction was another common industry that investors did not follow to trade in Central and East Europe markets. It is evident that investors followed to trade by industry. The industries by which investors herded were inconsistent by markets. For example, industry herding did not happen in construction in Central and East Europe but China. Also, the industries by which investors held on were inconsistent by markets. Given the inconsistency, there was a need to examine the relation between industry affiliation and individual herding for Taiwan market.

Three Characteristics of Stocks

In conclusion, investors herded in a volatile market to different extents by the three characteristics of stocks, market capitalization, P/B ratio, and industry affiliation. For example, herding in small capitalization stocks was common in some markets but not generalizable across all markets. Similarly, herding in low P/B or value stocks was common in some markets but invalid in other markets. Herding happened variably by industry affiliation by the market. There was not a consistently validated knowledge to leverage for Taiwan market. The knowledge seemed mostly market specific. Therefore, I included the three characteristics of stocks for this study which was Taiwan market specific.

Summary

Herding leads to excess market volatility by one to four times the normal volatility (Andrikopoulos et al., 2014). Excess market volatility causes mispricing and mispriced stocks that make the market less efficient (Huang et al., 2015). Institutional investors dealt with mispriced stocks competently and yielded superior investment performance (Ahmed, 2014). The general problem was individual investors who herded suffered from inferior investment performances and monetary losses (Chung & Wang, 2016). The negative effect of herding was likely broader in an individual investor dominant market like Taiwan where individual investors constituted 69% of the trading value (Lin et al., 2013). However, there was little research on individual herding while individual investors are less financially literate. The specific problem was a lack of knowledge about individual herding as related to characteristics of stocks. This lack of research about individual herding impeded individual investors in Taiwan from improving their financial literacy, which may alleviate herding and its effects. Investors herded to different extents by the three characteristics of stocks, market capitalization (Zheng et al., 2015), P/B ratio (Yao et al., 2014), and industry affiliation (Jame & Tong, 2014); however, the findings were not generalizable across all markets. There was not a consistently validated knowledge to leverage for Taiwan market. Therefore, I included the three characteristics of stocks and used the LSV measure to study individual herding in Taiwan market specifically. I will elaborate the research method, research design, data sources of the three characteristics of stocks, and the equation of the LSV measure which I used in next chapter.

Chapter 3: Research Method

Introduction

To examine herd behavior through an empirical approach, I reviewed peer-reviewed articles where researchers employed the same research designs and research measures as I did in this study. With the availability of trade data from stock exchanges, most authors of the peer-reviewed articles employed a causal-comparative design. I focused herding assessment on an investor segment than an entire market.

In this chapter, I will elaborate on the overarching research question and the three characteristics of stocks and will define the target population and sampling method. I will also integrate the details about the LSV measure including its mathematical formula described fragmentarily by Lakonishok et al. (1992), Wermers (1999), and Chang et al. (2012). I will describe from TWSE data the formations of the three independent variables and the dependent variable, herd behavior indicator, which will be the *t*-test outcomes of the estimated LSV measure results. I will also present the data collection and data analysis plans and will discuss the validity and reliability of the LSV measure.

Research Design

The overarching research question of this study was what differences in individual investor herd behavior, if any, existed by the following characteristics of the stock: market capitalization, P/B ratio, and industry affiliation. The three research questions and corresponding hypotheses I developed to guide this study were:

RQ1: What differences in individual investor herd behavior, if any, exist by market capitalization of stock?

H$_0$1: There is no statistically significant difference in individual investor herd behavior by market capitalization.

H$_a$1: There is a statistically significant difference in individual investor herd behavior by market capitalization.

The independent variable in RQ1 was market capitalization, which is the product by multiplying current stock price with the number of shares outstanding of a company (see Nasdaq, n.d.). I sorted all stocks by market capitalization in ascending order then divided them into quintiles, or five groups. The number of stock in each group was comparable with that of Yao et al.'s (2014) study. Moreover, the five groups collectively differentiated high market capitalization to low market capitalization. To assess herd behavior, I adopted the LSV measure, which in principle, is the difference between the actual proportion and the expected proportion of individual investor who net buys a stock. Based on my *t*-test results of the LSV measure, I indicated positive or no herd behavior in another variable, the herd behavior indicator, which became the dependent variable for RQ1.

RQ2: What differences in individual investor herd behavior, if any, exist by P/B ratio of stock?

H$_0$2: There is no statistically significant difference in individual investor herd behavior by P/B ratio.

H$_a$2: There is a statistically significant difference in individual investor herd behavior by P/B ratio.

The independent variable in RQ2 was P/B ratio, which is a financial ratio dividing a company's current market price by its book value. High P/B and low P/B stocks are characterized as growth stocks and value stocks respectively. The dependent variable in RQ2 was the herd behavior indicator of positive or no herd behavior.

RQ3: What differences in individual investor herd behavior, if any, exist by industry affiliation of stock?

$H_0$3: There is no statistically significant difference in individual investor herd behavior by industry affiliation.

H_a3: There is a statistically significant difference in individual investor herd behavior by industry affiliation.

The independent variable in RQ3 was an industry affiliation. I adopted the 28 industry affiliations defined by TWSE (2018a). The dependent variable in RQ3 was the herd behavior indicator of positive or no herd behavior.

The three independent variables and the dependent variable in this study were not manipulatable. A researcher, like I, had no way to manipulate any elements of the variables including market prices, numbers of shares issued, book values, industry affiliations, and herd behaviors in the real stock market. Under the condition of all variables beyond a researcher's manipulation, the causal-comparative design of a quantitative method was suitable (Brewer & Kuhn, 2010). The ex-post facto incident of herd behavior was another condition under which the causal-comparative design was suitable (Brewer & Kuhn, 2010). Any herd behavior in this study had already happened before I collected the data. I attempted to infer the relationships between herd behavior and the three characteristics of stocks through the causal-comparative design. The inference of this study could never be definitive due to the lack of random sampling and experimental intervention, so there could be other variables impacting the herd behavior (Brewer & Kuhn, 2010). Causal-comparative design and correlational design are similar in the sense that both are for examining relationships among variables without manipulations. The correlational design, however, was not applicable to this study due to its feature of only one group of samples. Such feature was not sufficient for the necessary comparisons among 28 industry affiliations. The experimental and quasi-experimental designs were not applicable to this study either. In the experimental design or quasi-experimental design, a researcher manipulates the independent variable of a treatment group. The three independent variables of this study were not manipulatable. In conclusion, the causal-comparative design was the most suitable for this study.

Definition of Target Population

Population Nature

The unit of analysis is the most elementary part of a phenomenon under study. Lakonishok et al. (1992) argued that a money manager rather than a fund is the most appropriate unit of analysis for institutional herding. In the same vein, I considered an individual investor as the most appropriate unit of analysis for individual herding. Due to the dominance of individual investors and occurrences of individual herding in Taiwan market (Lin et al., 2013), I selected individual investors in Taiwan as the target population. Individual herding occurs over a period of time rather than all at once, so either longitudinal data or repeated cross-sectional data of the target population were essential. Data spans of most previous herd behavior studies were no longer than 1 year (Chang et al., 2012). I adopted the same data span of 1 year and defined the sampling period as January to December 2016. As TWSE does not release the latest 12 months trade transaction data, December 2016 was the most recent month of trade transaction data available at the time of this study. In summary, I selected individual investors in Taiwan from January to December 2016 as the target population.

Population Size

As of December 2016, there were accumulated 9,772,316 investors with trading accounts in Taiwan (TWSE, 2018b). Among them, 1,072,236 investors had trading activities in December 2016 with 52.8% and 0.0% as domestic individuals and foreign individuals respectively (TWSE, 2018b). Altogether 565,711 individuals had trading activities, and these 565,711 individual investors constituted the target population of December 2016. Nevertheless, TWSE did not report the number of transactions traded by individual investors solely but stated there were a total of 12.807 million transactions traded by all investor types in December 2016. There is no linkage between the two sets of figures, so I could not estimate the number of transactions traded by individual investors in December 2016.

Consequently, I turned to another TWSE dataset: trade transaction data. Chang et al. (2012) and Lin et al. (2013) studied individual herding and institutional herding in the Taiwan market, and both used trade transaction data from TWSE. TWSE tags each trade transaction with an investor type but without a unique investor identifier. The investor type includes proprietary dealer, investment trust, individual, and foreign institution (see

TWSE, 2018b). In the absence of a unique investor identifier, it was not feasible to link trade transaction data up to the individual investor level. Neither Chang et al. nor Lin et al. stated whether they linked up the trade transaction data to an individual investor level. Also, neither stated the population size regarding individual investors. Chang et al. and Lin et al. included trade transaction data of almost all stocks in each month under their data spans. They implicitly assumed that each transaction in a month belonged to a unique individual investor. I made the same assumption in this study. I subscribed to the trade transaction data from January to December 2016 and focused on the data tagged as individual investors for this study.

Sampling Method

In a study, Ahmed (2014) included all daily aggregate investment flows made separately by individual and institutional investors in Qatar as samples. Similarly, Chung and Wang (2016) obtained all daily trading volume of each stock by individual investors from Korea Exchange as samples. Neither Ahmed nor Chung and Wang used probability sampling but purposive sampling, a nonprobability sampling. I adopted purposive sampling in this study too and included all TWSE trade transaction data of individual investors as the population. I implemented two exclusions. Chang et al. (2012) excluded a small number of stocks that market participants traded less than 20 days in 484 trading days from July 2006 to June 2008. Stocks traded in less than 4% of trading days in 1 year are not liquid. The stock prices may be less meaningful or even meaningless. I had the same exclusion of stocks traded in less than 4% of trading days in 1 year to avoid any jeopardy in herding detection. Lin et al. (2013) also excluded stocks with exchange sanctions for a more accurate herd behavior detection. I also excluded exchange sanctioned stocks. As such, the samples were the population excluding illiquid stocks and exchange sanctioned stocks. The samples consisted of a majority of the population.

Instrumentation

I encountered a challenge in deriving the LSV measure of Chang et al.'s (2012) version from Lakonishok et al.'s (1992) version. I brought in Wermers' (1999) version to connect the two. Lakonishok et al. (1992) developed the LSV measure and in the first place applied it to institutional investors only. Wermers enhanced the LSV measure by introducing the number of orders as an alternative to the number of investors in the formula. Chang et al. further enhanced the LSV measure by introducing a split between institutional and individual investors. I integrated the details about the LSV measure including its mathematical formula and considered the high applicability of the LSV measure by its extensive applications. I did not see any pertinent limitations of the LSV measure so found it appropriate for use in this study.

Lakonishok, Shleifer, and Vishny's Measure

Lakonishok et al. (1992) introduced the LSV measure to examine any disproportionate number of money managers buying or selling one stock in a given quarter. When half of the changes in holdings of individual stocks are increases and another half of the changes in holdings of individual stocks are decreases, there is no herding at the individual stock level. Alternatively, for example, when 70% of money managers increase their holdings of a particular stock but decrease their holdings of other stocks, the remaining 30% of money managers decrease their holdings of the same stock but increase their holdings of other stocks, then more money managers end up on the buy side of the market for the particular stock and the money managers have herded. The LSV measure represents the difference between an actual proportion and an expected proportion of *r*-type investor who net buys stock *i* at time *t*. Let B^r_{it} be the number of *r*-type investors who increase holdings of stock *i* at time *t*, and S^r_{it} be the number of *r*-type investors who decrease holdings of stock *i* at time *t*, and *t* be a quarter. The actual proportion, p^r_{it}, of *r*-type investors who net buy stock *i* at time *t* is:

$$p^r_{it} = B^r_{it} / (B^r_{it} + S^r_{it}). \tag{2}$$

With the actual proportion of *r*-type net buyer of stock *i* at time *t*, the LSV measure denoted by HM^r_{it} is as follows (Wermers, 1999):

$$HM^r_{it} = |p^r_{it} - E[p^r_{it}]| - E[|p^r_{it} - E[p^r_{it}]|]. \tag{3}$$

Wermers (1999) suggested estimating the expected proportion, $E[p^r_{it}]$, by dividing the number of purchase order across all stocks of one investor type by the total number of purchase and sale orders across all stocks of the same investor type in the same quarter. The expected proportion, $E[p^r_{it}]$, stays constant for all stocks in the same quarter. Wermers did not offer a similar suggestion of using the number of purchase orders and the number of sale orders of stock *i* by *r*-type investors at time *t* to estimate the actual proportion, p^r_{it}. The absence of such suggestion was probably due to unavailability of data at order level for each stock by investor type. Wermers followed Lakonishok et al.'s (1992) approach to estimate the actual proportion.

Chang et al. (2012) did not encounter the unavailability of data at order level for each stock by investor type in Taiwan market. Chang et al. considered B^r_{it} as the number of purchase order of stock i by r-type investors at time t while $S^r{}_{it}$ as the number of sale order of stock i by r-type investors at time t. The actual proportion, $p^r{}_{it}$, becomes the number of purchase orders of stock i by r-type investors divided by the total number of purchase and sale orders of stock i by r-type investors at time t. The use of the number of purchase and sale orders addresses the caveat that Wermers (1999) described. Wermers argued that, in Lakonishok et al.'s (1992) approach, funds report holdings were at the end of a calendar quarter. The change in holding between two quarters is unable to capture the funds that follow the trades of others in months or even weeks. The number of purchase and sale orders in an interval shorter than a quarter improves herding detection capability. Chang et al. used the number of purchase and sale orders from TWSE on a daily basis to detect herd behavior within a day.

$E[|p^r{}_{it} - E[p^r{}_{it}]|]$ in Equation 3 is an adjustment factor to illustrate the gap between $|p^r{}_{it} - E[p^r{}_{it}]|$ and its average in the long run. The adjustment factor becomes the expected value of $E[|B^r{}_{it} / (B^r{}_{it} + S^r{}_{it})|]$ under the null hypothesis of no herding (Lakonishok et al., 1992). The adjustment factor also takes into account the number of purchase and sale orders of stock i by r-type investors at time t. The adjustment factor follows a binomial distribution:

$$E\left[\left| p_{it}^r - E\left[p_{it}^r\right]\right|\right] = \sum_{k=0}^{n_{it}^r} \left| \frac{k}{n_{it}^r} - E\left[p_{it}^r\right] \right| C_k^{n_{it}^r} \left(E\left[p_{it}^r\right]\right)^k \left(1 - E\left[p_{it}^r\right]\right)^{n_{it}^r - k}, \qquad (4)$$

where $C_k^{n_{it}^r}$ denotes the number of possible combinations of selecting k number of stocks from a universe of $n^r{}_{it}$ number. The adjustment factor declines as the number of purchase and sale orders, $n^r{}_{it}$, of stock i by r-type investors at time t rises.

The LSV measure is a value between -1.0 and 1.0. When it is statistically significantly greater than zero, it implies an obvious propensity of the investors on accumulating trades on the purchase or sale side of the market. When it is not statistically significantly greater than zero, it implies no obvious propensity of the investors on accumulating trades on either side of the market.

LSV Measure upon Segregation of Stocks

Lakonishok et al. (1992) did not devise the LSV measure to show on the specific side, buying or selling, on which the strength of herding tends to accumulate. To achieve such purpose, Wermers (1999) applied the LSV measure upon segregation of stocks by whether the stocks had a higher or lower proportion of buyers than the average stock in the same quarter. The unconditional LSV measure, HM^r_{it}, is derived into two conditional herding measures, (a) buy-herding measure, BHM^r_{it}, and (b) sell-herding measure, SHM^r_{it} as follows:

$$BHM^r_{it} = HM^r_{it} \mid p^r_{it} > E[p^r_{it}], \tag{5}$$

$$SHM^r_{it} = HM^r_{it} \mid p^r_{it} < E[p^r_{it}], \tag{6}$$

where BHM^r_{it} is the number of buying orders of stock i by r-type investors as a proportion of all orders of stock i by r-type investors during period t; SHM^r_{it} is the number of selling orders of stock i by r-type investors as a proportion of all orders of stock i by r-type investors during period t.

The adjustment factor $E[|p^r_{it} – E[p^r_{it}]|]$ in Equation 3 is subject to the condition of $p^r_{it} > E[p^r_{it}]$ or $p^r_{it} < E[p^r_{it}]$ for BHM^r_{it} and SHM^r_{it} respectively under the null hypothesis of independent trading decisions. Averaging BHM^r_{it} (denoted by $\overline{BHM}$) and SHM^r_{it}

(denoted by $\overline{SHM}$) separately provides two means for assessing any herding by *r*-type investors. The former assesses any herding into stocks whereas the latter assesses any herding out of stocks. For example, when $\overline{SHM}$ is statistically significantly greater than $\overline{BHM}$, it means the *r*-type investors herd to sell than to buy. The two conditional herding measures are also useful for analyzing stock returns following buying versus selling by a herd.

LSV Measure Application in Previous Studies

Since the LSV measure development in the 1990s, researchers have been using and enhancing it for more than two decades. Choi and Skiba (2015) used a time-series average of the LSV measure and found institutional herding in 41 markets significant at 1% level. In ascending order, Nigeria was the lowest with the LSV measure at 3.13% while Ireland was the highest with the LSV measure at 12.61%. I referred to Choi and Skiba's LSV measure results as a yardstick despite the different investor type and the absence of Taiwan in Choi and Skiba's study. Lavin and Magner (2014) enhanced the LSV measure estimation with monthly instead of quarterly data. Lavin and Magner's estimation on Chile was 3.18%, 2.80% smaller than Choi and Skiba's estimation of 5.98%. Different data intervals may account for a part of the disparity in the LSV measure estimations. I estimated the LSV measure with daily data, hence noted any disparity between the results of quarterly data in previous studies. Chang et al. (2012) applied the LSV measure by investor type and found institutional herding more serious than individual herding in Taiwan market. The average LSV measure of institutional and individual investors were 16.3% and 4.7% respectively. I extended individual herding study to the front of characteristics of stocks that Jame and Tong (2014) did for the United States markets. I considered the high applicability of the LSV measure by its extensive applications.

Limitations of LSV Measure

There are two limitations of the LSV measure. First, it accounts for only the number of investors who buy or sell a particular stock, not for the trading volume, in the assessment of herding (Bikhchandani & Sharma, 2001). If the number of buyers and the number of sellers are close, and the sellers supply a small volume in the market, but the buyers collectively demand a big volume of stock and herd to bid, the LSV measure is not reflective of such herding. Although the TWSE's bid and ask orders data may account for bid- or ask-herding, I did not come across herd measures developed upon bid and ask orders. As such, the first limitation of not accounting for the trading volume was not relevant to this study. Second, the LSV measure is one herding assessment over a period regardless of the length of the period. The LSV measure is not for identifying herding in intertemporal trading patterns. Identification of herding in intertemporal trading patterns was not in the scope of this study. The second limitation was not relevant too. In brief, both limitations of the LSV measure were not pertinent to this study.

Data Collection Technique

I collected TWSE data, a secondary data, for this study. The data sources were (a) monthly trade transaction data files, (b) basic information of all stocks files, (c) monthly statistics files, and (d) a list of international securities identification numbers. I benefited from using secondary data in four aspects including a clear anticipation about the data nature in research design stage, having a reference for subsequent comparison, a higher confidence in the feasibility of the study, and a high cost efficiency. However, I inherited the limitation of data authenticity of secondary data but to a lessened extent.

Collection of Secondary Data

In the momentous herd behavior studies, Lakonishok et al. (1992) and Wermers (1999) did not collect primary data but secondary data. It signifies the sufficiency of secondary data for herd behavior studies. I collected secondary data with caution about a complete and reliable source. Lakonishok et al. used the data provided by SEI, a large consulting firm in financial services for institutional investors. Wermers used the database created by CDA Investment Technologies, Inc., of Rockville, Maryland. Furthermore, Chang et al. (2012) and Lin et al. (2013) used common secondary data, the intraday order data of TWSE. With Taiwan individual investors as the target population, I used TWSE as my data source too. TWSE is an order-driven market. Investors must submit orders into the stock exchange's system for call auction order matching before the market is open or for normal order matching when the market is open. As a result, TWSE data are official, complete, and reliable.

Monthly Trade Transaction Data File

TWSE avails monthly trade transaction data in its Chinese version website at the Data eShop option under the Products and Services drop-down. A monthly trade transaction data file captures the trade transactions of all stocks in a month. Monthly trade transaction data files of the latest 12 months are unavailable. I met the requirements of proposal oral defense in January 2018. December 2016 was the most recent month of which the monthly trade transaction data file was available. Given the 12 months data span, January 2016 was the first month of trade transaction data for this study. Hence, I collected the monthly trade transaction data files from January to December 2016. I followed the procedure stated in TWSE website to subscribe the data. In a monthly trade transaction data file, each row represents one trade transaction. For each trade transaction, there are 13 columns (see TWSE, 2004a). The 13 columns cover:

1. transaction date,
2. stock code,
3. transaction type that indicates buy or sell,
4. exchange code that indicates normal, huge volume, or odd volume transaction,
5. transaction time,
6. transaction record number,
7. broker's order number,
8. transaction price,
9. transaction share count,
10. broker's printer number,
11. order type,
12. investor type that indicates mutual funds, foreign investors, individual investors, and other institutional investors, and
13. broker code.

Basic Information of All Stocks File

TWSE avails basic information of all stocks in its English version website at the Data eShop option under the Products and Services drop-down. The basic information of all stocks of every trading day is available. The number of shares listed is one of the basic information. To estimate the market capitalization of a stock at the beginning of a month, I used the number of shares listed on the last trading day of the previous month. Therefore, I collected the basic information of all stocks file of last trading day of each month from December 2015 to November 2016. I followed the procedure stated in TWSE website to subscribe the data. In a basic information of all stocks file, each row represents one stock of a trading day. For each stock of a trading day, there are seven columns apart from the number of shares listed (see TWSE, 2004b). The eight columns cover:

1. trade date,
2. stock code,
3. stopping margin trading code,
4. stopping short sale code,
5. reference price,
6. closing price,
7. the number of shares issued, and
8. the number of shares listed.

Monthly Statistics File

TWSE avails monthly statistics of every stock in its English version website at the Listed Companies Monthly Statistics option under the Statistics of Market Information drop-down. A monthly statistics file consists of six columns (TWSE, 2018c). The six columns cover:

1. stock code,
2. name of the listed company,
3. latest price,
4. price-to-earnings ratio,
5. yield, and
6. P/B ratio.

In a monthly statistics file, each row represents one stock. Updated monthly statistics are available on the seventh trading day of a month. To link the latest P/B ratio of stock at the beginning of a month, I used the P/B ratio updated on the seventh trading day of the previous month. Therefore, I collected the monthly statistics files from December 2015 to November 2016. I followed the procedure stated in TWSE website to retrieve the data.

List of International Securities Identification Number

TWSE presents a list of international securities identification number (ISIN) code of listed equities in its English version website at the ISIN option under the Products and Services drop-down. The list consists of (a) security code and security name, (b) ISIN code, (c) date listed, (d) market, (e) industrial group, (f) classification of financial instruments (CFI) code, and (g) remarks (TWSE, 2018a). The list is, in fact, a table with industrial code and security code as two of the columns. I referred to the two columns to categorize the stock in other data sources by industry affiliation. The list of ISIN is updated every trading day. Outdated reference tables are not retrievable anymore. I collected a copy of the list of ISIN as of March 31, 2018.

Advantages and Limitations of Secondary Data

There were four advantages of using secondary data for this study. Secondary data are an utmost element with which a researcher can replicate a previous study and expect the same result (Frankfort-Nachmias et al., 2015). Chang et al. (2012) used TWSE data, a secondary data. Although I did not replicate Chang et al.'s study, I followed Chang et al.'s descriptions about TWSE data. I had a clear anticipation about the nature of data which I would collect at my research design stage. The clear anticipation of data nature was one advantage. If secondary data of the same sample at different times are available, a collection of such data forms longitudinal data (Frankfort-Nachmias et al., 2015). Chang et al.'s sampling period was July 2006 to June 2008 and my sampling period was January to December 2016. As there was no structural change in Taiwan stock market between the two periods, I could consider Chang et al.'s herding assessment result as a reference. The advantage of having a reference was that I could compare my result with the reference so as to have a sense about the correctness of my result. With a clear anticipation about the data nature and a reference, I developed a higher confidence in the feasibility of my study which was the third advantage. Furthermore, collecting secondary data is more cost efficient than collecting primary data (Frankfort-Nachmias et al., 2015). For this study, I subscribed TWSE data whose cost was affordable to me. If I had to incentivize hundreds of thousands of individual investors to reveal their stock trading data, the total cost would be way higher than I could afford. The high cost efficiency of secondary data was the fourth advantage. In summary, I benefited from using secondary data in four aspects including a clear anticipation about the data nature in research design stage, having a reference for subsequent comparison, a higher confidence in the feasibility of the study, and a high cost efficiency.

There was one limitation of using secondary data for this study. The major limitation of secondary data is data authenticity (Frankfort-Nachmias et al., 2015). Secondary data are usually a processed version of the raw data collected. A researcher usually does not have an access to the raw data collected, so any data validation that the researcher does is fundamentally limited. The validity of any research using secondary data is fundamentally limited too. TWSE data are no exception but likely with the limitation of data authenticity to a lessened extent because of their extensive use. TWSE has been operating for 57 years since its establishment in 1961 and is ranked the 19th in the globe. TWSE is an order-driven, completely computerized market. Bid and ask orders, buy and sell transactions, a full array of prices and dates are systematically captured or generated. TWSE's matured and sophisticated system likely generates secondary data of a reliable standard. Another limitation of secondary data is code variation (Frankfort-Nachmias et al., 2015) that might happen to TWSE data. The current code of a data field may denote finer or more precise than the old code of the same data field. I did not see any TWSE announcements about code variations of the data fields related to the three independent variables and dependent variable of this study, therefore code variation was not relevant to this study.

Data Analysis Plan

After data collection, I combined all files into a master data file. I sampled the trade transactions of all stocks except new issues and exchange sanctioned ones. I estimated the LSV measure and performed a *t* test on it for each stock-day. I captured the test result as positive or no in the herd behavior indicator, the dependent variable. I ran a logistic regression with the three independent variables against the dependent variable. I interpreted not only the logistic regression model result but the validity and reliability of the entire study.

Master Data File Creation

I created a master data file by combining the three secondary data sources, (a) monthly trade transaction data files, (b) basic information of all stocks files, and (c) monthly statistics files. I used the monthly trade transaction data file as the base and expanded it row by row. For example, there was a trade transaction in January 2016. I used the stock code as a matching key to link the closing price and the number of shares listed from the basic information of all stocks file of the preceding month, December 2015 in this case. I used the stock code again as a matching key to link the P/B ratio and industry code from the monthly statistics file of the preceding month, December 2015 again in this case. In the same manner, I expanded the monthly trade transaction data file for each of the remaining 11 months in the data span. At last, I combined the 12 expanded monthly trade transaction data files into a master data file.

Following Wermers' (1999) practice, there should not be new issues less than 1 year since first offering dates for analysis. I excluded new issues with first offering dates in or after January 2015 from the master data file. I also excluded all stocks delisted between January and December 2016 from the master data file. I had to exclude illiquid stocks, however there was a debate about the minimum trading activity required to generate a meaningful herding measure result (Ortiz et al., 2013). One theoretical argument is that more investors who trade probably lead to higher herding (Bikhchandani et al., 1992). Another theoretical argument is that herd can happen between as few as two investors (Wermers, 1999). Chang et al. (2012) suggested a pragmatic view that too few investors with trades probably lead to an extremely high herding measure result. Moreover, it is hard to generalize the *herd behavior* from too few investors with trades. Setting the minimum number of investors with trades was necessary. Wermers suggested examining institutional herding in stock with at least five active money managers in a quarter. Barber et al. (2009) suggested examining individual herding in one stock with at least 10 individual investors in a week. Barber et al.'s suggestions were more relevant to this study, because the target populations of both studies were individual investors. Nevertheless, there was a difference in time unit. The time unit in Barber et al.'s study was a week whereas that of this study was a day. The minimum number of investors with trades is constant regardless of the time unit. I excluded the stocks traded by fewer than 10 individual investors in a day.

The unit of analysis for this study was an individual investor. I identified all trade transactions of individual investors by referring to the column of investor type. TWSE data are anonymous. The data do not consist of any unique identifier of the investor. In the absence of unique identifier of investor, I was not able to identify and link up transactions of the same individual investor. Chang et al. (2012) and Lin et al. (2013) should have encountered the same situation of the inability of linking up transactions at the individual investor level. Both groups of authors did not explain how to deal with the situation. Both groups of authors implicitly assumed that every transaction belonged to a unique individual investor in herding measure estimations. In the United States which is another market, Jame and Tong (2014) implicitly assumed the same with the transaction data. Under such assumption, Jame and Tong might get an overstated number of individual investors buying and an overstated number of individual investors selling because, in reality, there should be at least some investors who bought or sold the same stock more than once. The overstatements would happen to this study too. As long as the repeated buying or selling happened randomly, the two overstatements would zero out each other. There would not be a distortion on the herding measure estimation. I held the same assumption of the uniqueness of individual investor for every transaction in a day for this study.

Independent Variable Development

The market capitalization of stock at the beginning of a month was one independent variable. I multiplied the closing price and the number of shares listed, two columns in the master data file, to produce it by stock-month. There were 12 market capitalizations for each stock. Yao et al. (2014) ranked all stocks of a month by market capitalization and constructed quintile portfolios. The one fifth of stocks ranked with the lowest market capitalizations were in the first quintile. The one fifth of stocks ranked just above the lowest market capitalizations were in the second quintile, and so on. I sorted all stocks by market capitalization in ascending order and by stock-month then divided them by quintile. The quintiles were low, mid-low, middle, mid-high and high market capitalization. P/B ratio at the beginning of a month was another independent variable. I used the P/B ratio column in the master data file by stock-month. There were 12 P/B ratios for each stock. P/B ratio was a continuous variable. The industry affiliation was the last independent variable. I referred to the industrial group in the master data file to classify each stock in one of the 28 industrial groups by stock-month.

LSV Measure Estimation and Herd Behavior Indication

With the specification of the master data file, I defined the parameters in Equation 2. Investor *r*-type was individual. Stock *i* was each stock listed in TWSE. Time *t* was a day. B^r_{it} was the number of buy transactions of stock *i* by individual investors in a day. S^r_{it} was the number of sell transactions of stock *i* by individual investors in the same day. p^r_{it}, was the proportion of buy transactions out of the sum of buy and sell transactions of stock *i* by individual investors in a day. I calculated the LSV measure for each stock-day in the master data file. The LSV measure is a value from -1.0 to 1.0. Then, I performed a *t* test on the LSV measure result. When statistically greater than zero, it meant positive herd behavior on either purchase or sale side of the market. When not statistically greater than zero, it meant no herd behavior on both sides of the market. I created the herd behavior indicator which was dichotomous at stock-day level. I set the herd behavior indicator to positive or no accordingly.

Level of Measurement

A researcher defines variable measurements based on three concepts: numerals, assignments, and rules (Frankfort-Nachmias et al., 2015). Market capitalization, P/B ratio, and industry affiliation of stock were the three independent variables for this study. Market capitalization was originally at ratio level of measurement. I derived another market capitalization at ordinal level for the sample division by quintile. The ordinal level of measurement indicates directions but has an uneven spacing. The numerals were one, two, three, four, and five. The assignments were from one as the lowest to five as the highest. The rule was to sort all stocks in ascending order, then divide them into quintiles. The quintile of the lowest market capitalization was one. The next quintile was two, and so on. P/B ratio was also originally at the ratio level of measurement. I derived another P/B ratio for the sample division by quintile. The numerals were one, two, three, four, and five. The assignments were from one as the lowest to five as the highest. The rule was to sort all stocks in ascending order, then divide them into quintiles. The quintile of the lowest P/B ratio was one. The next quintile was two, and so on. The industry affiliation was at the nominal level of measurement. I defined its numerals, assignments, and rules according to the list of ISIN, for example, 1101 for cement, 1213 for food (TWSE, 2018a). The numerals were for classification purpose only.

The herd behavior indicator was the dependent variable. Its level of measurement was binominal. The numerals were one and zero. The assignments and rules were related to the *t*-test result of the LSV measure which was Wermers's (1999) practice. Key assumptions about the parameters of a population are: (a) variables are at least on an interval scale and (b) samples are from a normally distributed population. Chang et al. (2012) found normal distribution but skewness and kurtosis in sell-herding institutional investors and buy-herding individual investors. Although the assumption of normality did not entirely hold for this study, I performed a *t* test on the LSV measure for each stock-day. I captured the *t*-test result in the herd behavior indicator. The assignments and rules were one and zero respectively for positive and no herd behavior. The LSV measure is the difference between the actual proportion and the expected proportion of buy transactions out of the total buy and sell transactions of one stock by individual investors in a day with an adjustment factor. The numerals of the LSV measure were from -1.0 to 1.0 with a natural zero. The LSV measure was at ratio level of measurement. The assignments and rules were as Equation 2 to Equation 4 set forth. The numerals, assignments, and rules of the buying and selling LSV measures were similar except the replacement of Equation 3 with Equation 5 and Equation 6 for the buying and selling LSV measures respectively.

Statistical Test

I examined any statistically significant difference in herd behavior by each of the three characteristics of stocks: market capitalization, P/B ratio, and industry affiliation through a chi-square test. Chi-square test is a statistical method suitable for examining any association between two categorical variables (Field, 2013). The herd behavior indicator indicated an occurrence of herd behavior or not and was a dichotomous, therefore categorical, variable. The derived market capitalization and derived P/B ratio indicated the respective characteristics in quintiles which were categorical.

Also, I examined the relationship between herd behavior and the three characteristics of stocks as a whole through logistic regression as specified in Equation 1. Logistic regression is a statistical method suitable for examining the relationships between a dependent variable and one or more independent variables. The dependent variable is dichotomous with only two possible outcomes. The independent variables can be dichotomous, nominal, or continuous. Logistic regression is to choose the parameters which maximize the likelihood of observing the sample values, rather than minimize the sum of squared errors as in ordinary regression (Field, 2013). The goal of Equation 1 was to find the best coefficients to describe the relationship between the herd behavior indicator and market capitalization, P/B ratio, and industry affiliation as a whole.

$$\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{1}$$

where

X_1 is the market capitalization of a stock and a continuous variable,

X_2 is the P/B ratio of a stock and a continuous variable,

X_3 is the industry affiliation of a stock and a nominal variable, and

logit(p) is the odds ratio of the positive herd behavior in logarithm form, ln[p/(1-p)], and p is the actual proportion of positive herd behavior.

I used SAS, the statistical analysis software, in Oracle Virtual Machine environment to process the files from all data sources and create the master data file. I estimated the LSV measure for each stock-day. I created and set the herd behavior indicator according to the *t*-test result of the LSV measure of each stock-day. I ran chi-square tests and a logistic regression. For the chi-square tests, the significance level was at 0.05.

For the logistic regression, I used Wald tests to examine independent variables. The significance level was at 0.05. Furthermore, I used Nagelkerke R^2 to assess the goodness of fit of the logistic regression model as a whole.

Validity and Reliability of LSV Measure

In a quantitative research, a researcher is seldom completely certain about a measuring variable in the best way to reflect a true picture (Frankfort-Nachmias et al., 2015). The same situation applied in regards to the LSV measure. Based on the previous studies, I discussed the validity of the LSV measure in three aspects, construct, content, and empirical. Also, I discussed the reliability of the LSV measure.

Construct validity. Construct validity is the logical and empirical tie of a measuring instrument to both the concept and assumptions of a theoretical framework (Frankfort-Nachmias et al., 2015). There was no measuring instrument like questionnaire but the LSV measure for this study. The construct validity of LSV measure was important. The rational view of intentional herding postulates market participants' revaluation and trades of stock after other market participants' trades. The trades of any market participants are the most direct evidence of behavior. Wermers (1999) suggested the interchangeability between buyers and buying transactions in the LSV measure. Intrinsically, there is a logical tie between the LSV measure and the concept and assumptions of herding theory. I discussed the empirical tie in the empirical validity section.

Content validity. A measuring instrument is content valid when it covers all attributes of a concept and leaves nothing relevant out (Frankfort-Nachmias et al., 2015). Sampling validity and face validity are the two components of content validity. Sampling validity refers to the extent which the domains in a sourced or developed measuring instrument represent the exhaustive list of domains of a concept. Again, there was no instrument like questionnaire but the LSV measure for this study. The exhaustive list of this study consisted of only two domains: herding presence and extent. The LSV measure addresses both domains, therefore is sampling valid. Face validity depends on researchers' opinion on the appropriateness of the instrument for measuring a concept. I considered the LSV measure face valid by its direct measure of herd behavior.

Empirical validity. Empirical validity is a strong relationship between a measuring instrument and its measured outcome (Frankfort-Nachmias et al., 2015). Among various tests to evaluate the empirical validity, predictive validity is the most common (Frankfort-Nachmias et al., 2015). In a predictive validity test, researchers obtain results from other measuring instruments as external criteria. The researchers compare the outcomes of their measuring instrument with the external criteria. For this study, the herd behavior measure was the LSV measure. Chang et al. (2012) used the same measure but did not report its predictive validity. I leveraged C. Y. Chang et al's study as the base and compared its finding with Lin et al.'s (2013) which were external criteria. Lin et al. used another herding measure by bootstrap run on the same target population. Notwithstanding the two different sampling periods, the consistent finding was a higher institutional herding over individual herding. Lin et al.'s study, the external criteria, provided a consistent finding that supported the predictive validity which implied the empirical validity of the LSV measure.

Reliability. Reliability refers to the extent of result consistency that a measuring instrument or procedure yields in repeated trials. Standard deviation or variance of the measured variable reflects result consistency. The smaller the standard deviation or variance, the more reliable is the measuring instrument or procedure. There was no questionnaire in this study, therefore respondents' momentary distractions in completing a questionnaire, one potential cause of variance, was irrelevant. Another potential cause of variance, biased sample, was minimal because of the inclusion of the majority of the population. Overall, the reliability of the LSV measure was acceptable.

Handling Methodological Limitation

I evaluated internal and external validity by each of their potential limitations. For internal validity, there are five potential limitations, namely selection effect, regression artifact, history, maturation, and experimental mortality. I did not recruit individual investors as samples. I did not need to set up and assign individual investors to experiment and control groups either. There was no selection per se, therefore selection effect was not relevant to this study. I used purposive sampling to exclude illiquid stocks and exchange sanctioned stocks only. The samples, in effect, consisted of the majority of the population. Hence, regression artifact was scarcely relevant to this study. History referred to any events happened in Taiwan stock market in 2016 including the measures implemented by Taiwan Securities and Futures Bureau (2018a). Some individual investors had to have changed their trade behavior as a result of the events, so history was relevant. Maturation referred to a natural process through which individual investors got more mature in the sampling period. Some individual investors had to have got more mature in 2016, therefore maturation was relevant. Experimental mortality referred to the drop-outs of individual investors in the sampling period. Some individual investors had to have traded in early 2016 but not anymore in late 2016. Such individual investors were the drop-outs, so experimental mortality was relevant.

For external validity, there are two potential limitations, namely reactive arrangement and representativeness of the sample. Reactive arrangement was irrelevant. All investor trades in the stock exchange were authentic actions. There was no experimental setting in this study for investors to react, therefore reactive arrangement was irrelevant. Representativeness of the sample was high. The samples of this study consisted of the majority, 94%, of the population. Illiquid or exchange sanctioned stocks were the only portion purposefully unrepresented, so representativeness of the sample was high.

I also evaluated reliability which meant the consistency of the LSV measurements in this study. I estimated the LSV measures by the standard mathematical formula of the LSV measure and the static TWSE data. The LSV measure results were consistent regardless how many times of estimations. Hence, the reliability of this study was high. I evaluated generalizability of this study to the population, a different time, or a different market. In this study, the samples consisted of the majority, 94%, of the population. The findings from such a majority were already generalized and should be generalizable to the individual investors of Taiwan. Besides, the data span of 1 year was long enough to capture many variations of the market. The findings from such variations were probably generalizable to certain months or years before and after the sampling period. Lastly, the findings of more serious herding in either high market capitalization or high P/B ratio stocks were already generalized between Taiwan and China and probably generalizable to different markets. Overall, the generalizability of this study could be to the population, a different time, and a different market.

Lastly, I evaluated two tiers of limitations, primary and secondary, of the use of secondary data. All three primary limitations, a gap between research purposes, limited data accessibility, and information insufficiency, were not relevant. Code variation of the secondary limitation was not relevant either. I evaluated the limitations specific to TWSE data. Data not connectable to individual investor level was a limitation to computing the actual number of buyers and actual number of sellers. Nonnormality of the LSV estimates, even after Box-Cox transformation, was another limitation of TWSE data.

There was a collective limitation of history, maturation, and experimental mortality to the internal validity but negligible limitations to the external validity. The overall validity was typical for causal-comparative design. In addition, there were negligible limitations to the reliability, generalizability, and use of secondary data. The limitations specific to TWSE data were the biggest which I mitigated with the adoptions of an implicit assumption. The overall limitations of this study were typical and mitigable, therefore acceptable.

Ethical Consideration

In the standard application form for research ethics review (Walden University Institutional Review Board, 2015), I declared to follow the procedures of analysis of existing public records or documents. TWSE is a quasi-governmental organization regulated by Financial Supervisory Commission R.O.C. (Taiwan). TWSE data were open for subscription, therefore public. I perused all terms and conditions about the use of TWSE data from data access control to data transferability between computers. I implemented measures to ensure full compliance. Also, I declared not applicable to the level of risk for all potential risk items in Section 3 in the standard application form. I opted to collect data in an anonymous way that would contain absolutely zero identifiers and would make it impossible to determine who participated and who did not in Question 25 under Section 4: data integrity and confidentiality. I declared no potential conflicts of interest in Section 5. According to Section 6: data collection tools, I asked Dr. Lakonishok, the primary author of the LSV measure, for the permission of my use of the LSV measure. Afterward, I received Dr. Lakonishok's permission through an email. After proposal approval, I subscribed TWSE data and received full terms and conditions about the use of TWSE data. TWSE did not state any requirements for destruction of the subscribed data. I set 7 years after my Ph.D. study publication as the data keeping duration which will meet Walden University minimum requirement of 5 years.

Summary

The overarching research question of this study was what differences in individual investor herd behavior, if any, existed by the three characteristics of the stocks, market capitalization, P/B ratio, and industry affiliation. All the three characteristics of stocks were not manipulable and herd behavior was ex-post facto; therefore, the causal-comparative design of quantitative method was suitable. The unit of analysis was the individual investor. All individual investors who traded between January and December 2016 in Taiwan constituted the target population. By purposive sampling, I included the target population after exclusion of the trade transactions of illiquid stocks and exchange sanctioned stocks. The samples consisted of the majority of the population. TWSE was the secondary data source. I aggregated trade transactions to estimate the LSV measure for each stock-day. Then, I performed *t* tests to assess herd behavior by stock-day. I captured positive or no herd behavior in the herd behavior indicator, the dependent variable. In next chapter, I will describe the data collection process covering the derivations of the independent and dependent variables in detail. I will also describe the results of the statistical tests between the three characteristics of stocks and herd behavior.

Chapter 4: Data Collection

Introduction

The purpose of this quantitative research in causal-comparative design was to examine individual investor herd behavior as related to the characteristics of stocks in Taiwan stock market. The characteristics of stocks included (a) market capitalization, (b) P/B ratio, and (c) industry affiliation. The overarching research question was what differences in individual investor herd behavior, if any, existed by the characteristics of stocks. I developed the following three research questions and corresponding hypotheses to guide this study:

RQ1: What differences in individual investor herd behavior, if any, exist by market capitalization of stock?

$H_0$1: There is no statistically significant difference in individual investor herd behavior by market capitalization.

H_a1: There is a statistically significant difference in individual investor herd behavior by market capitalization.

RQ2: What differences in individual investor herd behavior, if any, exist by P/B ratio of stock?

$H_0$2: There is no statistically significant difference in individual investor herd behavior by P/B ratio.

H_a2: There is a statistically significant difference in individual investor herd behavior by P/B ratio.

RQ3: What differences in individual investor herd behavior, if any, exist by industry affiliation of stock?

$H_0$3: There is no statistically significant difference in individual investor herd behavior by industry affiliation.

H_a3: There is a statistically significant difference in individual investor herd behavior by industry affiliation.

In this chapter, I will describe my data collection process from subscription to receipt and from integration into a master data file to reconciliation against TWSE official reports. With the master data file, I was able to address the research questions. I will also interpret the estimated LSV herd measure for each stock-day. The chapter will include a discussion of the hypothesis tests of general, buy-, and sell-herding separately and how based on the hypothesis test results, I assessed the needs for post hoc analysis and multivariate analysis. I will also discuss the characteristics and suitability of the logistic regression method and present the logistic regression model of herd behavior against market capitalization, P/B ratio, and industry affiliation as well as assess its statistical significance. Lastly, I will estimate the expected herd behavior at stock level by trading day.

Data Collection

TWSE was the only source of data for this study. I collected trade transaction data from January 2016 to December 2016, basic information of all stocks from December 2015 to November 2016, monthly statistics from December 2015 to November 2016, and the list of ISIN as of March 31, 2018. I integrated the collected data into a master data file to which I reconciled TWSE official reports.

Data Integration

I subscribed to and received two sets of data from TWSE. The first set was the monthly trade transaction data file from January 2016 to December 2016, and the second was the basic information of all stocks file from December 2015 to November 2016. Without the need of subscription, I downloaded the third set of data and a Status of Securities Listed on TWSE report from TWSE website. The third set was the monthly statistics file from December 2015 to November 2016. The Status of Securities Listed on TWSE report was issued December 2016.

The Worksheet 7 of the Status of Securities Listed on TWSE (2018d) report exhibited all the stocks listed in a month, and I referred to all the stocks on this worksheet of December 2016 as a basis. The stock number of a common stock was four digits and that of a preferred stock was four digits followed by an letter. With my reference to the stock number as the basis, I extracted transactions of only stocks from each of the monthly trade transaction data files that consisted of other securities such as convertible bonds, government bonds, warrants, and so on. The Worksheet 7 of 4 different months showed a blank transaction amount for six stocks. Accordingly, I did not extract those six stocks over those 4 months. During the extraction, I counted the number of transactions and also multiplied the transaction price and transaction share count to get a transaction amount. I aggregated the transaction amount by transaction date; stock code; transaction type, which indicated buy or sell; and investor type. I extracted the aggregate data of individual investors with the criterion of the investor type, I, into a subset. The subset data were at the stock-day level. Furthermore, the Worksheet 7 presented all stocks by industry affiliation, so I compared the industrial code and security code between the Worksheet 7 and the list of ISIN code for listed equities which I collected as of March 31, 2018. I found consistency across the two sources and adopted such an industry affiliation for each stock in the subset data.

In tandem, I developed two lists of stock for exclusion. I identified the stocks newly listed since January 2015 from the latest listed companies (see TWSE, 2018e) shown on the TWSE website. I also identified the stocks de-listed in or after January 2016 from the de-listed companies (see TWSE, 2018f) shown on TWSE website. From the subset stock-day data, I excluded the stock shown in either list. Then, I excluded all preferred stocks and the stock-day data with less than 10 trade transactions. Lastly, I identified the stocks that had been traded for less than 10 days out of 244 trading days in 2016 and excluded them.

I used the basic information of all stocks file of the previous month to produce the market capitalization for stock in a current month. I multiplied the closing price of a stock on the last trading day of the previous month with the number of shares listed. There were 32 stocks not traded on the last trading day of the previous month. I used the closing price of each of the 32 stocks on its last trading day within the previous month for market capitalization estimation. Similarly, I used the monthly statistics file of the previous month to extract the P/B ratio for stock in a current month. There were 13 stocks not reported in the previous month, so I used the P/B ratio of the month before the previous month for each of these stocks. I merged the market capitalization and the P/B ratio of the previous month with the subset stock-day data of a current month at stock level.

Data Reconciliation

Although I collected the monthly trade transaction data files directly from an official source, I reconciled the data before data analysis. I summed up the transaction amounts of all stocks by transaction month. There were 12 monthly transaction amounts, and again, I referred to the Worksheet 7 in the Status of Securities Listed on TWSE (2018d) report of the corresponding month as the basis. For each month, the summed transaction amount was 0.00% different than the basis. Therefore, I considered the monthly trade transaction data complete. The investor type in the monthly trade transaction data file was the focus of my next data reconciliation.

The worksheet named Form3 of the Statistics of Securities Market file (see TWSE, 2018b) reported the percentage of trading value by investor types. All individual investors were covered by two types: domestic individuals and foreign individuals. Throughout the 12 months I reviewed, there was 0.0% of trading value by foreign individuals. Domestic individuals were the sole contributors. I referred to the Form3 of the Statistics of Securities Market file as the basis. There were 12 monthly percentages of trading value by individual investors. I broke the summed transaction amounts of stocks by transaction month further down by investor type and computed the percentage contributed by individual investors. I performed *t* tests for two sample means assuming unequal variances. The percentages of trading value by individual investors between the two data sources, the basis (M = 0.52, SD = 0.02) and summed transaction amounts (M =

0.52, *SD* = 0.02), were not statistically significantly different at .05 significance level, t = 0.32, p = .75, so I considered the investor type in the monthly trade transaction data file complete and accurate. Then, I extracted the stock-day data of individual investors only into a subset.

The subset consisted of 209,251 stock-day data. The number of stock-day was smaller than 215,940 stock-days, the product of 244 trading days multiplied with 885 stocks. There were situations, namely temporary suspensions, new listings, and de-listings, under which stock was not subject to exchange every trading day in a year. New listings and de-listings were two situations whose stock-day data were not for my analysis. I excluded 8,099 stock-day data for newly listed stocks, 712 stock-day data for de-listed stocks, 569 stock-day data for preferred stocks, and 3,240 stock-day data for fewer than 10 individual investors in a trading day. After all exclusions, 196,631 stock-days or 94% of the subset total constituted the master data file for this study.

Measure and Test Results

Based on the master data, I described the characteristics of the 196,631 stock-day samples. I estimated the LSV measure and interpreted its result for each stock-day to determine whether individual herding occurred or not at the stock-day level. With determining whether the individual herding occurred or not at the stock-day level, I addressed the three research questions. I then performed hypothesis tests about general, buy-, and sell-herding separately and interpreted the hypothesis test results and the needs for post hoc analysis and multivariate analysis.

Descriptive Characteristics of the Sample

From the master data, the mean number of buy transactions per stock-day was 494 (*SD* = 1,128), whereas the mean number of sell transactions per stock-day was 495 (*SD* = 1,130). The mean transaction amount on the buy side per stock-day was NT$41,578,935 (*SD* = 144,279,968) or US$1,237,984 (*SD* = 4,295,837) at the 2016 exchange rate of NT$33.586 to US$1 (see Internal Revenue Service, 2018), whereas the mean transaction amount on the sell side per stock-day was NT$43,007,571 (*SD* = 147,539,110) or US$1,280,521 (*SD* = 4,392,875). The P/B ratio intervals with stock-days in quintiles were 0.01–0.69, 0.70–0.91, 0.92–1.25, 1.26–1.85, and 1.86–13.59. The intervals of market capitalization in millions of NT dollar with stock-days in quintiles were 1–2,059, 2,060–4,265, 4,266–7,749, 7,750–19,442, and 19,443–4,887,876. There were 28 industry affiliations. Electronic parts/components, the most common industry, consisted of 90 companies. Glass and ceramic, the least common industry, consisted of five companies. Table 1 shows the number of companies by industry affiliation in the master data file.

Table 1

Number of Companies by Industry Affiliations in 2016

Industry code and affiliation	#	Industry code and affiliation	#
01 – Cement	7	17 – Financial and insurance	33
02 – Food	21	18 – Trading and consumers’ goods	17
03 – Plastic	23	20 – Others	45
04 – Textile	45	21 – Chemical	25
05 – Electric machinery	46	22 – Biotechnology and medical care	26
06 – Electrical and cable	15	23 – Oil, gas, and electricity	8
08 – Glass and ceramic	5	24 – Semiconductor	65
09 – Paper and pulp	7	25 – Computer and peripheral	54
10 – Iron and steel	27	26 – Optoelectronic	66
11 – Rubber	11	27 – Communications and Internet	36
12 – Automobile	6	28 – Electronic parts/components	90
14 – Building material & construction	49	29 – Electronic products distribution	20
15 – Shipping and transportation	20	30 – Information service	12
16 – Tourism	14	31 – Other electronic	34

Herd Measure Estimation

I used the LSV measure to assess individual herding in this study. The number of buyers and number of sellers were the two key elements to the LSV measure. Nevertheless, TWSE did not release transaction data along with other information that would have enabled transaction consolidation at an investor level. I could not derive the actual number of buyers or sellers. I adopted the same approach as Chang et al. (2012)

and Lin et al. (2013) of applying the number of buy transactions and number of sell transactions instead in the LSV measure. The implicit assumption was that each transaction belonged to a unique individual investor.

In the master data file, there were 97,199,197 buy and 97,360,922 sell transactions by individual investors. Under the implicit assumption, the two figures became the numbers of individual investors who bought and sold respectively. There is a term of expected proportion, $E[p^r{}_{it}]$, of net buyer in the formula of the LSV measure. I estimated the expected proportion by dividing the number of buy transactions across all stocks, 97,199,197, by the total number of buy and sell transactions across all stocks, 194,560,119. The expected proportion was 49.96%, close to 50.00%. There were almost equal numbers of buy and sell transactions. I applied the expected proportion, $E[p^r{}_{it}]$, 49.96% for all subsequent LSV measure estimations.

There is a term of adjustment factor in the formula of the LSV measure. Among the five operands of the adjustment factor, the operand, $C_k^{n_{it}^r}$, denotes the number of possible combinations of k investors out of a universe of $n^r{}_{it}$. The adjustment factor is a summation for k from 0 to $n^r{}_{it}$. When k equals to ($n^r{}_{it}$ / 2), the number of combinations is a maximum. I encountered computation overflow problems on $C_k^{n_{it}^r}$ with SAS in Oracle Virtual Machine environment. There were notes of invalid argument to the function of computing combinations, COMB, in SAS when $n^r{}_{it}$ was larger than 1,029. The maximum was 1.429821×10^{308}. However, there were usually more than 1,029 individual investors trading one stock in a day. For example, 4,889 individual investors traded Taiwan Cement Corporation on January 4, 2016.

To circumvent computation overflow, I used the two operands, $E[p^r_{it}]^k$ and (1-$E[p^r_{it}])^{(n-k)}$, in the formula of the LSV measure. $E[p^r_{it}]$ and (1-$E[p^r_{it}]$) were 49.96% and 50.04% respectively. Both were less than one. A recursive multiplication of each operand would converge to zero. I leveraged the convergence to zero property of the two operands to offset the potentially huge number from $C_k^{n^r_{it}}$. I used the LCOMB function in SAS to take the logarithm form of $C_k^{n^r_{it}}$. Besides, I turned the two operands into k×ln(0.4996) and $(n^r_{it}-k)\times\ln(0.5004)$. The three operands in logarithm form were manageable. I added them and raised the sum to exponentiation for further computation. There was no more computation overflow problem.

I estimated the LSV measure, HM^r_{it}, with *r* already defined as individual, *i* set to stock level, and *t* set to per day. As such, the LSV measure was of the individual investor at stock-day level. Based on the LSV estimates of a group of stocks, Wermers (1999) produced and tested a mean LSV estimate. Similarly, I produced a mean LSV estimate of the 196,631 stock-days in 2016. The mean LSV estimate was 0.04 (SD = 0.08). I also estimated conditional LSV measures, BHM^r_{it} and SHM^r_{it} with the parameters, r, i, and t set identically. I segregated the stock-day data by the condition of whether the percentage of buy transactions out of total buy and sell transactions of a stock-day was bigger than the expected proportion, 49.96%, of the master data or not. There were 106,883 stock-days bigger and 89,748 stock-days smaller for buy- and sell-herding assessments respectively. The mean LSV estimate of buy-herding was 0.03 (SD = 0.08) whereas that the mean LSV estimate of sell-herding was 0.05 (SD = 0.08). With the LSV estimate at the stock-day level, I proceeded to perform statistical tests and interpret the herd measure estimate.

Interpretation of Herd Measure Estimate

For each of the 196,631 stock-days in the master data file, I performed a *t* test on the LSV measure estimate. The hypotheses were as follows:

H_0: Individual investors do not herd to either buy or sell the stock.

H_a: Individual investors herd to either buy or sell the stock.

If the LSV measure estimate was statistically significantly bigger than zero, I would reject the null hypothesis in favor of the alternative hypothesis. I would conclude that individual investors herded. Otherwise, individual investors did not herd. Given the directionlessness of the LSV measure between buy and sell, I focused on one side rather than two in the *t* test. Although the master data did not consist of illiquid stocks, the minimum number of transactions was 10 for a stock-day. I set the *t* test at 0.01, a higher significance level, to identify as accurately as possible the stock-days with herding. There were 110,237 stock-days in each of which individual investors herded and 86,394 stock-days in each of which individual investors did not herd. The percentage of stock-day in which individual investors herded was 56.11%. Individual investors herded in more than half of the stock-days. The mean LSV estimate of all stock-days was statistically significantly bigger than zero ($t(196{,}630) = 210.62$, $p < .0001$). Overall, individual investors herded.

Among the 106,883 stock-days for buy-herding assessment, there were 55,047 stock-days in each of which individual investors herded. There were 51,836 stock-days in each of which individual investors did not herd. The percentage of stock-day in which individual investors herded to buy was 51.50%. The mean LSV estimate of buy-herding was statistically significantly bigger than zero (t(106,882) = 122.06, $p < .0001$). Therefore, individual investors herded to buy. Among the 89,784 stock-days for sell-herding estimation, there were 55,190 stock-days in each of which individual investors herded. There were 34,558 stock-days in each of which individual investors did not herd. The percentage of stock-day in which individual investors herded to sell was 61.49%. The mean LSV estimate of sell-herding was statistically significantly bigger than zero (t(89,783) = 181.36, $p < .0001$). Therefore, individual investors herded to sell too.

Adoption of Chi-Square Test

Before the statistical tests of the three research questions, I assessed normality of the LSV estimates. The distribution of the 196,631 stock-days seemed in a bell shape but was not normal (Kolmogorov-Smirnov statistics = 0.13, $p < .01$). The stock-days were positively skewed (moment coefficient of skewness = 1.36) and platykurtic (moment coefficient of kurtosis = 2.10). The distribution of the subset stock-days for buy-herding assessment was not normal (Kolmogorov-Smirnov statistics = 0.13, $p < .01$). The distribution of the subset stock-days for sell-herding assessment was not normal either (Kolmogorov-Smirnov statistics = 0.14, $p < .01$). It was not proper to proceed to more complex parametric statistical tests with nonnormal data. I applied Box-Cox transformation over the estimates of LSV measure. The transformation was optimal when the lambda was -5 within the range of -5 to 5. I assessed normality of the transformed LSV estimates. The distribution of the 196,631stock-days was still not normal

(Kolmogorov-Smirnov statistics = 0.05, $p < .01$). The transformed LSV estimates were not skewed (moment coefficient of skewness = 0.00) but still platykurtic (moment coefficient of kurtosis = 0.25). Given the nonnormality of data, I resorted to chi-square test for the subsequent hypothesis tests. Chi-square test is a nonparametric test whose assumptions are fewer.

Hypothesis Test Result

General herding. The first research question was whether a difference in individual investor herd behavior existed by market capitalization of stock. I divided all stock-days by market capitalization in quintiles. I performed a chi-square test for the independence of the percentage of stock-day in which individual investors herded and the market capitalization in quintiles. The hypotheses were as follows:

$H_0 1$: There is no statistically significant difference in individual investor herd behavior by market capitalization.

$H_a 1$: There is a statistically significant difference in individual investor herd behavior by market capitalization.

If the chi-square statistic was bigger than 9.49, the critical value of the chi-square statistic with four degrees of freedom and 0.05 right-tail probability, I would reject the null hypothesis in favor of the alternative hypothesis. I would conclude that individual investors herded to different extents by market capitalization. Otherwise, individual investors did not herd to different extents. Table 2 shows the results of the chi-square test and descriptive statistics for herd behavior by market capitalization. With the result, χ^2 (4, 196,631) = 102,454.00, $p < .0001$, I concluded a statistically significant difference in individual investor herd behavior by market capitalization. I further performed pairwise comparisons among the quintiles of market capitalization as post hoc analysis. Given the quintiles of market capitalization, there were 10 pairs or subtables for subsequent chi-square tests. I used Bonferroni adjustments on the *p* values. For the 10 pairwise comparisons, the Bonferroni-adjusted *p* values for significance were 0.05/10 or 0.005. The *p* values of all pairs were smaller than 0.0001. The quintiles of market capitalization were statistically significantly different from each other. In the low market capitalization quintile, there was 5% of stock-days in which individual investors herded. The percentages progressively increased to 26% in the low-middle market capitalization, 62% in the middle, 90% in the middle-to-high, and 98% in the high. Individual investors herded most seriously in stocks with high market capitalization.

Table 2

Results of the Chi-Square Test for Herd Behavior by Market Capitalization

	# of stock-day	
Market capitalization	Not herded	Herded
Low	37,305 (0.95)	2,010 (0.05)[a]
Low-Middle	29,094 (0.74)	10,227 (0.26)[a]
Middle	15,145 (0.38)	24,225 (0.62)[a]
Middle-High	4,119 (0.10)	35,202 (0.90)[a]
High	731 (0.02)	38,573 (0.98)[a]

Note. χ^2 (4, N=196,631) = 102,454.00, p < .0001. Numbers in parentheses indicate row percentages.

[a]In the pairwise comparisons of the post hoc analysis, *p* value < .005, the Bonferroni-adjusted *p* values for significance in all pairs.

The second research question was whether a difference in individual investor herd behavior existed by P/B ratio of stock. I divided all stock-days by P/B ratio in quintiles. I performed a chi-square test for the independence of the percentage of stock-day in which individual investors herded and the P/B ratio in quintiles. The hypotheses were as follows:

$H_0$2: There is no statistically significant difference in individual investor herd behavior by P/B ratio.

H_a2: There is a statistically significant difference in individual investor herd behavior by P/B ratio.

If the chi-square statistic was bigger than 9.49, the critical value of the chi-square statistic with four degrees of freedom and 0.05 right-tail probability, I would reject the null hypothesis in favor of the alternative hypothesis. I would conclude that individual investors herded to different extents by P/B ratio. Otherwise, individual investors did not herd to different extents. Table 3 shows the results of the chi-square test and descriptive statistics for herd behavior by P/B ratio. With the result, χ^2 (4, 196,631) = 7,403.00, $p <$ .0001, I concluded a statistically significant difference in individual investor herd behavior by P/B ratio. I further performed pairwise comparisons among the quintiles of P/B ratio as post hoc analysis. Given the quintiles of P/B ratio, there were 10 pairs or subtables for subsequent chi-square tests. I used Bonferroni adjustment on the *p* values. For the 10 pairwise comparisons, the Bonferroni-adjusted *p* values for significance was 0.05/10 or 0.005. The *p* values of all pairs except the pair of low and low-middle were smaller than 0.0001. The quintiles of middle, middle-high, and high P/B ratio were statistically significantly different from each other. The quintiles of low and low-middle were not statistically significantly different between one another. There was 48% of stock-days in which individual investors herded in the low P/B ratio quintile and the low-middle P/B ratio quintile separately. The percentages progressively increased to 54% in the middle P/B ratio, 57% in the middle-high, and 74% in the high. Individual investors herded most seriously in stocks with high P/B ratio.

Table 3

Results of the Chi-Square Test for Herd Behavior by P/B Ratio

	# of stock-day	
P/B ratio	Not herded	Herded
Low	21,647 (0.52)	19,609 (0.48)
Low-Middle	19,819 (0.52)	18,507 (0.48)
Middle	18,017 (0.47)	20,730 (0.54)[a]
Middle-High	16,788 (0.43)	22,520 (0.57)[a]
High	10,123 (0.26)	28,871 (0.74)[a]

Note. χ^2 (4, *N*=196,631) = 7,403.00, $p < .0001$. Numbers in parentheses indicate row percentages.
[a]In the pairwise comparisons of the post hoc analysis, p value < .005, the Bonferroni-adjusted p values for significance, in all pairs.

The third research question was whether a difference in individual investor herd behavior existed by industry affiliation of stock. I performed a chi-square test for the independence of the percentage of stock-day in which individual investors herded and the industry affiliation. The hypotheses were as follows:

$H_0$3: There is no statistically significant difference in individual investor herd behavior by industry affiliation.

H_a3: There is a statistically significant difference in individual investor herd behavior by industry affiliation.

If the chi-square statistic was bigger than 40.11, the critical value of the chi-square statistic with 27 degrees of freedom and 0.05 right-tail probability, I would reject the null hypothesis in favor of the alternative hypothesis. I would conclude that individual investors herded to different extents by industry affiliation. Otherwise, individual investors did not herd to different extents. Table 4 shows the results of the chi-square test and descriptive statistics for herd behavior by industry affiliation. With the result, χ^2 (27, 196,631) = 12,776.84, $p < .0001$, I concluded a statistically significant difference in individual investor herd behavior by industry affiliation. I further performed pairwise comparisons among the industry affiliations as post hoc analysis. Given the 28 industry affiliations, there were 378 pairs or subtables for subsequent chi-square tests. I used Bonferroni adjustments on the p values. With the 378 pairwise comparisons, the Bonferroni-adjusted p values for significance was 0.05/28 or 0.0018. There were 347 pairs of industry affiliations independent from each other (p value < .0018) and 31 pairs related with each other (p value > .0018). The 31 pairs were across 21 industry affiliations. The remaining seven industry affiliations were statistically significantly different from each of other industry affiliations. Textile, electrical and cable, and information service were the three industry affiliations which individual investors herded less. Rubber, automobile, financial and insurance, and semiconductor were the four industry affiliations which individual investors herded more.

Table 4

Results of the Chi-square Test for Herd Behavior by Industry Affiliation

	# of stock-day	
Industry affiliation	Not herded	Herded
01 – Cement	621 (0.39)	986 (0.61)
02 – Food	2,232 (0.44)	2,844 (0.56)
03 – Plastic	2,110 (0.38)	3,434 (0.62)
04 – Textile	5,848 (0.58)	4,314 (0.42)[a]
05 – Electric machinery	5,527 (0.50)	5,599 (0.50)
06 – Electrical and cable	2,552 (0.74)	892 (0.26)[a]
08 – Glass and ceramic	763 (0.64)	429 (0.36)
09 – Paper and pulp	632 (0.37)	1,064 (0.63)
10 – Iron and steel	3,339 (0.54)	2,824 (0.46)
11 – Rubber	629 (0.24)	2,034 (0.76)[a]
12 – Automobile	103 (0.07)	1,361 (0.93)[a]
14 – Building material and construction	5,890 (0.51)	5,564 (0.49)
15 – Shipping and transportation	1,642 (0.34)	3,185 (0.66)
16 – Tourism	1,623 (0.52)	1,477 (0.48)
17 – Financial and insurance	1,101 (0.14)	6,949 (0.86)[a]
18 – Trading and consumers’ goods	1,311 (0.33)	2,656 (0.67)

(table continues)

	# of stock-day	
Industry affiliation	Not herded	Herded
20 – Others	4,126 (0.39)	6,437 (0.61)
21 – Chemical	3,160 (0.52)	2,872 (0.48)
22 – Biotechnology and medical care	3,050 (0.48)	3,286 (0.52)
23 – Oil, gas, and electricity	1,021 (0.64)	572 (0.36)
24 – Semiconductor	4,704 (0.30)	10,897 (0.70)[a]
25 – Computer and peripheral	4,278 (0.33)	8,567 (0.67)
26 – Optoelectronic	6,765 (0.43)	8,946 (0.57)
27 – Communications and Internet	3,726 (0.43)	5,035 (0.57)
28 – Electronic parts/components	10,486 (0.48)	11,221 (0.52)
29 – Electronic products distribution	2,950 (0.61)	1,910 (0.39)
30 – Information service	2,384 (0.85)	432 (0.15)[a]
31 – Other electronic	3,821 (0.46)	4,448 (0.54)

Note. χ^2 (27, N=196,631) = 12,776.84, $p < .0001$. Numbers in parentheses indicate row percentages.
[a]In the pairwise comparisons of the post hoc analysis, p value < .005, the Bonferroni-adjusted p values for significance, in all pairs.

Based on the results of the three hypothesis tests, I concluded that statistically significant differences in herd behavior existed by the three independent variables, market capitalization, P/B ratio, and industry affiliation. I proceeded to examine any statistically significant differences separately in buy- and sell-herding by the three independent variables. If statistically significant differences in buy- and sell-herding both existed by each independent variable, I would conclude no statistically significant difference in either buy- or sell-herding only by all independent variables.

Buy- and Sell-Herding. With the segregated stock-day data, I tested the hypotheses of the first research question twice more. One chi-square test was for the independence of the percentage of stock-day which individual investors herded to buy and the market capitalization in quintiles. Another chi-square test was the same but on individual investors herded to sell. The critical values of the chi-square statistic of both tests were the same, 9.49, with four degrees of freedom and 0.05 right-tail probability. Table 5 shows the results of the two chi-square tests and descriptive statistics for herd behavior by market capitalization. With the result, χ^2 (4, 106,883) = 58,650.43, $p < .0001$, I concluded a statistically significant difference in buy-herding by market capitalization. With the result, χ^2 (4, 89,748) = 43,035.95, $p < .0001$, I concluded a statistically significant difference in sell-herding by market capitalization too.

Table 5

Results of the Chi-Square Tests for Buy- and Sell-Herding by Market Capitalization

	# of stock-day		# of stock-day	
Market	Not herded	Herded to buy	Not herded	Herded to sell
Low	12,918 (0.92)	1,140 (0.08)	24,387 (0.97)	870 (0.03)
Low-Middle	12,896 (0.71)	5,370 (0.29)	16,198 (0.77)	4,857 (0.23)
Middle	6,561 (0.36)	11,848 (0.64)	8,584 (0.41)	12,377 (0.59)
Middle-High	1,870 (0.10)	17,710 (0.90)	2,249 (0.11)	17,492 (0.88)
High	313 (0.02)	19,122 (0.98)	418 (0.02)	19,451 (0.98)

Note. χ^2 (4, 106,883) = 58,650.43, $p < .0001$; χ^2 (4, 89,748) = 43,035.95, $p < .0001$
Numbers in parentheses indicate row percentages.

Similarly, I tested the hypotheses of the second research question twice more. One chi-square test was for the independence of the percentage of stock-day which individual investors herded to buy and the P/B ratio in quintiles. Another chi-square test was the same but on individual investors herded to sell. The critical values of the chi-square statistic of both tests were the same, 9.49, with four degrees of freedom and 0.05 right-tail probability. Table 6 shows the results of the two chi-square tests and descriptive statistics for herd behavior by P/B ratio. With the result, χ^2 (4, 106,883) = 4,733.57, p < .0001, I concluded a statistically significant difference in buy-herding by P/B ratio. With the result, χ^2 (4, 89,748) = 2,865.99, p < .0001, I concluded a statistically significant difference in sell-herding by P/B ratio too.

Table 6

Results of the Chi-square Tests for Buy- and Sell-Herding by P/B Ratio

	# of stock-day		# of stock-day	
Market	Not herded	Herded to buy	Not herded	Herded to sell
Low	13,153 (0.58)	9,577 (0.42)	8,494 (0.46)	10,032 (0.54)
Low-Middle	11,649 (0.56)	9,192 (0.44)	8,170 (0.47)	9,315 (0.53)
Middle	10,921 (0.53)	9,659 (0.47)	7,096 (0.39)	11,071 (0.61)
Middle-High	9,936 (0.46)	11,448 (0.54)	6,852 (0.38)	11,072 (0.62)
High	6,177 (0.29)	15,171 (0.71)	3,946 (0.22)	13,700 (0.78)

Note. χ^2 (4, 106,883) = 4,733.57, p < .0001; χ^2 (4, 89,748) = 2,865.99, p < .0001
Numbers in parentheses indicate row percentages.

Lastly, I tested the hypotheses of the third research question twice more. One chi-square test was for the independence of the percentage of stock-day which individual investors herded to buy and the industry affiliation. Another chi-square test was the same but on individual investors herded to sell. The critical values of the chi-square statistic of both tests were the same, 40.11, with 27 degrees of freedom and 0.05 right-tail probability. Table 7 shows the results of the two chi-square tests and descriptive statistics for herd behavior by industry affiliation. With the result, χ^2 (27, 106,883) = 8,041.72, $p < .0001$, I concluded a statistically significant difference in buy-herding by industry affiliation. With the result, χ^2 (27, 89,748) = 5,203.27, $p < .0001$, I concluded a statistically significant difference in sell-herding by industry affiliation too.

Table 7

Results of the Chi-square Tests for Buy- and Sell-Herding by Industry Affiliation

	# of stock-day		# of stock-day	
Industry affiliation	Not herded	Herded to buy	Not herded	Herded to sell
01 – Cement	351 (0.41)	507 (0.59)	270 (0.36)	481 (0.64)
02 – Food	1,388 (0.52)	1,304 (0.48)	844 (0.35)	1,540 (0.65)
03 – Plastic	1,242 (0.43)	1,622 (0.57)	868 (0.32)	1,812 (0.68)
04 – Textile	3,714 (0.62)	2,251 (0.38)	2,134 (0.51)	2,063 (0.49)
05 – Electric machinery	3,317 (0.53)	2,905 (0.47)	2,210 (0.45)	2,694 (0.55)
06 – Electrical and cable	1,552 (0.81)	363 (0.19)	1,000 (0.65)	529 (0.35)
08 – Glass and ceramic	377 (0.63)	217 (0.37)	386 (0.65)	212 (0.35)
09 – Paper and pulp	357 (0.39)	568 (0.61)	275 (0.36)	496 (0.64)
10 – Iron and steel	1,924 (0.67)	965 (0.33)	1,415 (0.43)	1,859 (0.57)
11 – Rubber	323 (0.24)	1,037 (0.76)	306 (0.23)	997 (0.77)
12 – Automobile	67 (0.09)	688 (0.91)	36 (0.05)	673 (0.95)
14 – Building material and construction	3,465 (0.58)	2,560 (0.42)	2,425 (0.45)	3,004 (0.55)
15 – Shipping and transportation	1,037 (0.34)	1,975 (0.66)	605 (0.33)	1,210 (0.67)
16 – Tourism	948 (0.54)	819 (0.46)	675 (0.51)	658 (0.49)
17 – Financial and insurance	663 (0.16)	3,457 (0.84)	438 (0.12)	3,492 (0.88)
18 – Trading and consumers' goods	771 (0.35)	1,426 (0.65)	540 (0.31)	1,230 (0.69)

(table continues)

	# of stock-day		# of stock-day	
Industry affiliation	Not herded	Herded to buy	Not herded	Herded to sell
20 – Others	2,596 (0.44)	3,259 (0.56)	1,530 (0.33)	3,179 (0.67)
21 – Chemical	2,160 (0.61)	1,377 (0.39)	1,000 (0.40)	1,495 (0.60)
22 – Biotechnology and medical care	1680 (0.47)	1,886 (0.53)	1,370 (0.49)	1,400 (0.51)
23 – Oil, gas, and electricity	594 (0.69)	270 (0.31)	427 (0.59)	302 (0.41)
24 – Semiconductor	2,590 (0.33)	5,377 (0.67)	2,114 (0.28)	5,520 (0.72)
25 – Computer and peripheral	2,729 (0.40)	4,027 (0.60)	1,549 (0.25)	4,540 (0.75)
26 – Optoelectronic	4,116 (0.47)	4,731 (0.53)	2,649 (0.39)	4,215 (0.61)
27 – Communications and Internet	2,215 (0.46)	2,560 (0.54)	1,511 (0.38)	2,475 (0.62)
28 – Electronic parts/ components	5,979 (0.52)	5,594 (0.48)	4,507 (0.44)	5,627 (0.56)
29 – Electronic products distribution	1,794 (0.68)	847 (0.32)	1,156 (0.52)	1,063 (0.48)
30 – Information service	1,528 (0.90)	175 (0.10)	856 (0.77)	257 (0.23)
31 – Other electronic	2,359 (0.51)	2,280 (0.49)	1,462 (0.40)	2,168 (0.60)

Note. χ^2 (27, 106,883) = 8,041.72, $p < .0001$; χ^2 (27, 89,748) = 5,203.27, $p < .0001$
Numbers in parentheses indicate row percentages.

The purpose of the six additional hypothesis tests was to identify any differences in either buy- or sell-herding but not both. Based on the results, I concluded that statistically significant differences in both buy- and sell-herding existed by the three independent variables, market capitalization, P/B ratio, and industry affiliation. No difference existed between one independent variable and either buy- or sell-herding but not both. I did not pursue further data analysis including post hoc analyses for buy- and sell-herding because any finding of general herding would be more understandable to individual investors. It became necessary to examine the relation between general herding and the three independent variables as a whole through a multivariate analysis. Any finding would be more pragmatic for creating a social change.

Multivariate Analysis

To introduce a multivariate analysis, I first studied the characteristics of independent and dependent variables. Then, I explained the suitability of logistic regression method. I regressed herd behavior against market capitalization, P/B ratio, and industry affiliation. I assessed the statistical significance of the logistic regression model. Lastly, I estimated expected herd behavior at stock level by trading day.

Characteristics of Independent and Dependent Variables

In the hypothesis tests, I had estimated the LSV measure and tested its statistical significance for each stock-day. I had also created a dichotomous variable to reflect the herding occurrence or not at stock-day level. I defined the dichotomous herd behavior variable to be the dependent variable in the multivariate analysis. Two independent variables, market capitalization and P/B ratio, were of ordinal scale in the hypothesis tests. Their scale was 5-point, from low, low-middle, middle, middle-high to high. There was a higher percentage of individual investors herded at the high side of the scale of each variable. In fact, the two independent variables were of a ratio scale originally. Intervals of the ratio scale are equal, hence advantageous over other scales including ordinal. I preserved the ratio scales of market capitalization and P/B ratio in the multivariate analysis. The third independent variable, industry affiliation, was of nominal scale. I included industry affiliation as was in the multivariate analysis. Given the characteristics of the independent and dependent variables, I adopted logistic regression for the multivariate analysis.

Suitability of Logistic Regression

Logistic regression is a multivariate analysis which I could use to explain the relationship between one dichotomous dependent variable and a combination of one nominal independent and two ratio independent variables. The nonnormality of the LSV estimates would no longer be a concern because of its transformation into a dichotomous variable. The dichotomous herd behavior variable indicated an occurrence or not at stock-day level. Individual investors could have herded for one stock in certain trading days but not all. In logistic regression, I would turn the herding occurrence into an extent or a probability. The results of a logistic regression would be probabilistic. I would relate the independent variables to the probabilistic results. A change in the extent or probability of herding might be related to the change of any ratio-scaled independent variables, market capitalization and P/B ratio. For the nominal-scaled independent variable, industry affiliation, the hypothesis test results indicated differences in herd behavior among industry affiliations. Logistic regression would account for the difference of each industry affiliation. The change in herding extent was on a relative basis between one industry affiliation and another. I defined the *others* industry affiliation whose code was 20 as the reference in logistic regression. Through logistic regression, I could understand the impact of each independent variable on the dependent variable with control for other independent variables. For example, there were high percentages of stock-days with individual herding, 98% and 86% respectively in high market capitalization and financial and insurance industry affiliation. However, financial and insurance industry was by nature high market capitalization. I would estimate the impact of market capitalization on the extent of herd behavior with control for industry affiliation and other independent variables.

Logistic Regression

The linearity between the log-odds of a dependent variable and an independent variable is important for logistic regression. I assessed the linearity between the log-odds of herd behavior and market capitalization in two forms, namely original and logarithmic. Then, I estimated log-odds of herd behavior at a stock-month level from 196,631 stock-days. There were 6,921 stock-months with either numerator or denominator zero. I could not take the logarithm of the 6,921 stock-months but the 2,939 stock-months with nonzero numerator and nonzero denominator. The log-odds of herd behavior was more correlated with the logarithmic market capitalization, $r(2{,}937) = 0.50$, $p < .0001$ than with the original market capitalization, $r(2{,}937) = 0.37$, $p < .0001$. Similarly, I compared the linearity between the log-odds of herd behavior and P/B ratio in two forms. The log-odds of herd behavior was more correlated with the original P/B ratio, $r(2{,}937) = 0.12$, $p < .0001$ than with the logarithmic P/B ratio, $r(2{,}937) = 0.06$, $p < .0001$. As a result, I transformed only one independent variable, market capitalization, into logarithm for logistic regression.

Based on Equation 1, I regressed by stepwise the herd behavior against the logarithmic market capitalization, P/B ratio, and industry affiliation with the *others* as the reference. I expressed the regression result in Equation 7.

$$\begin{aligned}\mathrm{logit}(p) = &-23.60 + 2.76X_1 - 0.30X_2 - 0.74X_{3a} + 0.24X_{3b} + 1.21X_{3c} + 0.37X_{3d} \\ &+ 0.14X_{3e} + 0.61X_{3f} - 0.17X_{3g} + 0.77X_{3h} + 0.59X_{3i} - 0.57X_{3j} + 0.09X_{3k} \\ &- 0.13X_{3l} - 0.66X_{3m} + 1.01X_{3n} + 0.30X_{3o} + 0.68X_{3p} - 2.21X_{3q} + 1.52X_{3r} \\ &+ 1.83X_{3s} + 1.62X_{3t} + 1.31X_{3u} + 1.02X_{3v} - 0.17X_{3w} - 0.79X_{3x} + 0.83X_{3y}\end{aligned} \tag{7}$$

where

X_1 is a stock's logarithmic market capitalization (in millions of NT dollar),

X_2 is a stock's P/B ratio,

X_{3a} is one for cement as the industry affiliation, otherwise zero,

X_{3b} is one for food as the industry affiliation, otherwise zero,

X_{3c} is one for textile as the industry affiliation, otherwise zero,

X_{3d} is one for electric machinery as the industry affiliation, otherwise zero,

X_{3e} is one for electrical and cable as the industry affiliation, otherwise zero,

X_{3f} is one for glass and ceramic as the industry affiliation, otherwise zero,

X_{3g} is one for paper and pulp as the industry affiliation, otherwise zero,

X_{3h} is one for iron and steel as the industry affiliation, otherwise zero,

X_{3i} is one for rubber as the industry affiliation, otherwise zero,

X_{3j} is one for automobile as the industry affiliation, otherwise zero,

X_{3k} is one for building material and construction as the industry affiliation, otherwise zero,

X_{3l} is one for shipping and transportation as the industry affiliation, otherwise zero,

X_{3m} is one for financial and insurance as the industry affiliation, otherwise zero,

X_{3n} is one for trading and consumers' goods as the industry affiliation, otherwise zero,

X_{3o} is one for chemical as the industry affiliation, otherwise zero,

X_{3p} is one for biotechnology and medical care as the industry affiliation, otherwise zero,

X_{3q} is one for oil, gas, and electricity as the industry affiliation, otherwise zero,

X_{3r} is one for semiconductor as the industry affiliation, otherwise zero,

X_{3s} is one for computer and peripheral as the industry affiliation, otherwise zero,

X_{3t} is one for optoelectronic as the industry affiliation, otherwise zero,

X_{3u} is one for communications and Internet as the industry affiliation, otherwise zero,

X_{3v} is one for electronic parts/components as the industry affiliation, otherwise zero,

X_{3w} is one for electronic products distribution as the industry affiliation, otherwise zero,

X_{3x} is one for information service as the industry affiliation, otherwise zero,

X_{3y} is one for other electronic as the industry affiliation, otherwise zero, and logit(p) is the odds ratio of the positive herd behavior in logarithm form, ln[p/(1-p)], and p is the actual proportion of positive herd behavior. I performed likelihood ratio chi-square test on the logistic regression model. The hypotheses were as follows:

H_0: The coefficients of all three independent variables are equal to zero.

H_a: The coefficient of at least one independent variable is not equal to zero.

If the likelihood ratio chi-square statistic was bigger than 42.56, the critical value of the chi-square statistic with 29 degrees of freedom and 0.05 right-tail probability, I would reject the null hypothesis in favor of the alternative hypothesis. I would conclude the coefficient of at least one independent variable unequal to zero. Otherwise, the coefficients of all three independent variables were equal to zero. With the result, χ^2 (29, N=196,631) = 139,702.44, $p < .0001$, I concluded the coefficient of at least one independent variable unequal to zero. Then, I assessed the statistical significance of the coefficient of each independent variable with its Wald chi-square statistic. All three independent variables, log market capitalization (χ^2 (1, N=196,631) = 41,914.75, $p < .0001$), P/B ratio (χ^2 (1, N=196,631) = 1,466.29, $p < .0001$), and industry affiliation (χ^2 (1, N=196,631) = 8,955.12, $p < .0001$), were statistically significant and in the final logistic regression model. For a unit change in log market capitalization, the odds ratio of herding was 15.73 with other independent variables controlled. The bigger the log market capitalization, the higher was the individual investor herding. For a unit change in P/B ratio, the odds ratio of herding was 0.74 with other independent variables controlled. The bigger the P/B ratio, the lower was the individual investor herding. For the eight industry affiliations including (a) cement; (b) paper and pulp; (c) automobile; (d) shipping and transportation; (e) financial and insurance; (f) oil, gas, and electricity; (g) electronic products distribution; and (h) information service, the odds ratio of herding was smaller than one with other independent variables controlled. Individual investor herding was relatively lower. For the 17 industry affiliations including (a) food; (b) textile; (c) electric machining; (d) electrical and cable; (e) glass and ceramic; (f) iron and steel; (g) rubber; (h) building material and construction; (i) trading and consumers' goods; (j) chemical; (k) biotechnology and medical care; (l) semiconductor; (m) computer and peripheral; (n) optoelectronic; (o) communications and Internet; (p) electronic parts/components; and (q) other electronic, the odds ratio of herding was larger than one with other independent variables controlled. Individual investor herding was relatively higher. Two industry affiliations, plastic and tourism, were not statistically significant. Table 8 shows the logistic regression model results.

Table 8

Logistic Regression Model of Herd Behavior against Logarithmic Market Capitalization, P/B Ratio, and Industry Affiliation

	Est. β	S.E. β	Wald χ^2	*p*	Odds ratio	95% C.I.
Intercept	-23.60	0.12	39,009.95	<.0001		
Log market capitalization	2.76	0.01	41,914.75	<.0001	15.73	15.32, 16.15
P/B ratio	-0.30	0.01	1,466.29	<.0001	0.74	0.73, 0.75
Industry affiliation			8,955.12	<.0001		
01 – Cement	-0.74	0.08	86.73	<.0001	0.48	0.41, 0.56
02 – Food	0.24	0.05	22.82	<.0001	1.28	1.15, 1.41
04 – Textile	1.21	0.05	687.37	<.0001	3.36	3.07, 3.68
05 – Electric machinery	0.37	0.04	76.03	<.0001	1.45	1.33, 1.58
06 – Electrical and cable	0.14	0.06	5.22	0.02	1.15	1.02, 1.30
08 – Glass and ceramic	0.61	0.09	46.90	<.0001	1.84	1.55, 2.20
09 – Paper and pulp	-0.17	0.07	5.26	0.02	0.84	0.73, 0.98
10 – Iron and steel	0.77	0.05	234.46	<.0001	2.17	1.97, 2.40

(table continues)

11 – Rubber	0.59	0.07	77.71	<.0001	1.81	1.59, 2.07
12 – Automobile	-0.57	0.12	22.18	<.0001	0.57	0.45, 0.72
14 – Building material and construction	0.09	0.04	5.18	0.02	1.10	1.01, 1.19
15 – Shipping and transportation	-0.13	0.06	4.80	0.03	0.88	0.78, 0.99
17 – Financial and insurance	-0.66	0.06	134.51	<.0001	0.52	0.46, 0.58
18 – Trading and consumers' goods	1.01	0.06	303.26	<.0001	2.75	2.45, 3.08
21 – Chemical	0.30	0.05	34.41	<.0001	1.35	1.22, 1.50
22 – Biotechnology and medical care	0.68	0.05	224.02	<.0001	1.97	1.81, 2.16
23 – Oil, gas, and electricity	-2.21	0.08	789.26	<.0001	0.11	0.09, 0.13
24 – Semiconductor	1.52	0.04	1,366.25	<.0001	4.55	4.20, 4.93
25 – Computer and peripheral	1.83	0.05	1,606.82	<.0001	6.24	5.70, 6.82
26 – Optoelectronic	1.62	0.04	1577.86	<.0001	5.07	4.67, 5.49
27 – Communications and Internet	1.31	0.05	765.11	<.0001	3.72	3.39, 4.09
28 – Electronic parts/ components	1.02	0.04	717.43	<.0001	2.76	2.56, 2.97
29 – Electronic products distribution	-0.17	0.05	10.39	0.00	0.85	0.77, 0.94
30 – Information service	-0.79	0.08	106.98	<.0001	0.45	0.39, 0.53
31 – Other electronic	0.83	0.04	339.75	<.0001	2.30	2.10, 2.50

With the statistical significance of each independent variable, I assessed the goodness-of-fit of the three independent variables as a whole. The Nagelkerke R^2 was 0.6815. I attributed 68.15% of total variability in the herd behavior to the collective variability in logarithmic market capitalization, P/B ratio, and industry affiliation. Last, I assessed the agreement between pairs of variables. There were 110,237 stock-days with individual herding and 86,394 stock-days without. The product between them was the number of pairs, 9,523,815,378. Among them, 93.3% was concordant and 6.7% was discordant. The difference between them was the Somers' D, 86.5%. The difference was close to the end of 100.0% that represented perfect agreements of all pairs of the variables than another end of -100.0% that represented perfect disagreements. With 68.15% herd behavior explainable and 86.5% of agreement between pairs of variables, I concluded the goodness-of-fit of the model. I would focus on the expected herd behavior and its implication.

Expected Herd Behavior

Expected herd behavior was a probability of individual investor herd behavior. The coefficients and the specific values of the logarithmic market capitalization, P/B ratio, and industry affiliation of a stock-day were inputs to expected herd behavior estimation. Based on Equation 7, I estimated expected herd behavior for each of the 196,631 stock-days. Table 9 shows the distribution of the stock-days by expected herd behavior. 31.63% or 62,201 stock-days were in the highest expected herd behavior interval (>0.9 – 1.0). Out of them, 97% or 60,241 stock-days were with actual herding.

49.06% or 96,454 stock-days were in next eight intervals. 19.31% or 37,976 stock-days were in the lowest expected herd behavior interval (0.0 – 0.1). Out of them, 96% or 36,610 stock-days were without actual herding.

Table 9

Distribution of Stock-Days by Expected Herd Behavior

Expected herd behavior	#	%	Cumulative #	Cumulative %	# with actual herding	% with actual herding
>0.9 – 1.0	62,201	31.63	62,201	31.63	60,241	97%
>0.8 – 0.9	16,941	8.62	79,142	40.25	15,145	89%
>0.7 – 0.8	11,896	6.05	91,038	46.30	9,505	80%
>0.6 – 0.7	10,226	5.20	101,264	51.50	6,745	66%
>0.5 – 0.6	8,476	4.31	109,740	55.81	4,178	49%
>0.4 – 0.5	10,050	5.11	119,790	60.92	4,189	42%
>0.3 – 0.4	10,261	5.22	130,051	66.14	3,331	32%
>0.2 – 0.3	11,816	6.01	141,867	72.15	3,068	26%
>0.1 – 0.2	16,788	8.54	158,655	80.69	2,469	15%
0.0 – 0.1	37,976	19.31	196,631	100.00	1,366	4%

Within the highest expected herd behavior interval, I identified 9 stock-days with the highest expected herd behavior, 1.00. Among the 9 stock-days, I randomly selected one for studying the three independent variables at a fundamental level. Taiwan Semiconductor Manufacturing Company Limited (TSM) was of stock code 2330. TSM's logarithm market capitalization was 15.37. It was equivalent to NT$4,750,213 million or US$141,434 million at 2016 exchange rate of NT$33.586 to US$1 (see Internal Revenue Service, 2018) market capitalization which was close to the upper bound of high market capitalization quintile. TSM's P/B ratio was 3.70 in the high P/B quintile. TSM's industry affiliation was semiconductor. In comparison with the industry affiliation of others, the odds ratio of herding was 4.55. Similarly, I identified the stock-day with the lowest expected herd behavior, 0.00. Sumagh High Technology Corporate (SHT) was of stock code 1475. SHT's logarithm market capitalization was 4.55. It was equivalent to NT$94.76 million or US$2.82 million at 2016 exchange rate of NT$33.586 to US$1 (see Internal Revenue Service, 2018) market capitalization which was in the low market capitalization quintile. SHT's P/B ratio was 0.94 in the middle P/B quintile. SHT's industry affiliation was textile. In comparison with the industry affiliation of others, the odds ratio of herding was 3.36.

Summary

With 196,631 stock-days in 2016, I estimated the LSV measure of individual investor at 0.04. The percentage of stock-day in which individual investors herded was 56.11%. There were statistically significant differences in individual herding by market capitalization, P/B ratio, and industry affiliation separately. I also estimated the LSV measures of buy- and sell-herding at 0.03 and 0.05 respectively. The percentages of stock-day in which individual investors herded to buy and to sell were 51.50% and 61.49% respectively. Similarly, there were statistically significant differences in individual herding by market capitalization, P/B ratio, and industry affiliation separately. I adopted logistic regression to study the combined effects of the three independent variables upon herding. For a unit change in log market capitalization, the odds ratio of herding was 15.73. For a unit change in P/B ratio, the odds ratio of herding was 0.74. For the eight industry affiliations including (a) cement; (b) paper and pulp; (c) automobile; (d) shipping and transportation; (e) financial and insurance; (f) oil, gas, and electricity; (g) electronic products distribution; and (h) information service, the odds ratio was smaller than one. For the 17 industry affiliations including (a) food; (b) textile; (c) electric machining; (d) electrical and cable; (e) glass and ceramic; (f) iron and steel; (g) rubber; (h) building material and construction; (i) trading and consumers' goods; (j) chemical; (k) biotechnology and medical care; (l) semiconductor; (m) computer and peripheral; (n) optoelectronic; (o) communications and Internet; (p) electronic parts/components; and (q) other electronic, the odds ratio of herding was larger than one. 31.63% stock-days were of higher than 90% probability to herd whereas 19.31% stock-days were of lower than 10% probability to herd. In conclusion, Taiwan individual investors herded in more than half of the stock-days in 2016. In almost one third of the stock-days, they herded at a great extent, 90%. On the other hand, they herded negligibly in almost one fifth of the stock-days. The results were evidence of individual investor herd behavior in Taiwan. There was variations in the results, therefore I will discuss the herd behavior in the context of the reality in next chapter.

Chapter 5: Discussion, Recommendations, and Conclusion

Introduction

The purpose of this quantitative research in causal-comparative design was to examine individual investor herd behavior as related to characteristics of stocks in Taiwan stock market. The characteristics of stocks included (a) market capitalization, (b) P/B ratio, and (c) industry affiliation. The findings of this study contribute to the overall knowledge of herding in the field of behavioral finance.

With 196,631 stock-days in 2016, I estimated the LSV measure of individual investor at 0.04. There were statistically significant differences in individual herding by market capitalization, P/B ratio, and industry affiliation separately. I also estimated the LSV measures of buy- and sell-herding at 0.03 and 0.05 respectively. Similarly, there were statistically significant differences in individual herding by market capitalization, P/B ratio, and industry affiliation separately. I adopted logistic regression and found the combined effects of the three characteristics of stocks with market capitalization logarithmically transformed on herd behavior. In almost one third of the stock-days, investors herded at a great extent of 90%.

In this chapter, I will interpret the findings of this study in the flow of comparability with extant literature, comparisons with extant literature, comparisons with the logistic regression results, and contribution to the knowledge. I will evaluate the limitations of this study from five perspectives: validity, reliability, generalizability, use of secondary data, and TWSE data. On the grounds of the limitations and strengths of this study, I will also make recommendations for further research. Lastly, I will discuss the implications of the findings and proposed action items through which the results of this study can be used towards achieving positive social change.

Discussion of the Findings

Before presenting my interpretations of the hypothesis test results, I will discuss the comparability between this study and the peer-reviewed literature described in Chapter 2. I will focus on the same herd measure and the same market. Having reconciled estimates of herd measure, I will compare the findings of this study with the relevant ones in previous studies. Then, I will describe in what ways the findings confirmed, disconfirmed, or extended knowledge of herding in the field of behavioral finance.

Comparability with Extant Literature

For the sampling period, the worksheet named Form3 of the Statistics of Securities Market file (TWSE, 2018b) reported 52% of the total trading value constituted individual investors. I reconciled the same level of constitution from the monthly trade transaction data of this study. The same worksheet of 2010 reported 68% (TWSE, 2018b) as the full-year total trading value by individual investors, a level comparable to the 69% cited by Lin et al. (2013). Individual investor constitution reduced 16% to 17% from 2010 to 2016. Despite the reduction, 52% was still the largest among the four types of investors.

Next, I focused on the LSV measure estimate distribution. Table 10 shows the descriptive statistics of this study and Wermers' (1999) study. Wermers' research samples were institutional investors in the United States. Although the two research samples were completely different, the two distributions are comparable. The two means, medians, and standard deviations are close. Both minima are negative and nearer to zeros than both maxima. The distribution of the LSV measure estimates in this study was positively skewed (moment coefficient of skewness = 1.36), whereas that of Wermers' study was likely positively skewed too with its median smaller than its mean. I reconciled the distribution of the LSV measure estimates in this study, and therefore, proceeded further.

Table 10

Distributions of the LSV Measure Estimates

	This study	Wermers' (1999)
Number of observations	196,631	109,486
Mean	0.039	0.034
Median	0.01	0.01
Standard deviation	0.08	0.12
Maximum	0.46	0.65
Minimum	-0.12	-0.16
Range	0.58	0.65

The mean LSV measure of the 196,631 stock-days in 2016 was 0.039, rounded up to 0.04 in Chapter 4. I used data from TWSE, the same source as Chang et al. (2012) did. Chang et al. reported a mean LSV measure of 0.057 by individual investors in Taiwan from July 2006 to June 2008. There is a difference of 0.018 between our findings. The conditional LSV measures of buy- and sell-herding in 2016 were 0.031, rounded down to 0.03, and 0.049, rounded up to 0.05, respectively in Chapter 4. Chang et al. also reported a mean conditional LSV measure of buy-herding at 0.056 and that of sell-herding at 0.057. There are also differences in the corresponding conditional LSV measures between this study and Chang et al.'s. Chang et al.'s set of standard deviations from 0.007 to 0.009 was about one tenth of the one in this study, 0.08. I performed *t* tests between the two sets of means to examine the significance of the differences. Chang et al. did not mention the number of stock-day data but 484 trading days total for 2 years. In 244 trading days of 2016 (1 year), there were 196,631 stock-days which I doubled and used as a proxy of the number of stock-days in Chang et al.'s study for one-sided *t* tests with unequal variances. There was not a significant difference in herding, $t(0) = 90$, $p = .5$; buy-herding, $t(0) = 90$, $p = .5$; and sell-herding, $t(0) = 90$, $p = .5$. These findings confirmed comparable levels of individual herding in Taiwan between the two periods of time. Additionally, the higher conditional LSV measure of sell-herding than that of buy-herding in this study confirmed Chang et al.'s statement of tending toward sell-herding rather than buy-herding.

With comparable levels and trends of herding, it was appropriate to dig into the hypothesis test results. I compared the hypothesis test results with the findings from the extant literature and also interpreted the comparison results.

Comparisons with Extant Literature

In this study, I found that the higher the market capitalization of a stock, the more serious the individual herding was. Yao et al. (2014) did not differentiate between institutional and individual investors but claimed a dominance by individual investors in China markets. Yao et al. had a similar finding of more serious herding in large stocks in

China. Analysts follow large stocks, especially blue chips, more closely, and individual investors may react to analyst recommendations whose coverages are more frequent. Zheng et al. (2015) also found more serious herding in large stocks by institutional investors in China. The findings of this study confirmed the knowledge of more serious herding in high market capitalization or large stocks. Yao et al. and Zheng et al. found more serious herding in small stocks but not in medium stocks. In this study, I found the opposite, and individual herding was less serious in small stocks and close to the mean in medium stocks. This finding disconfirmed the knowledge of more serious herding in low market capitalization or small stocks and no herding in middle market capitalization or medium stocks. Furthermore, I found that the relationship between the market capitalization of a stock and the log odds of individual herding were less linear than that between the logarithmic market capitalization and the log odds of individual herding. Both Yao et al. and Zheng et al. studied herding by market capitalization without transformation. The findings of this study extended the knowledge around more linearity between the logarithmic market capitalization and the log odds of individual herding.

In this study, I also found that the higher the P/B ratio of a stock, the more serious the individual herding was. Yao et al. (2014) used the reciprocal of P/B ratio, the BTM ratio, and had a similar finding with more serious herding in lower BTM or higher P/B ratio stocks in China. The price of a high P/B ratio stock is by definition high. The high price is because of an expected high growth rate of the stock; hence, a high P/B ratio stock is also known as a growth stock. Investors cannot simply extrapolate to estimate the intrinsic value of a growth stock from its historical performance; they have to come up with forward-looking assumptions and a complex model (Garcia & Oliveira, 2018). Some investors may find such estimation difficult and intentionally herd instead. Also, Gong and Dai (2017) found more serious herding in growth stocks than in value stocks. Prices of value stocks are usually on par with or below the market average. I found the least herding in value stocks. This finding confirmed the knowledge of more serious herding in high P/B ratio or growth stocks and less serious herding in low P/B ratio or value stocks.

In this study, I also found individual herding in all 28 industries in Taiwan. The broad industry herding happened in Europe as well with all 10 industries except consumer goods (Ouarda et al., 2013) and in Central and East Europe with all five industries except construction (Filip & Pochea, 2014). In China, industry herding was less broad with 15 industries out of a total of 21 found to be herding (Yao et al., 2014). In contrast, industry herding in the United States was minor, in only two industries, public utilities and transportation, out of a total of 12 (Litimi et al., 2016). The industry classifications across stock exchanges are inconsistent. There were 22 industries in this study compared to 10 in Yao et al.'s. Herding occurred commonly in Taiwan and China in the following 22 industries: (a) food; (b) textiles; (c) paper and pulp; (d) plastic, rubber, and chemicals as a group comparable to petrochemicals; (e) semiconductors, computers and peripherals, optoelectronics, communications and Internet, electronic parts/components, electronic products distribution, and other electronics as a group comparable to electronics; (f) iron and steel; (g) electric machinery, electrical and cable, glass and ceramic, and automobiles as a group comparable to other manufacturing; (h) cement, building material, and construction as a group comparable to construction; (i) shipping and transportation; and (j) trading and consumers' goods.

There were five other industries in this study compared to the corresponding ones in Europe. Herding occurred commonly in (a) oil, gas, and electricity and (b) financial and insurance in Taiwan, Europe (Ouarda, et al., 2013), and Central and East Europe (Filip & Pochea, 2014); in (c) biotechnology and medical care and (d) information service in Taiwan and Europe (Ouarda, et al., 2013); and in (e) tourism in Taiwan and Central and East Europe (Filip & Pochea, 2014).

It was not meaningful to compare the others industry with other markets; therefore, I left the others industry out of this study. Each industry is uniquely linked to the political environment and macroeconomy. Major political and economic events of Taiwan in 2016 included the inauguration of new president, Tsai Ing-wen; China ratcheting up the isolation of Taiwan; the diversion of tour groups away from Taiwan (Tsai's brighter side, 2017), a surge of foreign direct investment in electronics industry (Tsai's brighter side, 2017), the magnitude 6.4 earthquake in southern Taiwan, the cease of TransAsia Airways, and so on. It took the efforts of individual investors to anticipate any implications of each event to an industry. It probably took even more individual investor effort to estimate any impacts on intrinsic values of related stocks. Some individual investors behaved by industry herding, which might have been an easier decision process for them. The findings of this study confirmed the knowledge of broad industry herding like had occurred in Europe. On the contrary, the same findings disconfirmed the knowledge of minor industry herding like that in the United States.

Comparison with the Logistic Regression Result

In this study, I employed a logistic regression model. In my review of the literature, I did not find any previous studies with regression models of herd behavior against a series of stock characteristics as independent variables. Hence, I had no previous regression models to compare my results with. I contrasted the logistic regression model results with the findings for the three research questions. From the logistic regression model, the odds ratio of herding was larger than one for a unit change in log market capitalization. This implies more serious herding in larger log market capitalization. Due to the strictly increasing property of logarithm, it also implies more serious herding in larger market capitalization. Not only was this finding consistent with that of the first research question but even clearer with other independent variables controlled.

From the logistic regression model, the odds ratio of herding was smaller than one for a unit change in P/B ratio. This implies less serious herding in higher P/B ratio. This finding was inconsistent with and opposite to that of the second research question. Since other independent variables were controlled in the logistic regression model, the opposite result implies a positive correlation between the P/B ratio and other independent variables which were more related to herding. Market capitalization was one of those independent variables ($r(196{,}631) = .13$). Its logarithmic form was even more correlated with the P/B ratio ($r(196{,}631) = .35$). This finding disconfirmed the knowledge of more serious herding in high P/B ratio stocks and extended the knowledge of less serious herding in high P/B ratio stocks with other independent variables controlled.

From the logistic regression model, the odds ratios of herding of eight industry affiliations were smaller than that of the others with other independent variables controlled. This implies the least serious herding with the industry affiliation of oil, gas, and electricity, followed by cement; financial and insurance; automobile; paper and pulp; electronic products distribution; and shipping and transportation at last. This finding of less serious herding with the industry affiliation of financial and insurance seemed contradictory to the finding in post hoc analysis for the research question. In fact, the herding with the industry affiliation of financial and insurance was mainly due to the correlated high market capitalization. Among the industry affiliations with high market capitalization, individual herding of financial and insurance was low. Plastic and tourism were the two industry affiliations whose odds ratios of herding were at par with the others. There were 17 industry affiliations with odds ratios of herding greater than that of the others with other independent variables controlled. Individual investors herded most seriously with computer and peripheral, followed by optoelectronic; semiconductor; communications and Internet; textile; electronic parts/components; trading and consumers' goods; other electronic; iron and steel; biotechnology and medical care; glass and ceramic; rubber; electric machinery; chemical; food; electrical and cable; and building material and construction at last. The top four industry affiliations were electronics related. Their odds ratios of herding were from 3.72 to 6.24, substantially higher than from 1.10 to 3.36 of the following 13 industry affiliations. The findings from the logistic regression model extended the knowledge of broad industry herding to the differences by industry affiliations with other independent variables controlled.

Contribution to the Knowledge

In response to the overarching research question, the findings of this study were that statistically significant differences in individual investor herd behavior existed by each of the three characteristics of stocks: market capitalization, P/B ratio, and industry affiliation. The major findings contributed to the body of behavioral finance knowledge included:

a. the extension of the knowledge to comparable levels of individual herding in Taiwan between Chang et al.’s (2012) sampling period, July 2006 to June 2008 and the sampling period of this study, 2016;

b. the confirmation of the knowledge of higher sell-herding than buy-herding;

c. the confirmation of the knowledge of more serious herding in high market capitalization stocks, and the disconfirmation of the knowledge of more serious herding in low market capitalization stocks and no herding in middle market capitalization stocks;

d. the extension of the knowledge to more linearity between the logarithmic market capitalization and the log odds of individual herding;

e. the confirmation of the knowledge of more serious herding in high P/B ratio stocks and less serious herding in low P/B ratio stocks without other independent variables controlled, but the disconfirmation of the same knowledge and the extension of the knowledge to less serious herding in high P/B ratio stocks with other independent variables controlled; and

f. the confirmation of the knowledge of broad industry herding like in Europe, but the disconfirmation the knowledge of minor industry herding like in the United States, and the extension of the knowledge to the differences by industry affiliations with other independent variables controlled.

I may have set an example of applying logistic regression upon a series of stock characteristics as independent variables and a dichotomous herd behavior variable based on the LSV measure. An advantage of such application is the revelation of the net difference in herd behavior by one stock characteristic with other stock characteristics controlled. Although the natures of the four most herded industry affiliations, namely computer and peripheral; optoelectronic; semiconductor; and communications and Internet, are interrelated, I could use logistic regression to quantify any differences among them with market capitalization and P/B ratio controlled.

The findings of this study also supported the theoretical base of herding. In 2016, Hon Hai Precision's $3.8 billion acquisition of Japanese electronics manufacturer Sharp (Inagaki, 2016), the merger between Advanced Semiconductor Engineering and Siliconware Precision Industries (Wang, 2016), the rising competition from China electronics companies in global markets (Flannery, 2016), and the 6.4 magnitude earthquake at electronics maker hub, Tainan (Lin, 2016) are examples of social events related to electronics. When their news was released, investors tried to predict potential impacts. When there was a mix of positive and negative news, not all individual investors could develop a clear view and predict. Some individual investors followed other investors probably including institutional ones to trade. The individual herding in 57% to

70% stock-day of the four industry affiliations, computer and peripheral; optoelectronic; semiconductor; and communications and Internet, is evidence. There were disproportionally more individual investors either buying or selling one stock in a trading day than the mean of 2016 whole year. The social events are exogenous factors of the individual herding whereas the three characteristics of stocks are endogenous factors. The statistically significant differences in herd behavior by each characteristic of stocks are evidence. The empirical evidence in this study collectively supported the theoretical base of herding. Inevitably, there are limitations to the findings and conclusions of the study. It is important to understand the limitations.

Limitations of the Study

I evaluated limitations of this study from five perspectives, namely validity, reliability, generalizability, use of secondary data, and TWSE data. As the design of this study was causal-comparative, typical strengths and limitations of such design were bound to be applicable. A typical limitation is no manipulation of an independent variable by researchers, therefore the researchers must infer the direction of causation logically or theoretically (Frankfort-Nachmias et al., 2015).

Validity

There are two parts of validity, internal and external. For internal validity, there are five potential limitations, namely selection effect, regression artifact, history, maturation, and experimental mortality. Selection effect and regression artifact were two least relevant limitations to this study because of the use of secondary data. I did not recruit individual investors as samples. I did not need to set up and assign individual investors to experiment and control groups either. There was no selection per se, therefore selection effect was not relevant to this study. I used purposive sampling to exclude as few samples as possible from the population. The exclusions were illiquid stocks and exchange sanctioned stocks only. As a result, the samples consisted of the majority, 94%, of the population and were unlikely nonrandom from the population. Hence, regression artifact was scarcely relevant to this study.

History, maturation, and experimental mortality are three potential limitations. In the context of this study, history referred to any events happened in Taiwan stock market in 2016. The measures which Taiwan Securities and Futures Bureau implemented in 2016, such as raised minimum margin requirement for short sales, amended regulations on public tender offers, and others (Securities and Futures Bureau, 2018b) were the events. The purposes of all new measures were to create a fair, transparent, and efficient market. Some individual investors had to have changed their herd behavior as a result of the events, so history was relevant. However, the infeasibility of exposing the events to one group of individual investors but not another group was the limitation of history to this study. In the context of this study, maturation referred to a natural process through which individual investors got more mature in the sampling period. An improvement in financial literacy, exposure to objective investment information, and continuous reviews on investment strategy and performance were examples of natural processes through which individual investors might have changed their herd behavior. Hence, maturation was relevant. However, the lack of data about the natural processes at individual investor level was a limitation of maturation to this study. In the context of this study, experimental mortality referred to the drop-outs of individual investors in the sampling period. Some individual investors had to have traded in early 2016 but not anymore in late 2016. Such individual investors were the drop-outs, so experimental mortality was relevant. However, the drop-outs were unidentifiable, therefore unquantifiable, in the anonymous TWSE data. The unquantifiable drop-out was a limitation of experimental mortality to this study.

For external validity, there are two potential limitations, namely reactive arrangement and representativeness of the sample. Reactive arrangement was irrelevant for no experimental setting in this study. The stock exchange was a natural setting. All investor trades in the stock exchange were authentic actions. There was no experimental setting in this study for investors to react, therefore reactive arrangement was irrelevant. Representativeness of the sample was high. The samples of this study consisted of the majority, 94%, of the population. Illiquid or exchange sanctioned stocks were the only portion unrepresented. Furthermore, there was no experiment in this study. Refusal rate which could affect the representativeness of the sample was not relevant. Overall, representativeness of the sample was high.

Reliability

In the context of this study, reliability referred to the consistency of the herd behavior measurements. There was no instrument for the samples to respond and provide data for herd behavior measurement, therefore reliability assessment in a pretest or posttest approach was not applicable. I estimated the herd behavior with the mathematical formula of the LSV measure and the static TWSE data. The mathematical formula of the

LSV measure was standard since its publication. The LSV measure results would be consistent regardless how many times of estimation. Hence, the reliability of this study was high.

Generalizability

Generalizability referred to whether or not the findings of herd behavior could be generalizable to the population, a different time, or a different market. In this study, the samples consisted of the majority, 94%, of the population. The findings from such a majority were already generalized and should be generalizable to the individual investors of Taiwan. Besides, the data span of 1 year was long enough to capture many variations of the market. The findings from such variations were probably generalizable to certain months or years before and after the sampling period. Any relationship between herd behavior and industry affiliation unlikely held permanently due to the business cycle of each industry, therefore the farness of the certain months or years from the sampling period was the limitation to generalizability to a different time. The findings of more serious herding in either high market capitalization or high P/B ratio stocks were already generalized between Taiwan and China and probably generalizable to different markets. Overall, the limitations to the generalizability of this study were negligible.

Use of Secondary Data

There are primary limitations specific to the use of secondary data, namely a gap between research purposes, limited data accessibility, and information insufficiency (Frankfort-Nachmias et al., 2015). All three primary limitations were not relevant to this study. First, there was no gap between research purposes. TWSE did not collect the data for a research purpose but for its process of stock exchanges among investors. TWSE data were not biased for any research, therefore the gap was not even present. Second, limited data accessibility was least relevant. TWSE released an array of data from trade orders to trade transactions. TWSE did not make a part of data inaccessible for no reason, so limited data accessibility was not an issue in this study. Lastly, information insufficiency was least relevant too. TWSE described data in detail in the materials on its website. TWSE even answered questions about data through its inquiry hotline. Information insufficiency was not a limitation in this study either. There is a secondary limitation, code variation, specific to the use of secondary data (Frankfort-Nachmias et al., 2015). Code variation refers to that the current code of a data field represents finer or more precise than the old code of the same data field. I confirmed no code variation in the data fields of this study in the sampling period. Code variation was not an issue in this study. There is a limitation, authenticity, specific to the use of private records (Frankfort-Nachmias et al., 2015). TWSE was a financial institute regulated by Financial Supervisory Commission R.O.C. (Taiwan). Technically speaking, TWSE data were private records. TWSE did not release the private records unconditionally but after its approval of an application. TWSE data were authentic and official, therefore authenticity is not a limitation in this study.

TWSE Data

Data not connectable to individual investor level was a limitation specific to this study. The number of buyers and number of sellers are two key elements of the LSV measure. Nevertheless, TWSE did not release trade transaction data along with other information which enabled trade transaction connections to individual investor level. I could not compute the actual number of buyers and actual number of sellers. I adopted Chang et al.'s (2012) and Lin et al.'s (2013) approach of applying the number of buy transactions and number of sell transactions instead in the LSV measure. The implicit assumption was that each transaction belonged to a unique individual investor. Such assumption was a mitigation to the limitation of TWSE data not connectable to individual investor level. Nonnormality of the LSV estimates was another limitation from TWSE data. The distribution of the 196,631 stock-days' LSV estimates was not normal. The distributions of the subset LSV estimates for buy- and sell-herding assessments were not normal either. The distributions were not normal even after Box-Cox transformation. Given the nonnormality of LSV estimates, I resorted to chi-square tests for the subsequent hypothesis tests. Chi-square test is a nonparametric test whose assumptions are fewer.

Overall Limitation

To this study, there was a combined limitation of history, maturation, and experimental mortality to internal validity but least limitation to external validity. The overall validity was typical of causal-comparative design. In tandem, there were negligible limitations to reliability, generalizability, and use of secondary data. There were bigger limitations in TWSE data. I mitigated the bigger limitations with the adoptions of an implicit assumption and an alternative multivariate analysis method. The overall limitations of this study were typical and mitigable, therefore acceptable.

Recommendations

Although the overall limitations were acceptable, two key limitations can be the grounds for further research. They are data not connectable to individual investor level and the combined limitation of history, maturation, and experimental mortality. Besides, the combined strength of external validity, reliability, generalizability, and use of secondary data can be another ground for further research.

Combined Strength of the Study

In this study, I may have set an example of applying logistic regression upon a series of stock characteristics as independent variables and a dichotomous herd behavior variable based on LSV measure. I revealed the net difference in herd behavior by one stock characteristic with other stock characteristics controlled. Researchers may replicate this example with introductions of other stock characteristics. The combined strength of external validity, reliability, generalizability, and use of secondary data can be the ground for replication. Researchers may first replicate this example to get a reconciled baseline, then introduce other stock characteristics. TWSE publishes at stock-month level price-to-earnings ratio, yield, year-on-year change of trading sales revenue, year-on-year change of endorsed borrowing, and net profit after tax. Researchers may further include any appropriate stock characteristics as independent variables. Researchers can even go beyond TWSE data to stock characteristics data from other sources. Examinations of different stock characteristics will be a series of further research.

.

Limitations of the Study

On another hand, there were two key limitations to this study, data not connectable to individual investor level and a combined limitation of history, maturation, and experimental mortality. Number of buyers and number of sellers are investor level information which is fundamental to the LSV measure; however, data connectable to individual investor level are rare in stock exchange published data. I did not come across a previous study with a solution other than assuming a trade transaction as an individual investor. Choi (2016) used a different data source, a Korean securities brokerage firm. The data from such source were at individual investor level which was the best input to the LSV measure estimation. If researchers can observe personal data privacy ordinance and tap in investor data of a securities brokerage firm, the data will probably be connectable to individual investor level. Besides, the combined limitation of history, maturation, and experimental mortality was due to the nonexperimental nature of this study. By nonexperimental nature, there was no splitting of individual investors in groups and no intervention such as exposure to objective investment information. Tapping in the data of a securities brokerage firm may resolve a part of the combined limitation. For example, researchers can compare the extents of herd behavior of a particular group of investors between two times. It may become the maturation effect. In a similar manner, experimental mortality can be trackable. Nevertheless, it is difficult to mitigate the limitation of history.

Three Rooms for Further Research

On the grounds of strengths and limitations of this study, I recommend three rooms for further research. The first is inclusions of other TWSE published data and stock characteristics data of other sources. The second is to tap into the data connectable to individual investor level from a securities brokerage firm. With the data from a securities brokerage firm, the last is to examine maturation and experimental mortality of individual herding. Apart from the rooms for further research, I also recommend implementing the findings of this study in the society.

Implications

Empirical Implications

Individual herding happened from July 2006 to June 2008 (Chang et al., 2012). From this study, individual herding happened in 2016 too. Individual herding probably had been happening across the 8 to 10 years and will continue. Moreover, individual herding happened in more than half of the 2016 stock-days. Given its effect of inferior investment performance, it is necessary to alleviate individual herding. The finding of this study is a solution to the specific problem – a lack of knowledge about individual herding as related to characteristics of stocks.

Positive Social Change

A spread of the knowledge may improve individual investor financial literacy that may, in turn, alleviate herding and its effects. I will take a two-pronged approach. One, I will approach the management of nonprofit organizations whose missions are to help create a fair, transparent, and efficient market. One target will be Securities and Futures

Institute. I will propose to post the abstract and the full version of this study respectively in the education promotion, and research and development options under the organizational option of the menu bar on their website. Two, I will propose to Taiwan Securities and Futures Bureau whose directive is to ensure a fair and efficient market environment. I will present to its management the full version of this study. As TWSE does not release the latest 12 months trade transaction data, I will highlight the combined limitation of history, maturation, and experimental mortality to the validity and suggest validating my findings with data up to August 2017. Regardless the results of the additional validation, I will summarize all findings and propose to post the summary in the education promotion under investor area in their website. From the two websites, individual investors can access the abstract, summary, and the full version of this study to improve their financial literacy in the topic of individual herding. This may result in a positive social change of alleviations of individual investors' herd behavior and inferior investment performance. A portion of individual investors invests for the family living in future; hence, the potential impact for the positive social change will be at individual and family level. If Taiwan Securities and Futures Bureau proactively promotes the knowledge of individual herding to the over 500,000 individual investors trading monthly, the potential impact for the positive social change may become at a societal level as well.

Conclusion

In this study, I identified a lack of knowledge about individual herding as related to characteristics of stocks as the specific problem. Such lack of knowledge impedes individual investors in Taiwan from improving financial literacy that may alleviate herding and its effects. I then defined the overarching research question – what differences in individual investor herd behavior, if any, existed by characteristics of stocks. The characteristics of stocks included market capitalization, P/B ratio, and industry affiliation. With individual investor trade transaction data in 2016 from TWSE, I found the three characteristics of stocks separately and as a whole related to individual herding. I had six major findings to contribute to the body of behavioral finance knowledge, namely (a) the extension of the knowledge to comparable levels of individual herding in Taiwan between Chang et al.'s (2012) sampling period, July 2006 to June 2008 and the sampling period of this study, 2016; (b) the confirmation of the knowledge of higher sell-herding than buy-herding; (c) the confirmation of the knowledge of more serious herding in high market capitalization stocks; (d) the extension of the knowledge to more linearity between the logarithmic market capitalization and the log odds of individual herding; (e) the disconfirmation of the knowledge of more serious herding in high P/B ratio stocks and less serious herding in low P/B ratio stocks without other independent variables controlled and the extension of the knowledge to less serious herding in high P/B ratio stocks with other independent variables controlled; and (f) the confirmation of the knowledge of broad industry herding like in Europe, but the disconfirmation the knowledge of minor industry herding like in the United States, and the extension of the knowledge to the differences by industry affiliations with other independent variables controlled. The overall limitations of this study were typical and mitigable, therefore acceptable. I potentially can fill Taiwan individual investors in with this new knowledge. I propose to spread the new knowledge through posting on the websites of governmental organizations and nongovernmental organizations. A spread of the new knowledge may improve individual investor financial literacy that may, in turn, alleviate herding and its effects. This may result in a positive social change of alleviations of individual investors' herd behavior and inferior investment performance. A portion of individual investors invests for the family living in future; hence, the potential impact for the positive social change will be at individual and family level. If Taiwan Securities and Futures Bureau proactively promotes the knowledge of individual herding, the potential impact for the positive social change may become societal level as well.

References

Ahmed, W. M. A. (2014). The trading patterns and performance of individual vis-à-vis institutional investors in the Qatar Exchange. *Review of Accounting and Finance*, *13*(1), 24-42. doi:10.1108/RAF-09-2012-0089

Ahsan, A. F. M., & Sarkar, A. H. (2013). Herding in Dhaka stock exchange. *Journal of Applied Business and Economics*, *14*(2), 11-19. Retrieved from http://www.na-businesspress.com/JABE/AhsanAFMM_Web14_2_.pdf

Andersson, M., Hedesström, M., & Gärling, T. (2014). A social-psychological perspective on herding in stock markets. *Journal of Behavioral Finance*, *15*(3), 226-234. doi:10.1080/15427560.2014.941062

Andrikopoulos, P., Albin Hoefer, A., & Kallinterakis, V. (2014). On the impact of market mergers over herding: Evidence from EURONEXT. *Review of Behavioral Finance*, *6*(2), 104-135. doi:10.1108/RBF-12-07-10

Arjoon, V., & Bhatnagar, C. S. (2017). Dynamic herding analysis in a frontier market. *Research in International Business and Finance*, *42*, 496-508. doi:10.1016/j.ribaf.2017.01.006

Bakar, S., & Yi, A. N. C. (2016). The impact of psychological factors on investors' decision making in Malaysian stock market: A case of Klang Valley and Pahang. *Procedia Economics and Finance*, *35*, 319-328. doi:10.1016/S2212-5671(16)00040-X

Barber, B. M., Odean, T., & Zhu, N. (2009). Do retail trades move markets? *Review of Financial Studies*, *22*(1), 151-186. doi:10.1093/rfs/hhn035

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom and cultural change as informational cascades. *Journal of Political Economy, 100*, 992-1026.

Bikhchandani, S., & Sharma, S. (2001). Herd behavior in financial markets. *IMF Economic Review*, *47*(3), 279-310. doi:10.2307/3867650

Blasco, N., Corredor, P., & Ferreruela, S. (2012). Does herding affect volatility? Implications for the Spanish stock market. *Quantitative Finance*, *12*(2), 311-327. doi:10.1080/14697688.2010.516766

Brewer, E. W., & Kuhn, J. (2010). Causal-comparative design. In N. J. Salkind (Eds.), *Encyclopedia of research design volume 1* (pp. 124-131). Thousand Oaks, CA: SAGE. doi:10.4135/9781412961288.n42

Bucher-Koenen, T., & Ziegelmeyer, M. (2013). Once burned, twice shy? Financial literacy and wealth losses during the financial crisis*. *Review of Finance, 18*, 2215-2246. doi:10.1093/rof/rft052

Chang, C. Y., Chen, H. L., & Jiang, Z. R. (2012). Portfolio performance in relation to herding behavior in the Taiwan stock market. *Emerging Markets Finance and Trade, 48*(Sup2), 82-104. doi:10.2753/REE1540-496X48S205

Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, *24*(10), 1651-1679. doi:10.1016/S0378-4266(99)00096-5

Chen, J. J., Zheng, B., & Tan, L. (2013). Agent-based model with asymmetric trading and herding for complex financial systems. *PloS One*, *8*(11), 1-11. doi:10.1371/journal.pone.0079531

Chen, Y. F., Yang, S. Y., & Lin, F. L. (2012). Foreign institutional industrial herding in Taiwan stock market. *Managerial Finance*, *38*(3), 325-340. doi:10.1108/03074351211201442

Choi, N., & Skiba, H. (2015). Institutional herding in international markets. *Journal of Banking & Finance*, *55*, 246-259. doi:10.1016/j.jbankfin.2015.02.002

Choi, S. (2016). Herding among local individual investors: Evidence from online and offline trading. *Economics Letters, 144*, 4-6. doi:10.1016/j.econlet.2016.04.030

Christie, W. G., & Huang, R. D. (1995). Following the pied piper: Do individual returns herd around the market? *Financial Analysts Journal*, *51*(4), 31-37. Retrieved from http://www.jstor.org/stable/4479855

Chung, C. Y., & Wang, K. (2016). Short-term trading by individual investors in the Korean stock market. *Journal of the Asia Pacific Economy*, 1-13. doi:10.1080/13547860.2016.1187798

Cinar, D. (2014). A market analysis approach to portfolio theories. In H. Dinçer & Ü. Hacioğlu (Eds.), *Global strategies in banking and finance* (pp. 241-252). Hershey, PA: IGI Global.

Delfino, A., Marengo, L., & Ploner, M. (2016). I did it your way. An experimental investigation of peer effects in investment choices. *Journal of Economic Psychology*, *54*, 113-123. doi:10.1016/j.joep.2016.03.005

Demirer, R., Kutan, A. M., & Zhang, H. (2014). Do ADR investors herd?: Evidence from advanced and emerging markets. *International Review of Economics & Finance*, *30,* 138-148. doi:10.1016/j.iref.2013.10.006

Devenow, A., & Welch, I. (1996). Rational herding in financial economics. *European Economic Review, 40,* 603-615. doi:10.1016/0014-2921(95)00073-9

Durand, R. B., Limkriangkrai, M., & Fung, L. (2014). The behavioral basis of sell-side analysts' herding. *Journal of Contemporary Accounting & Economics, 10*(3), 176-190. doi:10.1016/j.jcae.2014.08.001

Duxbury, D. (2015). Behavioral finance: Insights from experiments I: Theory and financial markets. *Review of Behavioral Finance, 7*(1), 78-96. doi:10.1108/RBF-03-2015-0011

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383-417. doi:10.2307/2325486

Field, A. (2013). *Discovering statistics using IBM SPSS statistics, 4th edition*, 736-746, 762-763. London: Sage.

Filip, A. M., & Pochea, M. M. (2014). Herding behavior under excessive volatility in CEE stock markets. *Studia Universitatis Babes-Bolyai: Oeconomica, 59*(3), 38-47. ProQuest Central Database, Document ID: 1652350188

FINRA Investor Education Foundation. (2017). *Take the investor literacy quiz.* Retrieved from http://www.usfinancialcapability.org/quiz-literacy.php

Flannery, R. (2016, July 6.) Taiwan tech industry faces "greatest challenge" as rival China backs nat'l champions. *Forbes*. Retrieved from https://www.forbes.com/sites/russellflannery/2016/07/06/taiwan-tech-industry-faces-greatest-challenge-as-rival-china-backs-natl-champions/#7911fd052433

Fong, K., Gallagher, D., & Lee, A. (2014). Individual investors and broker types. *Journal of Financial and Quantitative Analysis*, *49*(2), 431-451. doi:10.1017/S0022109014000349

Frankfort-Nachmias, C., Nachmias, D., & DeWaard, J. (2015). *Research methods in the social sciences* (8th ed.). New York, NY: Worth Publishers.

Garcia, M. T., & Oliveira, R. A. A. (2018). Value versus growth in PIIGS stock markets. *Journal of Economic Studies,* accepted article. doi:10.1108/JES-06-2017-0160

Garg, A. K., & Mitra, S. K. (2015). A study of lead-lag relation between FIIs herding and stock market returns in emerging economies: Evidence from India. *Decision*, *42*(3), 279-292. doi:10.1007/s40622-015-0080-6

Gavriilidis, K., Kallinterakis, V., & Ferreira, M. P. L. (2013). Institutional industry herding: Intentional or spurious? *Journal of International Financial Markets, Institutions and Money*, *26*, 192-214. doi: 10.1016/j.intfin.2013.05.008

Gavriilidis, K., Kallinterakis, V., & Tsalavoutas, I. (2016). Investor mood, herding and the Ramadan effect. *Journal of Economic Behavior & Organization*, *132*, 23-38. doi:10.1016/j.jebo.2015.09.018

Gong, P., & Dai, J. (2017). Monetary policy, exchange rate fluctuation, and herding behavior in the stock market. *Journal of Business Research*, *76*, 34-43. doi:10.1016/j.jbusres.2017.02.018

Guvercin, A. (2016). Sentimental herding: The role of regional and global shocks in Egyptian and Saudi stock markets. *Sosyoekonomi, 24*(27), 25-32. doi:10.17233/se.53541

Hou, T. C. T., McKnight, P. J., & Weir, C. (2014). The impacts of stock characteristics and regulatory change on mutual fund herding in Taiwan. *Applied Financial Economics, 24*(3), 175-186. doi:10.1080/09603107.2013.866201

Huang, T. C., Lin, B. H., & Yang, T. H. (2015). Herd behavior and idiosyncratic volatility. *Journal of Business Research, 68*(4), 763-770. doi:10.1016/j.jbusres.2014.11.025

Huang, T. C., Wu, C. C., & Lin, B. H. (2016). Institutional herding and risk–return relationship. *Journal of Business Research, 69*, 2073-2080. doi:10.1016/j.jbusres.2015.12.011

Hwang, S., & Salmon, M. (2004). Market stress and herding. *Journal of Empirical Finance*, *11*(4), 585-616. doi:10.1016/j.jempfin.2004.04.003

Inagaki, K. (2016, August 12). Foxconn completes $3.8bn takeover of Sharp. *Financial Times*. Retrieved from https://www.ft.com/content/da1f3fc0-602c-11e6-b38c-7b39cbb1138a

Jame, R., & Tong, Q. (2014). Industry-based style investing. *Journal of Financial Markets*, *19*, 110-130. doi:10.1016/j.finmar.2013.08.004

Javaira, Z., & Hassan, A. (2015). An examination of herding behavior in Pakistani stock market. *International Journal of Emerging Markets, 10*(3), 474-490. doi:10.1108/IJoEM-07-2011-0064

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica, 47*, 263-291. doi:10.2307/1914185

Keynes, J. M. (1936). *The general theory of employment, interest and money.* London, England: Macmillan Publications.

Kindleberger, C. P. (2016). Bubbles in history. In: G. Jones (Eds.), *Banking crises* (pp. 37-38). London, England: Palgrave Macmillan. doi:10.1057/9781137553799_6

Internal Revenue Service. (2018, January 19). *Yearly average currency exchange rates*. Retrieved from https://www.irs.gov/individuals/international-taxpayers/yearly-average-currency-exchange-rates

Internet Society. (2017). *Global internet map.* Retrieved from http://www.internetsociety.org/map/global-internet-report/?gclid=CIWAxpCll9MCFU0HvAodfUcLTg#global-internet-penetration

Ivanov, A., Levin, D., & Peck, J. (2013). Behavioral biases in endogenous-timing herding games: An experimental study. *Journal of Economic Behavior & Organization*, *87*, 25-34. doi:10.1016/j.jebo.2012.12.001

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of Financial Economics*, *32*(1), 23-43. doi:10.1016/0304-405X(92)90023-Q

Lakshman, M. V., Basu, S., & Vaidyanathan, R. (2013). Market-wide herding and the impact of institutional investors in the Indian capital market. *Journal of Emerging Market Finance, 12*(2), 197-237. doi:10.1177/0972652713494046

Lam, K. S., & Qiao, Z. (2015). Herding and fundamental factors: The Hong Kong experience. *Pacific-Basin Finance Journal*, *32*, 160-188. doi: 10.1016/j.pacfin.2014.11.002

Lavin, J. F., & Magner, N. S. (2014). Herding in the mutual fund industry: Evidence from Chile. *Academia Revista Latinoamericana de Administración*, *27*(1), 10-29. doi:10.1108/ARLA-09-2013-013

Lin, A. (2016, February 6). Taiwan earthquake kills at least 14 as chipmaking hub rocked. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2016-02-06/over-120-rescued-as-buildings-collapse-in-taiwan-quake-tv-says

Lin, H. W. (2012). How herding bias could be derived from individual investor types and risk tolerance? In *Proceedings of World Academy of Science, Engineering and Technology* (Vol. 6, No. 6, pp. 831-836). World Academy of Science, Engineering and Technology. Retrieved from http://www.waset.org/publications/11343

Lin, W., Tsai, S. C., Lung, P. Y. (2013). Investors herd behavior: Rational or irrational. *Asia-Pacific Journal of Financial Studies, 42*, 755-776. doi:10.1111/ajfs.12030

Litimi, H., BenSaida, A., & Bouraoui, O. (2016). Herding and excessive risk in the American stock market: A sectoral analysis. *Research in International Business and Finance*, *38*, 6-21. doi:10.1016/j.ribaf.2016.03.008

Luchtenberg, K. F., & Seiler, M. J. (2013). The effect of exogenous information signal strength on herding. *Review of Behavioral Finance, 5*(2), 153-174. doi:10.1108/RBF-05-2012-0004

Mahmood, Z., Kouser, R., Abbas, S. S., & Saba, I. (2016). The effect of hueristics [*sic*], prospect and herding factors on investment performance. *Pakistan Journal of Social Sciences*, *36*(1), 475-484. Retrieved from http://www.bzu.edu.pk/PJSS/Vol36No12016/PJSS-Vol36-No1-42.pdf

Meier, S., & Sprenger, C. D. (2013). Discounting financial literacy: Time preferences and participation in financial education programs. *Journal of Economic Behavior & Organization*, *95*, 159-174. doi:10.1016/j.jebo.2012.02.024

Messis, P., & Zapranis, A. (2014a). Herding behaviour and volatility in the Athens Stock Exchange. *Journal of Risk Finance, 15*(5), 572-590. doi:10.1108/JRF-04-2014-0054

Messis, P., & Zapranis, A. (2014b). Herding towards higher moment CAPM, contagion of herding and macroeconomic shocks: Evidence from five major developed markets. *Journal of Behavioral and Experimental Finance, 4*, 1-13. doi:10.1016/j.jbef.2014.09.002

Nasdaq. (n.d.) Market capitalization. Retrieved from http://www.nasdaq.com/investing/glossary/m/market-capitalization

Noble Prize Organization. (2002, December 12). *Interview with Daniel Kahneman and Vernon Smith* (K. Lofgren & A. Crepin, Interviewers) [Interview transcript]. Retrieved from https://www.nobelprize.org/prizes/economics/2002/kahneman/25977-interview-transcript-2002-2/

Nguyen, T., & Schuessler, A. (2013). Do they trade as they say? Comparing survey data and trading records. *Journal of Asset Management*, *14*(1), 37-51. doi:10.1057/jam.2012.24

Ortiz, C., Sarto, J. L., & Vicente, L. (2013). Herding behaviour in Spanish global funds country allocations. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, *42*(157), 63-81. doi:10.1080/02102412.2013.10779740

Ouarda, M., El Bouri, A., & Bernard, O. (2013). Herding behavior under markets condition: Empirical evidence on the European financial markets. *International Journal of Economics and Financial Issues*, *3*(1), 214-228. Retrieved from http://www.econjournals.com/index.php/ijefi/article/viewFile/229/pdf

Park, H., & Sohn, W. (2013). Behavioral finance: A survey of the literature and recent development. *Seoul Journal of Business*, *19*(1), 3-42. Retrieved from http://s-space.snu.ac.kr/handle/10371/83787

Poshakwale, S., & Mandal, A. (2014). Investor behaviour and herding: Evidence from the national stock exchange in India. *Journal of Emerging Market Finance*, *13*(2), 197-216. doi:10.1177/0972652714541341

Ramadan, I. Z. (2015). Cross-sectional absolute deviation approach for testing the herd behavior theory: The case of the ASE Index. *International Journal of Economics and Finance, 7*(3), 188-193. doi:10.5539/ijef.v7n3p188

Ramli, I., Agoes, S., & Setyawan, I. R. (2015). Information asymmetry and the role of foreign investors in daily transactions during the crisis; A study of herding in the

Indonesian Stock Exchange. *Journal of Applied Business Research*, *32*(1), 269-288. doi:10.19030/jabr.v32i1.9537

Roszkowski, A. J., & Richie, N. (2016). The impact of mad money recommendations during bull and bear markets. *International Journal of Managerial Finance*, *12*(1), 52-70. doi:10.1108/IJMF-04-2014-0053

Royal Swedish Academy of Sciences. (2002). Foundations of behavioral and experimental economics: Daniel Kahneman and Vernon Smith. Retrieved from http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2002/advanced-economicsciences2002.pdf

Securities and Futures Bureau. (2018a). *Services and protections for investors*. Retrieved from https://www.sfb.gov.tw/ch/home.jsp?id=5&parentpath=0

Securities and Futures Bureau. (2018b). *Important measures*. Retrieved from https://www.sfb.gov.tw/en/home.jsp?id=79&parentpath=0,4

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, *71*, 421–436. doi:10.1016/j.najef.2013.10.003

Shin, S., & Yoon, S. (2012, February 20) South Korea police arrests 6 over spreading N. Korea rumors. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2012-02-21/six-arrested-in-s-korea-over-nuclear-rumors

Sias, R. W. (2004). Institutional herding. *Review of Financial Studies*, *17*(1), 165-206. doi:10.1093/rfs/hhg035

Simon, H. A. (1979). Rational decision making in business organizations. *American Economic Review, 69*(4), 493-513. Retrieved from http://www.jstor.org/stable/1808698

Solakoglu, M. N., & Demir, N. (2014). Sentimental herding in Borsa Istanbul: Informed versus uninformed. *Applied Economics Letters*, *21*(14), 965-968. doi: 10.1080/13504851.2014.902015

Spyrou, S. (2013). Herding in financial markets: a review of the literature. *Review of Behavioral Finance*, *5*(2), 175-194. doi:10.1108/RBF-02-2013-0009

Taiwan Stock Exchange. (n.d.). *Market capitalization weekly statistics* [Excel file]. Available from http://www.twse.com.tw/en/statistics/statistics_week.php

Taiwan Stock Exchange. (2004a). *Data eShop past transaction data file* [MS Word file]. Available from http://dataeshop.twse.com.tw/frontend/cht/index.jsp

Taiwan Stock Exchange. (2004b). *Data eShop past transaction data file* [MS Word file]. Available from http://dataeshop.twse.com.tw/frontend/en/index.jsp

Taiwan Stock Exchange. (2018a). *List of ISIN code for listed equities.* Retrieved from http://isin.twse.com.tw/isin/e_C_public.jsp?strMode=2

Taiwan Stock Exchange. (2018b). *Statistics of securities market: Monthly statistics* [Data file]. Available from http://www.twse.com.tw/en/statistics/statisticsList?type=02&subType=199

Taiwan Stock Exchange. (2018c). *P/E ratio and yield and P/B ratio of listed stocks: Monthly statistics* [Data file]. Available from http://www.twse.com.tw/en/statistics/statisticsList?type=04&subType=220

Taiwan Stock Exchange. (2018d). *Status of securities listed on TWSE* [Data file]. Available from http://www.twse.com.tw/en/statistics/statisticsList?type=04&subType=223

Taiwan Stock Exchange. (2018e). *Latest listed companies* [Data file]. Available from http://www.twse.com.tw/en/page/listed/listed_company/new_listing.html

Taiwan Stock Exchange. (2018f). *De-listed companies* [Data file]. Available from http://www.twse.com.tw/en/company/suspendListing

Tsai's brighter side. (2017, Jun 1). *Economist, 423*(9043), 61. Retrieved from https://www.economist.com/finance-and-economics/2017/06/01/taiwans-economy-has-defied-the-pessimists?zid=306&ah=1b164dbd43b0cb27ba0d4c3b12a5e227

Vo, X. V., & Phan, D. B. A. (2017). Further evidence on the herd behavior in Vietnam stock market. *Journal of Behavioral and Experimental Finance, 13*, 33-41. doi: 10.1016/j.jbef.2017.02.003

Walden University Institutional Review Board. (2015, October 20). *Standard application for research ethics review* [MS Word document]. Retrieved from http://academicguides.waldenu.edu/researchcenter/orec/application

Wang, S. (2016, May 26). ASE agrees to merge with Siliconware through new holding company. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2016-05-26/ase-agrees-to-merge-with-siliconware-through-new-holding-company

Wang, Y., & Ma, J. (2014). Excess volatility and the cross-section of stock returns. *North American Journal of Economics and Finance*, *27*, 1-16. doi:10.1016/j.najef.2013.10.003

Wermers, R. (1999). Mutual fund herding and the impact on stock prices. *Journal of Finance*, *54*(2), 581-622. doi:10.1111/0022-1082.00118

Yang, S. J., Hsu, A. C., Lai, S. Y., & Lee, C. C. (2015). Empirical investigation of herding behavior in East Asian stock markets toward the U.S. market. *International Journal of Business and Finance Research*, *9*(1), 19-32. Retrieved from http://www.theibfr2.com/RePEc/ibf/ijbfre/ijbfr-v9n1-2015/IJBFR-V9N1-2015-2.pdf

Yao, J., Ma, C., & He, W. P. (2014). Investor herding behaviour of Chinese stock market. *International Review of Economics & Finance*, *29*, 12-29. doi:10.1016/j.iref.2013.03.002

Zhang, Y., & Zheng, X. (2016). A study of herd behavior based on the Chinese stock market. *Journal of Applied Management and Investments*, *5(2)*, 131-135. Retrieved from http://www.jami.org.ua/Papers/JAMI_5_2_131-135.pdf

Zheng, D., Li, H., & Zhu, X. (2015). Herding behavior in institutional investors: Evidence from China’s stock market. *Journal of Multinational Financial Management*, *32*, 59-76. doi:10.1016/j.mulfin.2015.09.001

Appendix A: IRB Approval Number

02-19-18-0467816

Appendix B: Email to Prof. Lakonishok

From: Josef Lakonishok
Sent: Tuesday, July 31, 2018 11:42 PM
To: Tze Sun Wong
Subject: Re: My Use of Lakonishok, Shleifer, and Vishny's measure

I approve

Sent from my iPhone

On Jul 31, 2018, at 08:09, Tze Sun Wong wrote:

Dear Prof. Lakonishok,

I am currently working on my PhD dissertation. The topic is Characteristics of Stocks and Individual Investor Herd Behavior: A Causal-Comparative Study. I would like to seek your approval of my adoption of Lakonishok, Shleifer, and Vishny's measure as the herd measure in my study.

I would like to thank you for your magnificent work which set a new course for subsequent researchers including me.

Yours truly,
Tze Sun