Vaasan yliopisto
UNIVERSITY OF VAASA

Elias Heikkilä

Investor Herding and Stock Market Bubbles

An Investigation of the S&P 500 Index

School of Accounting and Finance
Master's Thesis in Finance
Master's Degree Programme in Finance

Vaasa 2025

UNIVERSITY OF VAASA
School of Accounting and Finance
Author: Elias Heikkilä
Title of the thesis: Investor Herding and Stock Market Bubbles : An Investigation of the S&P 500 Index
Degree: Master of Science in Economics and Business Administration
Discipline: Master's Degree Programme in Finance
Supervisor: John Kihn
Year: 2025 **Pages:** 68

ABSTRACT :

This thesis studies investor herding as it is a prominent bias in human behavior, generally leading to irrational outcomes. The tendency to herd also concerns investors, which causes the widespread effects regarding asset markets, thus spreading to economies. Stock market bubbles are another prominent phenomenon in the stock markets. Still, they have gained surprisingly limited attention in academic literature. Bubbles may cause excessive valuations from fundamental values and eventually lead to crashes when the bubble bursts.

Investor herding and stock market bubbles are generally considered to arise from irrational investor behavior which leads to inefficient outcomes. These observations serve as motivation for this thesis. Investor herding is examined on S&P 500 index during the period 1989-2025, with two herding measures, the cross-sectional absolute deviation, and the quantile regression method. The possible relationship between investor herding and bubbles is examined by incorporating a bubbly dummy variable into the quantile regression equation. The objective is then to determine whether there are differences in herding behavior during normal and bubble periods.

This thesis investigates investor herding and stock market bubbles in the S&P 500 index by forming a fixed and a dynamic index composition. This enables one to study possible differences in herding on the separate index compositions. Herding is also studied in sub-periods for the dynamic index composition.

The results for this study are somewhat inconclusive. There are differences in herding between the used measure of CSAD and the quantile regression. Quantile regression seems to better display differences in herding during varying levels of stock return dispersion. Sub-period quantile regression analysis reveals the notable variation in herding results for different sub-periods. There seems to be significant differences in herding between normal and bubble periods based on the quantile regression with a bubble dummy variable.

Based on the results, it seems that in some circumstances, herding is a matter of concern for stock markets. Still, the inconclusive results signal that there is an apparent need for new methods which are better oriented to measure herding. In addition, herding and bubbles seem to have a relationship, which requires increased academic attention.

KEYWORDS: Behavioral finance, Herding behavior, S&P 500, Stock markets, Cross-sectional absolute deviation, Quantile regression, Stock market bubbles

VAASAN YLIOPISTO
School of Accounting and Finance

Tekijä:	Elias Heikkilä
Tutkielman nimi:	Investor Herding and Stock Market Bubbles: An Investigation of the S&P 500 Index
Tutkinto:	Kauppatieteiden maisteri
Oppiaine:	Rahoitus
Työn ohjaaja:	John Kihn
Valmistumisvuosi:	2025 **Sivumäärä:** 68

TIIVISTELMÄ:

Tässä tutkielmassa tarkastellaan sijoittajien laumakäyttäytymistä, joka on saanut viime vuosina kasvavaa huomiota. Laumakäyttäytyminen on yleinen käyttäytymisvinouma, joka usein johtaa epärationaalisiin lopputulemiin. Taipumus laumakäyttäytymiseen koskee myös sijoittajia, joka aiheuttaa laajalle leviäviä vaikutuksia pääomamarkkinoilla, täten leviten talouksiin. Osakemarkkinakuplat ovat toinen huomiota herättävä ilmiö osakemarkkinoilla. Silti, ne ovat saaneet yllättävän rajattua huomiota osakseen. Kuplat voivat aiheuttaa liiallisia arvostuksia fundamentteihin nähden ja lopulta johtaa romahduksiin, kun kupla puhkeaa.

Sijoittajien laumakäyttäytymisen ja osakemarkkinakuplien muodostumisen nähdään yleensä alkavan sijoittajien epärationaalisesta käyttäytymisestä, joka johtaa epätehokkaisiin lopputulemiin. Nämä havainnot toimivat motivaationa tälle tutkielmalle. Tarkoituksena on tutkia sijoittajien laumakäyttäytymistä S&P 500 indeksissä, käyttäen kahta laumakäyttäytymismittaria. Mahdollista yhteyttä laumakäyttäytymisen ja kuplien välillä tutkitaan sisällyttämällä kuplamuuttuja laumakäyttäytymisanalyysiin. Tavoite on sen jälkeen määrittää, onko normaalien- ja kuplaperiodien välillä eroa.

Tässä tutkimuksessa laumakäyttäytymistä ja osakemarkkinakuplia tutkitaan muodostamalla kiinteä ja dynaaminen indeksikoostumus. Tämä mahdollistaa laumakäyttäytymisen eroavaisuuksien tutkimisen eri indeksikoostumusten välillä. Laumakäyttäytymistä tutkitaan myös alaperiodien osalta dynaamisesta indeksistä.

Tämän tutkimuksen tulokset ovat jossain määrin epäjohdonmukaisia. Laumakäyttäytymisessä on tulosten osalta eri metodien välillä. Eri kvantiileille suoritettu regressio paljastaa erovaisuuksia tulosten välillä eri kvantiileissa. Alaperiodien analysointi paljastaa merkittäviä erovaisuuksia tuloksissa eri alaperiodien välillä. Laumakäyttäytymisanalyysi kuplamuuttujalla osoittaa eroavaisuuksia tuloksissa normaalien- ja kuplaperiodien välillä.

Tulosten perusteella laumakäyttäytyminen on joissain tapauksissa merkittävä tekijä osakemarkkinoilla. Silti, epäjohdonmukaiset tulokset viestittävät, että uusille laumakäyttäytymistä paremmin mittaaville menetelmille on selvä tarve. Lisäksi laumakäyttäytymisen ja kuplien välillä vaikuttaa olevan yhteys, joka vaatii lisää akateemista huomiota.

AVAINSANAT: Behavioral finance, Herding behavior, S&P 500, Stock markets, Cross-sectional absolute deviation, Quantile regression, Stock market bubbles

Contents

Figures

Tables

Abbreviations

ADF	Augmented Dickey-Fuller
AMEX	American Stock Exchange
APT	Arbitrage Pricing Theory
BE	Book Value of Equity
CAPM	Capital Asset Pricing Model
CML	Capital Market Line
CSAD	Cross-Sectional Absolute Deviation
CSSD	Cross-Sectional Standard Deviation
E/P	Earnings-to-Price
EMH	Efficient Market Hypothesis

EUT	Expected Utility Theory
EZC	Eurozone Crisis
FPR	False-Positive Rate
GDP	Gross Domestic Product
GFC	Global Financial Crisis
GSADF	Generalized Sup Augmented Dickey-Fuller
HML	High Minus Low
IPO	Initial Public Offering
M&A	Mergers and Acquisitions
ME	Market Equity
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
QR	Quantile Regression
S&P	Standard and Poor’s
SADF	Sup Augmented Dickey-Fuller
SVI	Search Volume Index
TPR	True-Positive Rate
US	United States

1 Introduction

Herding behavior is the human tendency of following the decisions and behavior of others in a “herd”. Herding may cause people to rely excessively on decisions of others instead of making necessary considerations and analysis themselves. This may lead to making decisions that can be considered irrational. The effect of herding has recently been increasingly recognized in behavioral finance as it is a prominent behavioral bias behind human behavior. Herding behavior is a significant matter of concern if it affects financial decision making and this way also investing decisions and behavior, and thus whole economies.

Stock market bubbles and crashes are financial market phenomena which have not gained the amount of academic attention which could be expected. This might be surprising as they may have a significant impact on the functioning of asset markets and the economic cycles. One explanation could be the complexity of the matter and the difficulty of quantitatively measuring bubbles and crashes effectively. The need to understand the phenomena better and to create more suitable scientific methods is high in the increasingly complex economies and the turbulent world state.

An observation worth mentioning regarding academic finance is the tendency to explicate complicated matters with unrealistic assumptions. The reason for academic approval of theoretical models that do not represent reality, could stem from obviating the need to confront the complex reality, which often contradicts with widely accepted mathematical models that rely on expected rationality.

The objective of this thesis is to deepen the understanding behind investor herding, and to examine whether it has a relationship with market bubbles. The objective is pursued by using historical data of S&P 500 as it is a significant index and the results could be generalized to other mature markets that have similar market dynamics. The aspiration is to measure the occurrence of herding with the cross-sectional absolute deviation

(CSAD) and quantile regression (QR) method. There will also be a QR analysis with a bubble dummy variable to examine if there is a relationship between investor herding and stock market bubbles.

1.1 Purpose of the Study

The purpose of this thesis is to study if investor herding behavior can be detected in the United States (US) on a fixed S&P 500 index and it will be investigated whether there are differences when compared to a dynamic S&P 500 composition. Possible historical variation in herding is studied for different sub-periods. This study also investigates the relationship between herding and stock market bubbles. The relationship between investor herding and stock market bubbles is investigated to increase understanding in the field that has gained limited academic interest.

1.2 Research Hypotheses

The previous research on this topic seems to be limited. Bekiros et al. (2017) and Wang et al. (2024) are among the first to have studied quantitatively whether there is a relationship between herding behavior and stock market bubbles. To the best of my knowledge this is the first study to examine investor herding on a fixed and dynamic S&P 500 index composition with a QR, but there have been other QR studies on other markets, such as the studies by Bekiros et al. (2017) and Chiang et al. (2010).

When studying investor herding and its relationship with stock market bubbles and crashes, a logical expectation is that herding behavior exists in the financial markets. The null hypothesis is derived from the rational asset pricing theories. Random walk hypothesis by Fama (1965) claims that stock returns follow a random pattern, and by time the returns create a normal distribution. The efficient market hypothesis (EMH) can be considered an expansion to the random walk hypothesis. EMH shares the expectation of normally distributed returns, adding the assumption of normally distributed stock return dispersions. Thus, if the mentioned traditional asset pricing theory principles do not hold, this would enable herding to exist. These demonstrations lead us to first null hypothesis:

H_0^1: Based on the descriptive statistics of the S&P 500 index returns and the stock return dispersions, measured by the CSAD, the market returns, and the stocks return dispersions are normally distributed.

If the null hypothesis H_0^1 can be rejected successfully, we determine whether the opposite holds for market returns and CSADs. Thus, the alternative hypothesis will be the following:

H_a^1: According to the descriptive statistics of the S&P 500 returns and the stock return dispersions, measured by the CSAD, the market returns, and the stock return dispersions are not normally distributed.

If we can accept the alternative hypothesis H_a^1, this would suggest that investor herding could be a factor explaining why market returns and stock return dispersions do not follow a normal distribution. This would raise the need to examine whether herding occurs in the stock markets, but this is concurrently problematic as the CSAD method would require the market returns and stock return dispersions to be normally distributed. If the market returns and CSADs are not normally distributed, it would exclude using the CSAD measure. When using the CSAD method, we need to assume that the market returns and stock return dispersions are normally distributed, even though accepting the alternative hypothesis suggests the contrary.

The used methodology has two separate measures for herding, first one is a modified version of the original CSAD equation, and the second is a QR, also based on the CSAD measure. As the foundational measure is similar on the separate methods, the results could be expected to have similarities. These foundations enable to state the second null hypothesis as seen below:

H_0^2: The regression results between the used CSAD equation and the QR method are in line together.

If we can reject the null hypothesis H_0^2, we need to examine if the opposite seems to hold. The alternative hypothesis is formulated as follows:

H_a^2: The regression results between CSAD equation and the QR differ.

In the literature there is evidence which suggests that herding occurs in the financial markets. Bohl et al. (2017) argue that the approach of Chang et al. (2000) is biased against finding evidence for herding as they use a coefficient of zero for null hypothesis of no herding, when it should be positive. This suggests that the null hypothesis of no herding is rejected at the 5% significance level. According to Bohl et al. (2017), with the positive coefficient for the null hypothesis there is evidence for herding instead of the anti-herding. Wang et al. (2024) claim that in the Chinese stock market, investor herding has a significant effect on the market bubble formation and burst in extreme market conditions which are defined as the bull and the bear market. The continuous increase of investor herding will accelerate the formation of a price bubble especially in the bull market, and the decrease of the herding will dampen the creation of a bubble. If the bull market switches to bear market, the formed bubble may burst quickly. To investigate whether the previously described could hold, we first need to state the third null hypothesis as seen below:

H_0^3: Based on the methodology used, of QR with a bubble dummy variable, there is no difference in investor herding results between non-bubble and bubble periods.

Provided that the null hypothesis H_0^3 can be rejected, there could be a relationship between herding and stock market bubbles. Thus, the alternative hypothesis will be the following:

H_a^3: Based on the methodology used, there is a difference in herding coefficients for non-bubble and bubble periods.

1.3 Structure of the Study

There will be eight main chapters in this study. Chapter 1 is the introduction chapter. Chapter 2 covers traditional finance theories. Chapter 3 introduces non-traditional finance as an alternative to traditional finance. The literature review is covered in chapter 4. Chapter 5 states the used methodology. Data and descriptive statistics are addressed in chapter 6. The empirical results are presented in chapter 7. Chapter 8 covers the conclusions, after which will be the references.

2 Traditional Finance

Traditional finance means the standard finance theories that have been widely adopted and had a great significance in the financial world. The theories focus on financial market efficiency, investor rationality and the asset pricing. Traditional finance theories work well in theory, but they have limitations in practice, for example in terms of actual investor behavior and its biases.

Traditional finance theories are based on the expectation of a positive relationship between risk and expected return of an asset. This means that an investor could only achieve a higher expected return by tolerating more risk. For the rational asset pricing theories to hold, this relationship must always persist.

2.1 Random Walk Hypothesis

The concept behind random walk hypothesis dates back over 100 years, and it was popularized by Fama (1965). They state that there seem to be two different schools of thought for estimating future stock prices, the first of which are *technical* or *chartist* theories, and the second is analysis of fundamental or intrinsic value. The analysis techniques of chartists can be considered to lack concrete support for the decisions, and thus it probably is not used solely by many stock analysts. Most analysts can supposedly be considered fundamentalists, but for this method to work, the assumption is that a security has a true fundamental value at all times. The fundamental value of a security is based on all of the expected future earnings discounted to present. The only difficulty is that this would require the analyst to predict the future to know how factors affecting the earnings such as sales, costs, and management will change later. The security markets, consisting of all the separate market participants, are considered rational and intelligent, which allows for the market to account for events that have happened and events that are expected to happen when pricing assets. But there is always a varying amount of uncertainty in the future, which affects the asset prices. Thus, the author claims that the market prices will overreact as often as they underreact when new information arises. In addition, the change in prices to new expected intrinsic value is an independent random variable, and the adjustment in market prices may either precede the new information based on the expectations of the market or following the new information, which suggests that *cancellation* should correct pricing. This can be considered a catalyst for the random walk of asset prices, that causes unpredictable fluctuations in the asset values (Fama, 1965).

2.2 Efficient Market Hypothesis

The EMH by Fama (1970) suggests that efficient market prices should always “fully reflect” all fundamental information in an unbiased way. They claim that the assumption is so general that it cannot be tested empirically, which makes it necessary to define what “fully reflect” all the available information means. The EMH suggests three different levels of efficiency. The study suggests the lowest level of efficiency is the weak-form test, next level is the semi-strong form test, and the highest level of efficiency is the strong-form test. The weak form test focuses on past prices and previous returns, semi-strong form test concerns how quickly prices adjust to publicly available information, and the strong-form test expects that all available public and private information is accounted into prices (Fama, 1970).

When considering the three levels of efficiency of the EMH, it would be imprudent to claim that all levels of efficiency hold, if any. It is interesting how widely accepted a financial theory can be, although it often collides with reality. The strong-form of the EMH is practically impossible in practice, as all private information is not available to the market, the semi-strong form can also be questioned when prices can over- or underreact or the reactions may be delayed. The weak-form could hold in some cases but investors still indulge in biased behavior based on past prices which causes predictable return histories that collide with the theory.

2.3 Portfolio Theory

The portfolio theory by Markowitz (1952) was created for choosing an investment portfolio. The study states that the portfolio selection process consists of two stages of which first is experience and observation of the investment possibilities and the following expectations of future performance. The portfolio theory addresses the second stage that begins with the return expectations of possibilities and ends with choosing the investment portfolio. The study claims that the assumption of investors maximizing their discounted expected return is rejected as a hypothesis explaining investment behavior and as limitations guiding it. This is due to the rule not suggesting that there is a diversified investment portfolio which is preferred to any non-diversified investment portfolio. They suggest that there is support for an investment rule of aiming to maximize expected returns and minimizing the variance of the returns. Such portfolio is diversified as it is composed of multiple securities with the maximum expected return and minimum return variance and thus it is superior to any other investment portfolio (Markowitz, 1952).

2.4 Capital Asset Pricing Model

Sharpe (1964) claims that before the capital asset pricing model (CAPM) there has not been a financial model that accounts for risk, as the previous models have addressed investments under conditions of certainty. The study states that there is a positive relationship between risk and expected return which is depicted by the capital market line (CML). The capital asset prices are expected to adjust in equilibrium so that a rational investor following best available procedures, especially diversification can achieve any point on CML. A higher expected rate of return to investor could only be achieved by tolerating a higher level of risk. The market is comprised of two prices to investor, which are the price of risk and the price of time. They claim that previously no theory described how the price of risk is comprised from capital assets' physical attributes and the investor preferences. The study suggests that the model presented is consistent with multiple assumptions in traditional finance theories. Sharpe (1964) claims that the CAPM could be considered a model for determining capital asset prices as it notably explains the relationship between components of asset's total risk and the price.

Fama and French (2004) state that the asset pricing theory commenced from the CAPM of Sharpe (1964) and Lintner (1965). According to Fama and French (2004) the CAPM is a widely used model as in theory it generates a predictable relationship between the asset's risk and expected return. The CAPM can be derived as an equation:

$$E(r_i) = r_f + \beta_i[r_m - r_f], \tag{1}$$

where $E(r_i)$ is the expected return of an asset i, r_f is the risk-free rate, β_i is asset i beta, r_m is the expected market return, which means $[r_m - r_f]$ is risk-premium of the market.

Conversely to theory, Fama and French (2004) claim the CAPM empirical performance has been weak which questions the way it is applied in practice. The empirical shortcomings of CAPM may be caused by many simplifying assumptions of the model. In addition it is difficult to conduct robust tests for CAPM which could be other cause for the empirical problems. The unrealistic simplifying assumptions of CAPM such as the risk-free unrestricted borrowing and lending, and the unrestricted short selling probably are causing some of the empirical problems of the model (Fama & French, 2004).

2.5 Multifactor Risk Models

The arbitrage pricing theory (APT) by Ross (1976), can be considered the beginning for multifactor asset pricing models. The APT provides the possibility to use multiple risk factors when estimating an asset's expected return based on a linear relationship. This enables expanding the APT into multifactor models, for which the common expression is the following:

$$R_i = E(R_i) + \beta_{i1}F_1 + \beta_{i2}F_2 + \cdots + \beta_{ik}F_k + e_i, \tag{2}$$

where R_i is the excess return on asset i, $E(R_i)$ is the expected excess return on asset i, β_{i1} is the beta of risk factor F_1, β_{i2} is the beta of risk factor F_2, k is the number of risk-factors, and e_i is the company-specific surprise in the return of asset i. Roll and Ross (1980) state that the APT is suitable for multiperiod and single-period cases, unlike the CAPM which is limited to a single period. Another difference between the models is, the APT does not require the market portfolio to be mean variance efficient, but the CAPM does (Roll & Ross, 1980).

2.5.1 Three-Factor Model

The three-factor model by Fama and French (1992) is possibly the most well-known multifactor model to this date. The model has been widely adopted and used in finance to study the relationship between expected returns and factors affecting the risk of an asset. The univariate relationships seem to be strong between size i.e. market equity (ME) and average return, leverage, earnings to price (E/P) ratio, and book-to-market equity. The relationship between the company size and average returns has been negative in multivariate tests. Conversely, the relationship is positive between average return and book-to-market equity. The positive relationship also holds in comparison to other variables. They suggest that attention has been greater for the size effect but the explaining power for average returns has been higher with book-to-market equity. The average stock returns' cross-section cannot be explained with the beta, and during the sample period of 1963-1990, the average stock returns' leverage and E/P factors seem to be absorbed by the combination of book-to-market equity and size factors. They claim that stock risks seem to be multidimensional when assets are priced rationally. ME represents one dimension of stock risk, and other dimension is book value of equity (BE) to market value (BE/ME) of the company (Fama & French, 1992). Their three-factor model expression is the following:

$$R_i = E(R_i) + \beta_{i1}\left[E(r_m) - r_f\right] + \beta_{i2}(SMB) + \beta_{i3}(HML) + e_i, \qquad (3)$$

where in addition to the common multifactor model equation (2), β_{i1} is the beta for the excess market return $\left[E(r_m) - r_f\right]$, β_{i2} is the beta for the size-factor small minus big (SMB) portfolio, and β_{i3} is the beta for the value factor high minus low (HML) portfolio.

2.5.2 Four-Factor Model

Carhart (1997) created a four-factor model based on the three-factor model by Fama and French (1992) and added momentum as fourth factor based on the momentum anomaly for one-year period, observed by Jegadeesh and Titman (1993). The four-factor model has four risk factors in market equilibrium. Optionally, the four-factor model could be considered a performance attribution model where the factor-mimicking portfolios' coefficients and premia distribute the mean return proportions to four basic investment strategies. These four investment strategies are for stocks with high versus low beta, large versus small market capitalization, value versus growth, and return momentum versus contrarian for one-year period. The four factors of the model are return in excess of a value-weighted aggregate market proxy , portfolio returns of small minus big market capitalization companies, HML book to market value companies, and one-year momentum. The author finds that the average pricing errors are reduced by the four-factor model error of 0,14 percent, in comparison to errors of CAPM 0,35 percent, three-factor model 0,31 percent per month. Moreover, the pricing error patterns seem to be eliminated almost totally, which suggests that the cross-sectional variation is well described in average stock returns (Carhart, 1997).

2.5.3 Five-Factor Model

Fama and French (2015) created the five-factor model by adding two factors to the original three factor model by Fama and French (1992). The added factors are the difference between portfolio returns of companies with robust minus weak profitability, and the difference between portfolio returns of companies with low and high investments. The sample period for comparing the performance of the five-factor model to three-factor model and four-factor model is from July 1963 to December 2013, which means 606 months. The authors state that they detect patterns in average returns concerning the portfolios based on size, book to market value, profitability and investment. The five-factor model is estimated to explain the cross-section variance of expected returns by 71-94% for the size, book to market value, profitability and investment portfolios investigated (Fama & French, 2015).

2.5.4 Six-Factor Model

The six-factor asset pricing model of Fama and French (2018) is created by adding an up minus down momentum factor to the five-factor model by Fama and French (2015). The market factor in the six-factor model is the excess return of the market to risk-free rate. Other factors are the spreads of portfolio returns of small and big stocks, e.g. HML is the return difference between portfolios of HML book-to-market equity of small stocks, minus HML portfolios of big stocks. In a comparison of the nested models which are CAPM, three-factor model, five-factor model and the six-factor model, the latter is superior to others (Fama & French, 2018).

3 Non-Traditional Finance

Non-traditional finance is generally based on the thought that the positive relationship between risk and expected return in traditional finance may not hold in certain circumstances. This contradiction between traditional finance and non-traditional finance stems from unrealistic expectations of efficient market and asset pricing theories. In addition there is proof that investor behavior can be irrational and biased, which causes asset market anomalies. These findings in non-traditional finance suggest that the relationship between expected risk could be negative in certain situations, which could mean that there are arbitrage possibilities of gaining higher expected return without additional risk, or even risk-free profits.

This chapter will cover financial concepts and theories which have been used to explain the shortcomings of traditional finance theories and to offer alternatives for them. First, limits to arbitrage are covered, as they create the basis for non-traditional finance, as the asset price deviations from equilibrium could not persist without them. Next, momentum effect, and noise trading will serve as introduction to prospect theory, followed by behavioral finance and lastly behavioral biases. These will provide the basis for the following literature review chapter.

3.1 Limits to Arbitrage

Unlimited arbitrage is an important expectation for financial markets as theoretically it should drive asset prices deviating from their fundamental value, back to the expected fundamental value. In reality there are numerous limits to arbitrage. Shleifer and Vishny (1997) suggest that in reality there are limits to arbitrage which arise from differences between theories and reality. They claim that in reality it may be difficult or impossible to find similar assets in two different markets that differ in price, and generally there are differences in the trading hours, delivery terms, and settlement dates. Theoretically arbitrage should be riskless but in reality there are risks that that can arise if the mispricing that is arbitraged increases even more, which can cause liquidity issues and increased costs for arbitrageurs (Shleifer & Vishny, 1997).

Ljungqvist and Qian (2016) claim limits to arbitrage may also arise the shorting fees for lending a stock are high and only some of the stocks are available for shorting or the amount of capital available for arbitrageurs is not sufficient to cover possible margin calls. The possible margin call could force the arbitrageur to liquidate their position at a loss. This creates a significant risk, which is also against the theoretical assumption that arbitrage would be riskless. The study suggests that an informational arbitrage strategy has gained more popularity as it aims to decrease previously mentioned limits of arbitrage. Informational arbitrage is the reversal of conventional arbitrage strategy as the arbitrageur publicly announces their information that is credible to the public instead of searching for a suitable target company and shorting its stock without public announcement. This should encourage the current shareholders who are "long" to act and sell the stock. If the current shareholders act based on the provided information, this will drive prices towards their fundamental value, and accelerates price discovery which shortens the duration of the arbitrage cap and decreases the risk to arbitrageur (Ljungqvist & Qian, 2016).

Brav et al. (2010) test whether the limits of arbitrage are a credible explanation for asset-pricing anomalies. They sort securities with the Fama and French four factor model, based on their residual variability that represents idiosyncratic risk, to test the hypothesis that the existence of significant levels of mispricing are protected by high idiosyncratic volatility. The first tests reject the limits of arbitrage explanation for positive return anomalies for small value stocks and value stocks generally, positive earnings surprises and recent winners, as they do not find the relationship in accordance with the limits of arbitrage explanation. Conversely, when idiosyncratic volatility is low, the anomalies seem to be more significant. Although, the limits of arbitrage explanation is strongly supported for abnormal returns to small growth stocks, other growth stocks, negative earnings surprises, and recent losers. Their second tests reject the limits to arbitrage explanation for overvaluation and undervaluation anomalies. They test whether mispricing is amplified by noise trader momentum risk for growth and value companies, when previous returns have been good for growth firms or bad for value firms. Lastly, they test if a factor premium exists for portfolios of size, book-to-market, and momentum sorted stocks with lowest residual variability. Limits to arbitrage do not seem to explain the factor premiums for described portfolios, as the covariation between returns and the factors should be very low. The opposite seems to describe reality better, as the factors have high explanation power of returns to lowest residual variability stocks' zero cost portfolios (Brav et al., 2010).

3.2 Momentum Effect

The momentum effect is used to represent the phenomena where stocks previously experiencing positive returns tend to do so in the future, and correspondingly stocks with negative previous returns tend to follow this trend in future. These forecastable return trends defy the expectation of efficient stock markets. Investor herding can be suggested to be a factor behind momentum effect, as increasing number of investors may steer their interest towards stocks with positive past returns, thus creating a herd of investors buying the stock at prices over fundamental value.

Jegadeesh and Titman (1993) were arguably first to generalize knowledge of the momentum effect. They studied historical data on the New York stock exchange (NYSE) and the American stock exchange (AMEX) stocks, categorized on their past returns during 3-month period and examined their returns during the following 3-months, and similarly for 6-, -9 and 12-month periods. Predictable return patterns are detected from 3- to 12-month periods but the positive abnormal returns for stocks do not seem to last beyond the 12-month period, and instead negative abnormal returns are experienced following the period. For the portfolio formed based on the past 6-month returns, the average cumulative return is 9,5% for the following 12 months, but over half of this return is lost during the next 24 months. This pattern persists also for market expectations, as past winners realize higher returns consistently on their earnings announcements than past losers during the 7 months from portfolio formation. This effect is reversed for each of the following 13 months when returns around earnings announcements are higher for past losers than for past winners (Jegadeesh & Titman, 1993).

3.3 Noise Trading

Ramiah et al. (2015) define *noise trader* as an investor who does not use fundamentals as a basis for investment decisions, does not time the market well, usually overreacts or underreacts to new information and is prone to follow trends. Rationality wise, investors can be categorized to rational or information traders that base their investments on fundamentals, and irrational noise traders that do not consider fundamentals and trade on irrelevant signals. Such trend following and trading on irrelevant noise has been claimed to cause excessive volatility and a market component called *noise trader risk* (Ramiah et al., 2015).

Black (1986) claims that investors trading on information are rational they can expect to make profits with the trades. On the contrary, it is irrational to trade based noise in the way as if it was information. Noise traders should not expect to make profits from their trades. The study states that noise traders are still essential for market liquidity and thus a component affecting market efficiency. In their model noise is a component creating imperfections in the observations. Noise also creates uncertainty for example in the expected returns of assets or the relationship between monetary policy and inflation (Black, 1986).

De Long et al. (1990) state that noise traders have been recognized to be a significant group on the asset market but their effect on asset price formation has often been disregarded. They claim that noise traders may create asset price mispricing but there are also limits to arbitrage for the arbitrageurs to act based on the mispricing. Arbitrageurs can be assumed to be risk averse and their investment horizons to be relatively short. This limits their tolerance for taking positions that are contrary to noise traders' beliefs. The arbitrageurs' short horizons also create other significant risk as the mispricing created by noise traders may persist for a long time or even increase before reverting to mean eventually (De Long et al., 1990).

3.4 Survivorship Bias in Stock Index Returns

Survivorship bias in finance is often considered to affect mutual funds and their performance but also stock indexes are affected by it. The survivorship bias on stock indexes and their performance is of interest for this study. Stock indexes such as the S&P 500 are updated at certain intervals when some stocks are removed, and some are added to the index. Most common reasons for stocks to be either dropped from or added to the index are changes in market capitalizations, mergers and acquisitions (M&A) or significant initial public offerings (IPOs). These changes in index compositions favor the companies that have high market capitalization, are financially stable and growing, and correspondingly replace companies which have financial problems, are decreasing in size or go bankrupt. Index returns consist of the returns of stocks in the index, and by time, the winner stocks tend to stay in index and loser companies leave the index. Thus, surviving companies continue to contribute to the index return and dropped companies seize to do so, creating the survivorship bias.

According to Garcia and Gould (1993) the issue of backtesting the historical performance of a present portfolio arises from the need to have been able to forecast future in order to form the portfolio in the past. Not managing to quantify the impact of index constituent removals and additions could statistically bias their effect on index returns. Backtesting index returns creates multiple possible survivorship biases depending on the method and the selection criteria. The authors additionally found that including a stock in the S&P 500 index does not imply stronger performance in the future, which is logical as the additions are based on historical or present performance instead of unknown future performance (Garcia & Gould, 1993).

3.5 Prospect Theory

The prospect theory by Kahneman and Tversky (1979) can be considered as a beginning theory for behavioral finance. The study states that before the prospect theory, the expected utility theory (EUT) was generally accepted for analyzing decision making under risk. They suggest that in several different types of choice problems the participants' preferences have violated the EUT assumptions. The authors claim that the EUT cannot be suitably applied or interpreted as a model for choice under risk and instead the prospect theory is proposed for such situations (Kahneman & Tversky, 1979).

Kahneman and Tversky (1979) conducted multiple tests for prospect theory to study the participants' preferences when faced with choice under risk. If a person prefers a certain prospect (x) over any risky prospect (x), they are risk averse. The study states that choice problems in prospect theory systematically violate the EUT where the outcome utilities are weighted by the probabilities. They find support for phenomenon that they label certainty effect where certain gains are generally overweighted in relation to probable gains. When people in choice problems are faced with losses instead of gains, the preferences reverse and people tend to choose uncertain loss with lower expected value over a certain loss, which is called the reflection effect. Isolation effect refers to the phenomenon where people simplify choice between alternatives by focusing on differences between options and ignore the components that they share. This can lead to inconsistent preferences as the alternatives' common and distinctive components can be decomposed in multiple ways that could cause different preferences (Kahneman & Tversky, 1979).

3.6 Behavioral Finance

Lux (1995) states that the functioning of stock markets was explained with the EMH for a long time, but the credibility of these explanations has suffered, as increasing factual and empirical evidence has surfaced questioning the theory that does not allow stock prices systematically deviating from their fundamental values. Empirically one of the most significant facts questioning the efficiency of stock markets is the observation that the volatility for stock prices is higher than for the expected returns or fundamentals, which the stock prices should be based on.

Shiller (2003) states there has been significant progress in academic finance since the efficient markets theory was generally considered to have been proved without a question. According to the study, behavioral finance combines finance and a wider social sciences view consisting of psychology and sociology. They claim that nowadays behavioral finance is one of the most relevant research areas in academic literature and it provides significant critique to efficient markets theory. The efficient markets theory dominated the academic literature in the 1970s although the first evidence of anomalies also began to arise. The questioning of the efficient markets theory began to increase in the 1980s and the first signs of behavioral finance literature can be noticed. Pochea et al. (2017) state that confidence, greed or fear and other psychological factors are prominent in investor decision making and thus affect the overall market movements.

According to Statman (2018), in behavioral finance people are assumed to be normal as opposed to being fully rational. They suggest that in behavioral portfolio theory by Shefrin and Statman (2000), investors do not only build portfolios based on high expected return and low risk but also preferences such as social status and social responsibility. Statman (2018) states that investments’ expected returns are determined by behavioral asset pricing theory which in addition to differences in risk also accounts e.g. social status and preferences. According to the study, markets are not efficient in a way that the asset prices are same as the actual values of them but they are efficient in a way that it is difficult to make excess returns in them.

3.6.1 Behavioral Biases

According to Statman (2018) normal people that are not rational can commonly commit emotional or cognitive errors. This is due to people getting benefits that are utilitarian, expressive, and emotional when satisfying their wants considering all kind of products and services, also financial ones. Utilitarian benefit is the overall benefit to the individual or others, expressive benefits are about what something signals to others or you about yourself, and emotional benefits are about how something affects our feelings. The study states that humans have two cognitive systems, the quick intuitive system is implicit which is automatic and effortless, and the slow reflective system is explicit which is controlled and effortful. Implicit system often leads to sufficiently good choices with cognitive shortcuts in everyday life. But these shortcuts may also cause errors, especially in the more complicated matters, when the explicit system would generally lead to better choices. The decision errors caused by implicit system could also be one factor in creating behavioral biases in human behavior (Statman, 2018).

4 Literature Review

The literature review goes through previous literature about behavioral finance. Previous literature about behavioral biases is addressed. Herding behavior is being explained in terms of previous research. Stock market bubbles and crashes are explained, and the relevant previous studies are covered.

4.1 Herding Behavior

BenSaïda (2017) claims that if herding behavior occurs in a stock market, investors trading on a stock supposedly have deeper information about a stock, and opt out of investing to other stocks. These investors would be followed by other imitating investors without specific information, but they can consider trading on the stock rational based on the investment decisions of the well-informed trader. This will lead to higher than average trading volume for the specific stock, and lower for other stocks that are omitted from investment decisions for the stock that is considered superior to others (BenSaïda, 2017).

Park and Sabourian (2011) state that herding is a form of social learning where the agents' behavior switches from buying to selling or the contrary based on the crowd. Herding behavior may be rational when information externalities appear and the private information of the agents is dispersed by the information that is observed from others' actions. The characterization result finding is that in financial markets, social learning can only arise if the information received by investors has an intuitive and compelling nature. They find that herding and contrarianism both can lead to higher volatility and lower liquidity than in the case of no herding or contrarianism. There is still a significant difference between herding and contrarianism, as herding is self-enforcing when investors trade in the direction of the crowd. Conversely, contrarianism is self-defeating as a significant amount of contrarian trades will act against the crowd and end the contrarianism (Park and Sabourian, 2011).

Wanidwaranan and Padungsaksawasdi (2022) examine whether Google Search Volume Index (SVI) explains unintentional retail investor herding. Unintentional or spurious herding is based on investors having the same information and accidentally or unintentionally making similar investment decisions, which subsequently leads to efficient outcomes. This is considered information driven behavior that is consistent with rational herd behavior. The behavior of rational investors is similar when they receive and rationally analyze the same information about a stock or the market. Thus, stock prices will be driven closer to their fundamental values. If investors use information from internet searches as described to make trading decisions, this will create a relationship between herding and investor attention. However, intentional herding is not necessarily in-line with this, as it can cause market inefficiency, when investors disregard their own private information and instead decide to follow others decisions e.g. overall market consensus and peer recommendations. This follows the irrational herding concept, that is driven by psychological factors. The study finds that the Google SVI has a statistically significant positive impact on herding, which means that the herd behavior is stronger, when more online searches are made. This increased unintentional herding leads to more informationally efficient stock prices (Wanidwaranan & Padungsaksawasdi, 2022). The authors state that during downward market and crisis periods, investor attention seems to decrease, where the ostrich effect is offered as an explanation as suggested by Karlsson et al. (2009). Investors being more attentive during upward market and less attentive during a downward market is claimed to be caused by cognitive bias (Wanidwaranan & Padungsaksawasdi, 2022).

Li et al. (2022) study the effect of mobile trading on herding in China. They find that mobile trading seems to increase investor herding. The authors suggest that the mobile devices enable individual investors to respond to stock market events more quickly and based on impulses and emotions. These actions are more likely irrational and not well considered. With the increasing amount of mobile traders, this kind of impulse trading could also increase irrational herding behavior. Increasing mobile trading seems to decrease stock market dispersion, which suggests an increase in herding. Moreover, it is found that stocks with higher ratio of mobile trading seem to have a higher amount of investor herding (Li et al., 2022).

According to Chiang and Zheng (2010) there has been numerous researches on herding in different aggregate markets, they have mostly focused on certain financial market area. Thus, herding has not been studied on an international level to study the differences and possible relations between different markets. They state that this may create econometric limitations for the studies, as important variables could be excluded from ordinary least squares (OLS) estimates, which may bias the results. Other weakness of the researches stems from studying the herding behavior of a limited area, as the results cannot be adapted globally for the international behavior. Their study investigates herding behavior on an international level by testing the relationship between cross-sectional stock return dispersions and explanatory variables domestic market absolute stock returns, market conditions, and influences of foreign market. Herding behavior is examined in 18 economic units sorted into advanced, Asian, and Latin American markets. Investors in national markets are noticed to mostly herd following the US market. The findings suggest herding behavior in advanced and Asian markets, but less clear evidence for Latin American markets. Herding behavior seems to be more pronounced in financial crisis periods. Findings in Latin America support this as there is no support for herding except during the 1994 crisis in Mexico and 1999 crisis in Argentina (Chiang and Zheng, 2010).

4.2 Empirical Measures of Herding Behavior

This chapter will cover traditionally used empirical measures for market-wide herding. A more robust herding measure QR will also be covered. Of the methods introduced in this chapter, the application of cross-sectional absolute deviation (CSAD) method and the QR are described in the methodology chapter.

The widely used empirical measures of herding are somewhat problematic, as they are based on the notions and expectations of the rational asset pricing models, although herding is often considered irrational behavior and in conflict with the standard finance expectations. Main functioning logic behind herding measures is based on the expectations of rational asset pricing models how stock return dispersions should behave, and the violation of the expected relationship is seen as support for herding. Thus, one could argue whether stock return dispersions are a suitable measure for herding to begin with. If the stock return dispersions do not actually represent herding, regardless of the results, there could be herding even if the results suggest that there is not, and vice versa. This evokes the apparent need for new herding measures, that better depict the mechanisms behind herding and measure it more accurately.

4.2.1 Cross-Sectional Standard Deviation

Christie and Huang (1995) are considered the first to use the stock return dispersions as a measure for herding in the stock markets, with the cross-sectional standard deviation (CSSD) method. Stock return dispersions represent the average difference of individual returns from the mean return. The dispersions' lower bound is at zero, when all returns would move perfectly in line with the market, and accordingly, increasing dispersions signal individual stock returns deviating further from the market return. Their expectation of return dispersions increasing in case of no-herding during turbulent market periods, when fluctuation in market returns is higher, stems from the expectation of rational asset pricing models. Conversely, if the return dispersions decrease during periods of higher market price variation, this would contradict the expectations of rational asset pricing theories and suggest herding (Christie and Huang, 1995). Their formula used to calculate CSSD can be expressed as follows:

$$CSSD_t = \sqrt{\frac{\sum_{i=1}^{N}(R_{i,t}-R_{m,t})^2}{N-1}}, \tag{4}$$

where N is the number of firms, $R_{i,t}$ is the return on stock i on day t, $R_{m,t}$ is the cross-sectional average of the N returns in the index on day t. As Christie and Huang (1995) emphasize, CSSD itself is not a measure of herding but return dispersions are still expected to be low when herding exists. Instead, herding during turbulent market periods is measured with the following regression model:

$$CSSD_t = \alpha + \beta_1 D_t^L + \beta_2 D_t^U + \varepsilon_t, \qquad (5)$$

where D_t^L is a dummy variable which has value 1 when the market return falls in the extreme lower tail of the return distribution and 0 otherwise, D_t^U is a dummy variable that correspondingly gets value 1 when the market return reaches the extreme upper tail of the return distribution and 0 otherwise, α is the coefficient representing the average dispersion of the sample period not covered by either of the dummy variables. The coefficients β_1 and β_2 represent herding in different market states, and statistically significant positive values would be in line with rational asset pricing theories, and conversely statistically significant negative values would imply the existence of herding (Christie & Huang, 1995).

Christie and Huang (1995) find that dispersions increase significantly during turbulent market periods. This increase in the stock return dispersions is stronger in rising than in falling markets. They tested whether the asymmetry would suggest herding in decreasing markets, by using a rational asset pricing model to estimate the predicted returns' dispersions. Actual and the predicted return dispersions are practically similar, which supports their hypothesis that the lower increase in dispersion during falling stock markets supports rational pricing instead of herd behavior (Christie & Huang, 1995).

4.2.2 Cross-Sectional Absolute Deviation

Chang et al. (2000) create the CSAD method by extending the CSSD method by Christie and Huang (1995). They define the stock return dispersions as the average absolute deviation between stock returns and the market return. In addition to traditional asset pricing models' expectation of positive relationship between market return fluctuation and return dispersion, it is emphasized that the relationship should be linear. Thus, non-linear or negative relationship between market return and return dispersion indicates herding. In the following herding regression by Chang et al. (2000) the possible non-linear relationship is expected to be captured by the squared market return $R_{m,t}^2$, and a negative γ_2 coefficient would suggest the presence of herding:

$$CSAD_t = \alpha + \gamma_1 R_{m,t} + \gamma_2 R_{m,t}^2 + \varepsilon_t \qquad (6)$$

In the study by Chang et al. (2000), investor herding is examined in the US, Hong Kong, Japan, South Korea, and Taiwan. Return dispersions are expected to decrease when herding is strong, and to increase at a decreasing rate, when herding is moderate. Their findings for the US are uniform with the findings of Christie and Huang (1995) as the herding coefficients are significantly positive and the equity return dispersions generally increase in turbulent market periods. These findings do not suggest herding, which should decrease the return dispersions, by definition. The results are similar in Hong Kong and Japan and mostly similar for South Korea. Taiwan has contradicting results for herding, as in two of the three models, the coefficient measuring changes in investor behavior is significantly negative for periods with considerable upward price movements. This supports the occurrence of herding in Taiwan (Chang et al., 2000).

Since the non-linear herding regression of Chang et al. (2000) the model has been modified by Chiang and Zheng (2010) by adding an absolute market return factor $|R_{m,t}|$ which accounts for asymmetric investor behavior during different market conditions. Consistently to equation (5), negative herding coefficient γ_3 implies herding in the following equation:

$$CSAD_t = \gamma_0 + \gamma_1 R_{m,t} + \gamma_2 \left| R_{m,t} \right| + \gamma_3 R_{m,t}^2 + \varepsilon_t \quad (7)$$

4.2.3 Quantile Regression Method

Koenker and Bassett (1978) created the QR method as an extension of the OLS method, as the latter is very sensitive to outlier contamination in modest amounts. This makes OLS method weak estimator in many non-Gaussian events, especially for distributions with long tails. They created the QR model from the linear model by defining the sample quantiles, which enables to avoid the common reliance on a group of ordered sample observations. The model has slightly lower efficiency in terms of the Gaussian distribution, when compared to OLS model but the QR method seems to have a significant advantage over the OLS model in a situation with a non-Gaussian sample distribution (Koenker & Bassett, 1978).

According to Duygun et al. (2021) QR is used to estimate a collection of conditional quantile equations. The general QR equation can be expressed as follows:

$$y_i = \alpha_\tau + \beta_\tau x_i' + \varepsilon_{\tau,i}, \quad (8)$$

where y_i is the dependent variable, α_τ is the constant coefficient, β_τ is the vector of estimated coefficients for the vector of predictors x_i', and ε_τ is the error term.

Bekiros et al. (2017) state that the difference between standard OLS method and QR method, is the parameter identification in OLS when squares of deviations are minimized from the sample's conditional mean. Whereas, the estimators are defined for linear QR by designating weights based on the quantile, and minimizing a weighted sum of absolute errors based on the weights (Bekiros et al., 2017).

Pochea et al. (2017) use the QR method to study herding in Central and East European stock markets. They claim that the QR method is well suited for herding research as it explains the relationship between the explanatory variables and the dispersion measure. The estimators of QR are obtained when the weighted sum of absolute errors are minimized for specific quantiles' values, since under the presence of extreme values it is appropriate estimation method. Based on previous studies by Christie and Huang (1995) and Chang et al. (2000), the herding behavior is expected to be more pronounced in a declining and turbulent time than in increasing stock market. The expected asymmetry in herd behavior will be researched using the QR method as it is capable to explain these asymmetries between increasing and decreasing market (Pochea et al., 2017).

Bekiros et al. (2017) investigate herding with QR method on S&P100 and DJIA indexes. Additionally, they provide findings by QR method on the time-variation of herding i.e. temporal changes. Market sentiment variable is introduced to observe possible effects it has on herding behavior. The study period is from January 2000 to July 2015. Herd behavior variations are expected over time, and different sub-periods of two and seven years are studied consequently. Standard OLS model is used as a benchmark to detect possible differences in herding results in comparison to QR model. The QR model results suggest that herding decreases over time, which implies that the irrational following of other investors behavior is becoming less significant. However, herding seems to be more significant in a turbulent market and following the global financial crisis (GFC). Splitting the sample into sub-periods provided contradicting findings with the previous studies, as no significant herding was detected during the GFC. The authors claim that such difference in empirical results could be due to using QR method instead of the OLS model often used in the literature. Additionally, a significant positive correlation between investor herding and market volatility was observed. This provides support that market sentiments e.g. fear increase investor herding (Bekiros et al., 2017).

Duygun et al. (2021) investigate herding in the US and Eurozone stock markets using QR and the generally used CSAD methods to discover possible differences between the results. They measure herding on the market level and separately on the financial industry level. Evidence of herding during the GFC is found with both study methods for US and Eurozone stock markets and financial industries, except for Eurozone banks. Additionally, a herding spillover effect is detected in both economic areas, in Eurozone, spreading from banks to domestic market, and from insurance sector to the domestic market in the US (Duygun et al., 2021).

4.2.4 Herding During Extreme Market States

According to Duygun et al. (2021), during the GFC, herding coefficients in OLS estimates are significantly negative for equity markets and diversified financials in the Eurozone and the US, and also in the real estate industry for the latter. Herding estimates of the QR show that the return dispersions decreased significantly which suggests increase in herding during turbulent market period across all financial sectors. The results differ for the Eurozone crisis (EZC), as there is no evidence of herding in both markets. The OLS and QR methods provide a positive herding coefficient, which implies the absence of herding. When analyzing the financial industry, the herding coefficient is significantly negative in middle quantiles for banks, for Eurozone and the US. The OLS model suggests herding in the US insurance industry, until the 95 *th* quantile, whereas in the Eurozone real estate industry, herding arises in lower quantiles during this period (Duygun et al., 2021).

4.2.5 Herding in Rising and Falling Stock Markets

Possible asymmetric herding effect between up- and down-market states has recently been a topic of interest for researchers. Bekiros et al. (2017) studied herding in stock markets with QR and the OLS methods and analyzed effect of bull and bear markets on investor herding, and found it to be more prominent in declining than in rising markets and more significant in top quantiles above 90%. This suggests that turbulent market periods intensify herding, as the higher quantiles' dispersion accounts for larger price changes. Thus, herding could have been one catalyst of the GFC. They state that herding asymmetries were more noticeable at the monthly level than daily.

4.2.6 Different Forms of Herding

Venezia et al. (2011) claim that herding is often considered to be based on information or the lack of it. The information availability and quality varies by different investor groups. This informational asymmetry between different investors may cause varying herding behavior (Venezia et al., 2011).

Nofsinger and Sias (1999) claim that herding is sometimes divided based on investor group, into individual or institutional herding. Individual herding is considered irrational, but the reactions to trends or sentiment are systematic. Institutional herding is often considered more rational, and it could be caused by security features, agency problems, trends or how information spreads in the market. They investigate the differences in herding between individual and institutional investors, and the post-herding returns. Annual changes in institutional ownership and returns seem to have a significant positive relationship, as on average the stock decile increasing most in institutional ownership has abnormal returns of 18,38 percent, and the stock decile decreasing most in institutional ownership has negative abnormal returns of -13,12 percent, which means that the return difference between the deciles is over 31 percent. This implies that institutional investors either do intrayear positive-feedback trading more than individual investors or the institutional herding has more significant impact on asset prices than individual herding. Although, post-herding return examinations do not suggest irrationality in institutional herding, as no return reversals are detected in the following two years after herding. The subsequent returns for stocks bought by institutions are higher than for stocks sold by them. In addition, the findings suggest that institutional investors utilize positive-feedback trading.

Bikhchandani and Sharma (2000) divide herding into “spurious” and “intentional” herding. Herding can be defined as a recognized intent for investor to copy the decisions of other investors, which is called “intentional” herding. When investors are in similar decision situations and have symmetrical information, and unknowingly end up making identical decisions, the herding is “spurious”. “Spurious” herding can be considered an efficient outcome, the opposite holds for “intentional” herding that is considered inefficient. Theoretically straightforward distinction between “intentional” and “spurious” herding, is empirically complicated, as many different factors or combinations of them may affect trading decisions (Bikhchandani & Sharma, 2000).

Duygun et al. (2021) investigated “intentional” and “spurious” herding using QR approach during the EZC and in the US found support for “intentional” herding i.e. non-fundamental for equity markets and whole financial sector, most notably in lower quantiles. The authors suggest that there may be a distinction between type of herding in different crises, as during the GFC, investor herding was mostly “spurious”, unlike in the EZC.

4.2.7 Industry-Specific Herding

Ukpong et al. (2021) examine industry herding in the U.S with the CSAD methodology of Chang et al. (2000). The extent of stock market dispersion is estimated for the model for every industry and the aggregate market. Market conditions such as market returns, trading volume and volatility are expected to affect the industry herding. Thus, it is examined, if herding is influenced by high and low market returns, volume and differences in volatility. Their research finds no evidence of market-wide herding, but instead there is evidence of anti-herding as the CSAD is significantly positive. On a sector level, there is confined evidence of herding, as the herding coefficients are negative and significant for industrials, financials and real estate. Still, the coefficients for the remaining 7 sectors are positive and statistically significant, which suggests anti-herding. The results are similar for market-wide herding in rising and falling markets, as there is no support for herding but anti-herding is detected. Accordingly, there is evidence for herding in increasing and decreasing markets on industry-level for industrials, financials, real estate, telecoms and utilities. This suggests that, regardless of the market state, investors herd in the same industries (Ukpong et al., 2021).

Litimi et al. (2016) investigate the occurrence of herding in the US market, and whether possible herding causes excessive market volatility. Understanding the relationship between herding and volatility would be of significant value for risk management purposes, such as in derivatives trading the spot market herding could be used as a signal for investment decisions. Their sample includes four major crisis periods, which are the black Monday in 1987, the dot-com bubble from 1997 to 2000, the 2002 stock market downturn, and the GFC beginning in 2008. They claim that if there is no herding which suggests that investors are rational and estimate the price in terms of the CAPM, the relationship between CSAD and $R_{m,t}$ should be linear and positively increasing. If investors herd, the stock returns should converge towards the market return trend, and the relationship between CSAD and $R_{m,t}$ will be negative and nonlinear. The study extends the CSAD model by introducing two additional variables; a dummy variable for signaling significant bubbles and crashes, and a sentiment index to represent the fear or greed of the market. The authors observe herding behavior during extreme market states, these are financial market crises with negative returns, and market bubbles with increasing returns. The original CSAD model regression results suggest that the linear relation between the variables is violated, as the market returns are in absolute values. Based on the nonlinear coefficient γ_2 for $R_{m,t}^2$ there is a heterogenous market structure in the US, and a significantly negative herding coefficient exists only in two sectors: Public utilities and Transportation. The modified CSAD model seems to provide more accurate representation of herding, as it suggests herding for the whole market, resulting from herding in 8 out of 12 sectors (Litimi et al., 2016).

BenSaïda (2017) studies how herding affects idiosyncratic volatility on a sector level, instead of the entire US stock market. The research also addresses relationship of herding and trading volume and the market sentiment, and deepens the understanding of the cross-relationships effect on idiosyncratic volatility in normal and crisis periods. They modify the CSAD model by adding three variables to the original model. Two of the introduced variables are dummy variable representing major bubble and crisis periods, and a sentiment index to express the degree of investor fear, which are in line with the CSAD modifications of Litimi et al. (2016), and the third volume turnover variable, as a new addition. The original CSAD model results of BenSaïda (2017) are mostly in line with the results of Litimi et al. (2016), except that there is no evidence of herding in any industry, whereas the latterly mentioned detected herding for two sectors. The modified CSAD model results suggest herding as the coefficients are significantly negative for 10 sectors out of 12, and negative for 2, although not significantly. The difference in herding between the original and the modified CSAD model imply that during bubbles and financial crises, investors bypass their own information and attempt to imitate the decisions of presumably better informed investors, consequently beginning to herd (BenSaïda, 2017).

4.3 Stock Market Bubbles and Crashes

Bubbles and crashes are terms used for asset market phenomena, where asset values increase or decrease excessively in comparison to actual fundamental values. These asset price increases and decreases, not fully supported by changes in the actual values often happen in a relatively short period of time. Vogel and Werner (2015) claim that over the last forty years a significant amount of literature about bubbles and crashes has been published. The literature regarding bubbles and crashes has mostly focused on describing and testing for them mathematically but the assumptions are based on rationality and market efficiency. Extreme market events such as bubbles and crashes have not been explained sufficiently by the existing models and the criticism and questions have not been answered or resolved adequately (Vogel & Werner, 2015).

Barro and Ursúa (2017) claim that stock-market crashes often precede economical depressions. They state that stock-market crashes, defined as multi-year cumulated real returns of -25 percent or lower, are followed by minor economic depressions 31 percent of time, defined as multi-year gross domestic product (GDP) or consumption decline of 10 percent or more, and by major depressions of 25 percent declines or more, 10 percent of the cases. Conversely, minor depressions include a stock-market crash in 71 percent of the cases, and for major depressions the probability of including a stock-market crash is 92 percent. This suggests that minor and major economical depressions are highly conditional on a stock-market crash happening (Barro & Ursúa, 2017).

Whitehouse et al. (2023) state that in financial markets the asset price bubbles and crashes are a prevalent phenomenon. Examples of significant historical bubbles are Dot-com bubble that affected technology stocks in late 1990s, the mid-2000s sub-prime mortgage bubble in the US housing market, and the recent bubbles in the cryptocurrency markets. The possibility to recognize the emergence and collapse of asset price bubbles would be of great importance as the effect often is not limited to the market where the bubble occurs and is widespread across economies globally. Methods to early detection and warnings of bubbles and crashes would be needed, and thus the authors provide a procedure to monitor asset price bubble crashes in real-time. The previous literature in identification of bubbles has mostly focused on historical detection of bubble episode from a sample of observed data after the emergence of the bubble. Rational bubbles have mostly been the focus of these studies, which implies a scenario where the value of an asset exceeds the fundamental value but the investors purchase the asset as they expect that other investor will pay more for the asset later (Whitehouse et al., 2023).

Astill et al. (2017) study the possibility of creating a method for one-trial test of detecting an end-of-sample bubble. Astill et al. (2018) continue investigating the topic with a method which is not limited to one-shot testing and it allows the theoretical false positive rate (FPR) to be determined at any point during the bubble monitoring period. Astill et al. (2018) find that the true-positive rate (TPR) is appealing for the alternative hypothesis when a bubble arises during the monitoring period. Their objective is to recognize an explosive regime rapidly, by detecting the possible deviation from the null hypothesis. Based on simulations, the procedures allow for real-time detection of an emerging bubble. The methods empirically applied to five indices of major stock markets, indicated presence of bubbles for all indices during the test period from January 1997 to January 2002 (Astill et al., 2018).

Phillips et al. (2015) create a method for studying bubbles and crashes in stock markets. They apply the method to S&P 500 stock market data from January 1871 to December 2010. The new research application was able to successfully identify all the generally known historical bubble and crash events during the sample period, e.g. the great crash, Black Monday in October 1987, and the dot-com bubble. However, Astill et al. (2018) claim that when the test of Phillips et al. (2015) is sequentially applied, it would not be size controlled, as the overall false-positive rate (FPR) will be unknown, defined by the probability, that at least one test in the sequence rejects the null hypothesis, when it was actually true. Such test in sequence would falsely suggest that there was a bubble in the monitoring period, when there was not. Performing these tests sequentially as described would cause the FPR to probably be significantly over the nominal rate, which is used for the individual tests. If the monitoring horizon grows, the FPR would increase even more, due to the multiple testing problem (Astill et al., 2018).

4.3.1 Testing for Bubbles and Crashes

Diba and Grossman (1988) were knowingly the first to test for explosive financial bubbles with method based on autoregressive unit root tests by Dickey and Fuller (1981). They investigate the possibility for rational bubbles and aim to measure such bubbles and crashes. The existence of rational bubbles is based on a consideration that fundamentally irrelevant variables that are not observable on the market, affect asset prices. Their model assumes that the intrinsic value consists of an unobservable variable and the present value of expected dividends, which are discounted at a constant rate. Rational bubbles are then defined as a self-enforcing deviation of stock prices from fundamentals, based on external variables. Based on the tests, rational bubbles are not detected in the stock prices (Diba & Grossman, 1988).

The framework by Phillips et al. (2015) can be used to test for bubble phenomena and date them also with multiple possible bubbles in the sample. Their methods are extended from method by Phillips et al. (2011) by enabling flexible window widths in the recursive regressions that the methods of the test are based on. Phillips et al. (2011) test for bubble by using the sup augmented Dickey-Fuller (SADF) method which uses a sequence of forward recursive right-tailed augmented Dickey-Fuller (ADF) unit roots tests. They proposed a dating strategy based on a backward regression technique to identify bubble origination and termination points. The previous test worked satisfactorily in terms of structural breaks against other recursive procedures such as the Chow tests, cumulative sum tests and model selection, and was particularly effective in real-time bubble detection, when Homm and Breitung (2012) conducted extensive simulations. The test has limitations, when the sample period has multiple episodes of bubbles and crashes, as it reduces the effectiveness of the procedures and it may become inconsistent and fail to detect bubbles. This weakness is especially significant when analyzing long time series or when the market data has a lot of turbulence and multiple episodes of exuberance are suspected. To address this limitation and manage multiple breaks of exuberance and collapse, a Generalized Sup Augmented Dickey-Fuller (GSADF) method will be used to test for presence of bubbles, and to time-stamp the bubble emergence and termination dates with a recursive backward regression technique. Similarly to Phillips et al. (2011), the new framework is based on recursive right-tailed ADF tests but in implementation, flexible window widths will be used (Phillips et al., 2015).

Test method for bubbles by Astill et al. (2018) will be used to detect possible bubbles and crashes in the time series data of the analyzed markets. Their method is based on the monitoring procedure developed by Harvey et al. (2018) for predictive regime detection tests. Harvey et al. (2018) consider a scenario where the null hypothesis of no predictive behavior is rejected if the separate t-tests signal contiguous rejections by exceeding a certain threshold value when performed at assigned nominal significance level. The individual t-test critical values are achieved by using the sub-sampling method by Andrews (2003) and Andrews and Kim (2006) where training period observations for which the null of no predictive behavior holds are assumed to be available in practice. The method of Astill et al. (2018) detects an explosive asset price bubble in the monitoring period if the test by Astill et al. (2017) signals that the number of contiguous rejections on a certain significance level exceeds a threshold value. Their application also modifies the methodology of Harvey et al. (2018), where bubble is identified if any monitoring period test statistic exceeds the highest analogous statistic of sub-sample on training period, and thus making it unnecessary to calculate critical value for training period.

5 Methodology

The construction of the dynamic index in this thesis seems to be fairly distinctive in comparison to the conventional index construction for similar studies. A common practice for index construction seems to create a fixed index composition of stocks which will be analyzed for the sample, as in the fixed composition of this study. The dynamic index composition has been created by separately selecting each year's quarterly updated S&P 500 compositions which have daily price data for the constituents. These separate yearly compositions are then combined from year beginning of 1989 until the beginning of 2025 to form a continuous sample that follows the historical changes in the S&P 500 index.

The CSSD method by Christie and Huang (1995) has knowingly been the first attempt to measure investor herding quantitatively. Since then, multiple different factors have been added, and different methods have been created to measure herding more accurately. Nevertheless, the idea of how herding can be measured has remained mostly similar. Well known extension to the method has been the CSAD method by Chang et al. (2000). The CSADs are defined according to them as follows:

$$CSAD_t = \frac{1}{N}\sum_{i=1}^{N} |R_{i,t} - R_{m,t}|, \tag{9}$$

where $R_{i,t}$ is the return of the stock i at time t, $R_{m,t}$ is the return of the market at day t, N is the number of stocks in total. After the calculation of daily CSADs, the CSAD regression by Chiang & Zheng (2010) is defined with the equation:

$$CSAD_t = \gamma_0 + \gamma_1 R_{m,t} + \gamma_2 |R_{m,t}| + \gamma_3 R_{m,t}^2 + \varepsilon_t, \tag{10}$$

where γ_0 is the constant coefficient, γ_1 is the coefficient for market return on day t, γ_2 is the coefficient for absolute market return on day t, $|R_{m,t}|$ is the absolute market return on day t, γ_3 is the coefficient for squared market return on day t, $R_{m,t}^2$ is the squared market return on day t, and ε_t is the error term.

To study the occurrence of herding in stock markets, the QR method of Koenker and Bassett (1978) will be used. The CSAD formula by Chang et al. (2000) will be used for the QR, applied by Bekiros et al. (2017) using the equation seen below:

$$Q_\tau(\tau|CSAD_t) = \gamma_{0,\tau} + \gamma_{1,\tau}|R_{m,t}| + \gamma_{2,\tau}R_{m,t}^2 + \varepsilon_{t,\tau}, \tag{11}$$

where τ is the quantile (0,1), $\gamma_{0,\tau}$ is the intercept representing the baseline level of dispersion when market return is zero, $\gamma_{1,\tau}$ is the coefficient for absolute market return on day t, $\gamma_{2,\tau}$ is the coefficient for squared market return on day t, and $\varepsilon_{t,\tau}$ is the error term.

In a QR, all regressions at different quantiles are applied to all observations, instead of pre-assigned observations. The observations are ordered to different quantiles based on the dependent variable, in this study the daily CSAD values. The estimations are done on the whole dataset but with different weightings applied to each quantile, and the weighted sum of absolute residuals is being minimized. The weightings are done in such a way that e.g. quantile $\tau = 10\%$ has the weighting $1 - \tau$ below the estimated level, equaling 90%, and accordingly weight τ above the estimated level, equaling 10%. The same logic applies to quantiles above $\tau = 50\%$, e.g. the weighting for quantile $\tau = 90\%$ is 10% below the specified level, and 90% above it. In this study, the pseudo-R^2 is applied to QR to measure goodness of fit, as it is better suited for non-linearity and other properties of QR, as reasoned by Koenker and Machado (1999).

The relationship between herding and bubbles is examined by incorporating a bubble dummy variable into the QR equation. The QR equation with bubble dummy will be used to investigate whether the herding results differ between non-bubble and bubble periods by using the following regression:

$$Q_\tau(\tau|CSAD_t) = \gamma_{0,\tau} + \gamma_{1,\tau}|R_{m,t}| + \gamma_{2,\tau}(1-D)R_{m,t}^2 + \gamma_{3,\tau}DR_{m,t}^2 + \varepsilon_{t,\tau}, \quad (12)$$

where in addition to equation (11), D is the bubble dummy which gets the value 1 during bubbles, and 0 otherwise. The bubble dummy has been constructed by measuring rolling cumulative returns of the S&P 500 index. The used rolling window is 90 trading days, and the return threshold is 20% cumulative return during the rolling window. Thus, if the cumulative rolling return is under 20%, $\gamma_{2,\tau}$ coefficient represents herding during such non-bubble periods, and when the cumulative rolling return is over 20%, $\gamma_{3,\tau}$ represents herding during such bubble periods.

The used cumulative rolling return method is a simplified measure, but it still serves its purpose to distinguish between periods of moderate, and increased return periods. Such high return periods could be a sign of excessive valuations from fundamentals, and thus a signal of a bubble. The rolling window length of 90 trading days, equaling about 4,5 months, is relatively short, but this increases its sensitivity to short-term bubbles. The return threshold of 20% has been chosen, as it could be considered an excessive short-term return of 90 trading days.

6 Data and Descriptive Statistics

The following chapters will describe the data used for empirical analysis and the basis for the decisions regarding the data. Descriptive statistics for the relevant measures of the data will be provided following the data description.

6.1 Data

The data for empirical analysis has been downloaded from Datastream database. It consists of daily closing price data for market capitalization weighted S&P 500 index and for the index constituent stocks. The price data is adjusted for stock splits, and all prices are in local currency, US dollars ($). The sample period is from 03.01.1989 until 13.01.2025, equaling slightly over 36 years with 9076 daily observations in total. The S&P 500 index composition is generally updated quarterly, and thus there have been numerous changes during the observation period. Two separate datasets are used; the main data set is a dynamic S&P 500 index composition that follows the historical index constituent changes during the observation period. This means that constituent stocks have changed according to the historical changes in S&P 500 index composition, varying at approximately 500 stocks for whole sample period. The other data set used for comparison, is a fixed composition of S&P 500 constituent stocks that have been in the index at the beginning and at the end of the observation period which means that some stocks have remained in the index continuously, and some stocks have left the index at some point but have re-entered it again before the end of study period. The fixed composition equals 236 stocks, for which the daily data is gathered for the whole sample period. The comparisons of descriptive statistics and results between the two index compositions are used to detect possible differences in the datasets, which may be affected by survivorship bias among other factors.

It has been determined that the main analysis is done on the dynamic index composition. This should enable us to follow the actual changes in index as accurately as possible. This is due to focusing the study on stocks that have remained in the index for a long period of time and those that have left it for a period before being added to the index again. This allows us to study how herding has affected a group of stocks with varying index histories to see if herding has emerged, and if there is a relationship between possible herding and bubbles and crashes. The data has been cleaned by removing days with only zero returns for the index and all the stocks.

In addition to acquiring daily price data for S&P 500 index and the stocks, the daily returns need to be calculated for the index and the stocks. The daily returns are calculated for both, by using the following formula:

$$R_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \tag{13}$$

where R_t is the daily return between day $t-1$ and day t, $\ln$ is a natural logarithm, P_t is the closing price of an asset on day t, and P_{t-1} is the closing price of an asset one day before day t.

6.2 Descriptive Statistics

As mentioned in the research hypotheses chapter, the null hypothesis states that the S&P 500 index returns and the stock return dispersions measured by CSAD, are normally distributed. The alternative hypothesis is the opposite as it states that the index returns and return dispersions are not normally distributed. The null hypothesis and the alternative hypothesis can be tested by providing descriptive statistics of the underlying data for the S&P 500 return and the stock return dispersions and examining the measures of variability. Thus, based on skewness and kurtosis it can be detected if the stock return dispersions and index returns do not follow a normal distribution, which would mean that the null hypothesis will be rejected, and the alternative hypothesis is accepted. The descriptive statistics for stock return dispersions of fixed and dynamic index constituents, and S&P 500 market returns can be seen in the following table 1.

Table 1. Descriptive statistics for CSADs and market returns

Index	**S&P 500 fixed constituents**	**S&P 500 dynamic constituents**	**S&P 500**
	CSAD	CSAD	Rm
Mean	0,0118	0,0126	0,0003
25th Percentile	0,0085	0,0093	-0,0044
Median	0,0110	0,0114	0,0006
75th Percentile	0,0136	0,0140	0,0057
Maximum	0,0578	0,0641	0,1096
Minimum	0,0033	0,0039	-0,1277
Standard Deviation	0,0049	0,0052	0,0113
Skewness	2,1771	2,3834	-0,4169
Kurtosis	9,1824	9,8345	10,7721
Observations	9076	9076	9076

This table reports the descriptive statistics of daily cross-sectional absolute deviations (CSAD) for S&P500 with fixed and dynamic constituents, and S&P 500 index returns (Rm).
The sample period is 03.01.1989-13.01.2025.

When examining the descriptive statistics in terms of null hypothesis, the skewness for normal distribution should equal zero but it is over two for (*CSAD*) of both index compositions, and slightly negative for (*Rm*) which means that they are asymmetrical. Kurtosis of a normal distribution should be three, when in the data it is over nine for (*CSAD*) of both index compositions, and over ten for (*Rm*) which means that the distributions are tail-heavy. These findings indicate that the stocks return dispersions, and the market returns do not follow a normal distribution, and therefore the null hypothesis H_0^1 is rejected, which denotes that herding could exist. Thus, the alternative hypothesis H_a^1 which states that the stock return dispersions and the market returns do not follow a normal distribution, will be accepted.

In a similar manner as in the study by Chang et al. (2000), scatter plots are used to represent the relationship between market returns and the stock return dispersions. Figure 1 below shows the relationship between *CSAD* and *Rm* for the fixed index constituents, and figure 2 depicts the relationship for dynamic constituents.

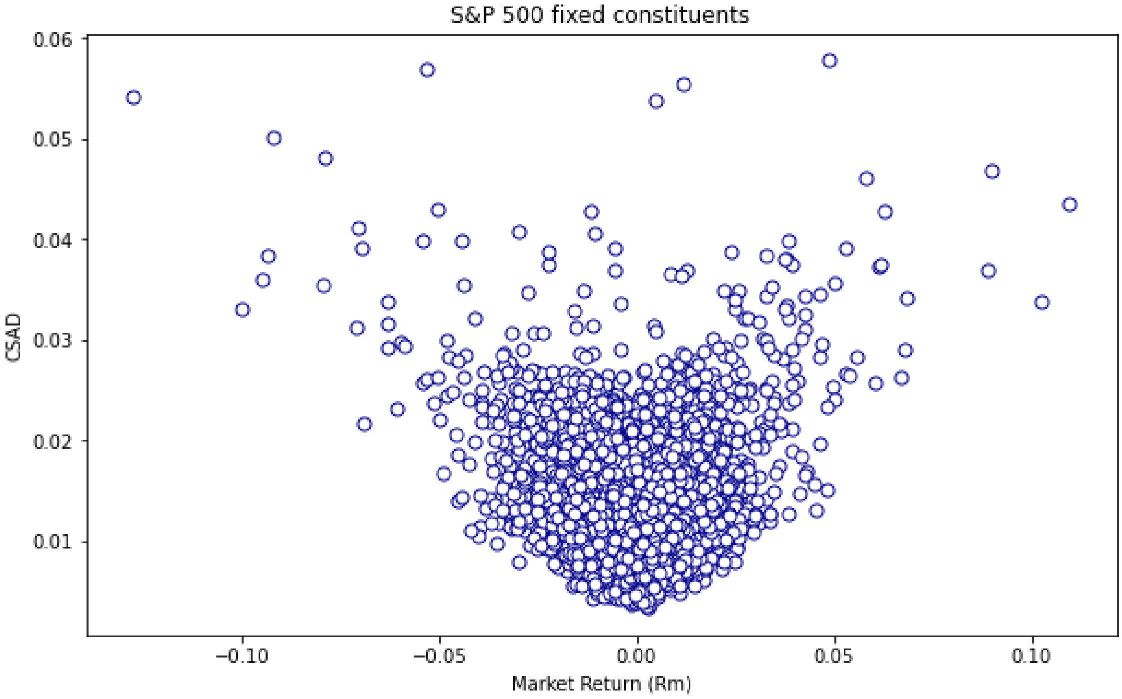

Figure 1. The relationship between daily cross-sectional absolute deviations (CSAD) of the fixed constituents, and S&P 500 market returns (Rm)

Based on figure 1 above, the CSADs are generally lower when market returns are low, and they increase when market return increases. There are no significant asymmetries in the graphs which would suggest herding. The market returns are slightly positively skewed which applies also to CSADs. Some relatively high CSAD values can be detected even when market return is close to zero.

The figure 2 below does not seem to display significant changes in comparison to previous figure. There is possibly still slightly more dispersion, and the highest CSAD value is higher than in figure 1.

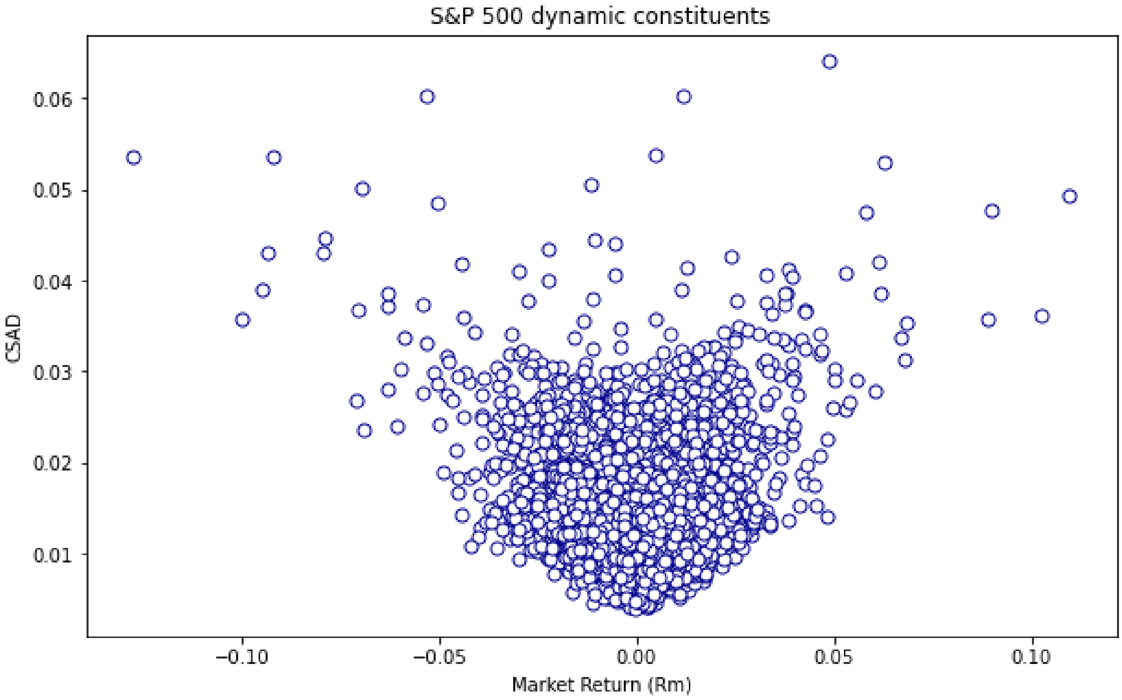

Figure 2. The relationship between daily cross-sectional absolute deviations (CSAD) of the dynamic constituents, and S&P 500 market returns (Rm)

7 Empirical Results

In the empirical results section, the regression results are provided based on the equations stated in the methodology section. First the results are provided for the equation (10), which will be followed by results for QR of equation (11). From now on, in the result tables, the S&P 500 fixed constituents are referred to as S&P 500 fixed, and the S&P 500 dynamic constituents are referred to as S&P 500 dynamic.

Below in table 2 will be the regression results for equation (10). As a reminder, a statistically significant negative γ_3 coefficient would be an indication of herding.

Table 2. Regression results for herding over the entire sample period

Regression results of herding over the sample period 03.01.1989-13.01.2025.

	γ_0	γ_1	γ_2	γ_3	Adj. R^2
Index					
S&P 500 fixed	0,010 ***	0,008 **	0,275 ***	0,932 ***	0,297
	(142,577)	(2,211)	(30,510)	(5,589)	
S&P 500 dynamic	0,010 ***	0,008 **	0,289 ***	0,995 ***	0,296
	(142,660)	(2,011)	(30,441)	(5,659)	

The T-statistics are in parentheses.
* Statistically significant at the 10% level
** Statistically significant at the 5% level
*** Statistically significant at the 1% level

As seen in table 2, there is no implication of herding as the (γ_3) coefficient is positive and statistically significant at the 1% level. Instead, this suggests negative herding, which is generally referred to as anti-herding.

The QR results of the whole sample period for equation (11) are in the following table 3. A statistically significant negative γ_2 would suggest herding.

Table 3. QR results for herding over the entire sample period

QR results for herding over the sample period 03.01.1989-13.01.2025.

		γ_0	γ_1	γ_2	Pseudo R^2
Index	Quantile				
S&P 500 fixed	τ=10%	0,006 ***	0,133 ***	1,360 ***	0,610
		(111,197)	(16,744)	(8,021)	
	τ=25%	0,007 ***	0,168 ***	1,548 ***	0,276
		(123,732)	(20,332)	(9,452)	
	τ=50%	0,009 ***	0,224 ***	1,441 ***	0,107
		(124,044)	(22,248)	(7,739)	
	τ=75%	0,011 ***	0,293 ***	1,840 ***	0,295
		(135,502)	(27,513)	(11,118)	
	τ=90%	0,013 ***	0,460 ***	0,009	0,582
		(88,413)	(24,737)	(0,036)	
S&P 500 dynamic	τ=10%	0,007 ***	0,127 ***	1,625 ***	0,573
		(132,654)	(16,734)	(10,019)	
	τ=25%	0,008 ***	0,161 ***	1,661 ***	0,252
		(148,406)	(20,258)	(10,021)	
	τ=50%	0,010 ***	0,197 ***	2,508 ***	0,114
		(153,595)	(22,837)	(15,719)	
	τ=75%	0,012 ***	0,335 ***	1,687 ***	0,287
		(122,549)	(28,009)	(9,399)	
	τ=90%	0,014 ***	0,546 ***	-0,658 *	0,567
		(68,309)	(21,689)	(1,824)	

The T-statistics are in parentheses.
* Statistically significant at the 10% level
** Statistically significant at the 5% level
*** Statistically significant at the 1% level

Based on the QR results above, generally there is no sign of herding, except for the S&P 500 dynamic index at quantile 90% for which the γ_2 coefficient is negative and statistically significant at the 10% level. This herding is revealed due to QR attributes of being able to examine specific quantiles.

Next, the second null hypothesis H_0^2, which states that the results of used CSAD equation and QR are similar, is being examined. Based on tables 2 and 3, one can detect that the herding results of coefficient γ_3 for equation (10) and γ_2 for equation (11) are relatively similar. The γ_3 coefficients are still lower than the γ_2 coefficients, and the previously mentioned statistically significant negative γ_2 coefficient for dynamic index QR differs from the herding results of the general CSAD equation. Due to this, the null hypothesis H_0^2 is being rejected, and the alternative hypothesis H_a^2 which states that the results between CSAD equation and the QR differ, is accepted.

The QR results of equation (11) for sub-periods will be in the following tables 4, 5, and 6. The sub-periods are otherwise 5 years, except for the last sub-period, which is slightly over 6 years. The sub-periods reveal significant variation in the results for herding coefficient γ_2, as will be discussed after the tables.

Table 4. QR results of herding for first three sub-periods

QR results of herding for different sub-periods.

Sub-period 03.01.1989 – 31.12.1993								
		γ_0		γ_1		γ_2		Pseudo R^2
Index	Quantile							
S&P 500 dynamic	τ=10%	0,010	***	0,055		3,769	*	0,049
		(59,760)		(1,238)		(1,794)		
	τ=25%	0,010	***	0,070	***	3,279	***	0,052
		(88,942)		(2,626)		(2,981)		
	τ=50%	0,012	***	0,088	***	4,298	***	0,077
		(102,439)		(4,015)		(5,781)		
	τ=75%	0,013	***	0,115	***	5,709	***	0,118
		(78,058)		(3,887)		(7,063)		
	τ=90%	0,014	***	0,216	***	3,316	***	0,169
		(60,074)		(5,177)		(3,278)		

Sub-period 03.01.1994 – 31.12.1998								
		γ_0		γ_1		γ_2		Pseudo R^2
Index	Quantile							
S&P 500 dynamic	τ=10%	0,009	***	0,108	***	1,907	**	0,099
		(84,179)		(4,734)		(2,476)		
	τ=25%	0,010	***	0,125	***	1,506	***	0,111
		(112,608)		(7,729)		(3,456)		
	τ=50%	0,011	***	0,114	***	3,536	***	0,132
		(124,743)		(7,540)		(9,428)		
	τ=75%	0,012	***	0,166	***	3,971	***	0,164
		(87,007)		(7,446)		(8,238)		
	τ=90%	0,014	***	0,182	***	4,952	***	0,191
		(46,481)		(3,952)		(6,138)		

Sub-period 04.01.1999 – 31.12.2003								
		γ_0		γ_1		γ_2		Pseudo R^2
Index	Quantile							
S&P 500 dynamic	τ=10%	0,010	***	0,122	**	3,312	**	0,069
		(30,800)		(2,492)		(2,287)		
	τ=25%	0,012	***	0,181	***	2,020		0,079
		(34,980)		(3,505)		(1,371)		
	τ=50%	0,016	***	0,197	***	1,243		0,070
		(38,831)		(3,447)		(0,805)		
	τ=75%	0,020	***	0,034		8,634	***	0,092
		(50,142)		(0,634)		(5,880)		
	τ=90%	0,023	***	0,095		8,527	***	0,126
		(36,953)		(1,116)		(3,703)		

The T-statistics are in parentheses.
* Statistically significant at the 10% level
** Statistically significant at the 5% level
*** Statistically significant at the 1% level

Table 5. QR results of herding for the following three sub-periods

QR results of herding for different sub-periods.

Sub-period 02.01.2004 – 31.12.2008								
		γ_0		γ_1		γ_2		Pseudo R^2
Index	Quantile							
S&P 500 dynamic	τ=10%	0,007	***	0,119	***	2,277	***	0,123
		(70,467)		(8,064)		(7,867)		
	τ=25%	0,008	***	0,164	***	1,953	***	0,161
		(83,441)		(12,527)		(8,585)		
	τ=50%	0,009	***	0,208	***	3,033	***	0,225
		(76,715)		(14,177)		(14,517)		
	τ=75%	0,010	***	0,382	***	2,346	***	0,311
		(45,831)		(14,033)		(7,363)		
	τ=90%	0,011	***	0,683	***	-1,807	***	0,414
		(30,940)		(14,663)		(-3,437)		

Sub-period 02.01.2009 – 31.12.2013								
		γ_0		γ_1		γ_2		Pseudo R^2
Index	Quantile							
S&P 500 dynamic	τ=10%	0,007	***	0,089	***	2,268	***	0,107
		(55,136)		(4,116)		(3,640)		
	τ=25%	0,008	***	0,057	***	4,156	***	0,132
		(71,256)		(3,404)		(9,908)		
	τ=50%	0,009	***	0,046	**	6,673	***	0,167
		(73,032)		(2,541)		(15,372)		
	τ=75%	0,010	***	0,169	***	6,245	***	0,223
		(47,083)		(5,691)		(9,521)		
	τ=90%	0,011	***	0,580	***	-0,729		0,278
		(22,557)		(8,675)		(-0,564)		

Sub-period 02.01.2014 – 31.12.2018								
		γ_0		γ_1		γ_2		Pseudo R^2
Index	Quantile							
S&P 500 dynamic	τ=10%	0,006	***	0,089	***	1,293		0,051
		(53,457)		(3,259)		(1,189)		
	τ=25%	0,007	***	0,108	***	0,769		0,064
		(69,653)		(4,434)		(0,764)		
	τ=50%	0,008	***	0,150	***	-0,104		0,078
		(64,209)		(5,490)		(-0,104)		
	τ=75%	0,009	***	0,141	***	1,805	*	0,093
		(60,797)		(4,332)		(1,663)		
	τ=90%	0,011	***	0,075		7,414	***	0,103
		(50,014)		(1,554)		(4,288)		

The T-statistics are in parentheses.
* Statistically significant at the 10% level
** Statistically significant at the 5% level
*** Statistically significant at the 1% level

Table 6. QR results of herding for the last sub-period

QR results of herding for different sub-periods.

Sub-period 02.01.2019 – 13.01.2025					
		γ_0	γ_1	γ_2	Pseudo R^2
Index	Quantile				
S&P 500 dynamic	τ=10%	0,007 ***	0,100 ***	1,837 ***	0,103
		(62,062)	(7,010)	(7,805)	
	τ=25%	0,008 ***	0,131 ***	1,745 ***	0,114
		(66,438)	(8,908)	(8,760)	
	τ=50%	0,010 ***	0,178 ***	1,301 ***	0,141
		(73,321)	(11,351)	(5,763)	
	τ=75%	0,012 ***	0,204 ***	2,233 ***	0,163
		(54,693)	(8,574)	(7,069)	
	τ=90%	0,013 ***	0,428 ***	-0,353	0,228
		(39,886)	(11,686)	(-0,816)	

The T-statistics are in parentheses.
* Statistically significant at the 10% level
** Statistically significant at the 5% level
*** Statistically significant at the 1% level

As seen in the result tables above, some sub-periods possess quite low variation in herding coefficient γ_2 between the different quantiles, and some have significant differences between the quantiles. One period with relatively constant herding results for all quantiles is the subperiod 03.01.1989 – 31.12.1993, which suggests that the amount of stock return dispersion has not affected herding significantly. Whereas for the sub-period 04.01.1999 – 31.12.2003 herding coefficients have been moderate for lower quantiles, but the quantiles 75% and 90% have experienced strong anti-herding.

As mentioned before, the statistically significant negative γ_2 coefficient would suggest the existence of herding. Probably the most significant finding from the results is the herding coefficient (-1,807) at the quantile 90% for the sub-period 02.01.2004 – 31.12.2008. It is important to note that this period includes the beginning and some of the duration of the GFC which supports the assumption that herding intensifies during crisis periods. In addition, there are some other slightly negative γ_2 coefficients during other sub-periods, but as they are not statistically significant, they give no support for existence of herding.

Table 7 below provides the results for equation (12) of QR with a dummy variable for bubbles. The coefficient γ_2 denotes herding during non-bubble periods, and the coefficient γ_3 during bubble periods.

Table 7. QR results of herding for non-bubble and bubble periods

QR results of herding for non-bubble and bubble periods during 03.01.1989-13.01.2025.

		γ_0	γ_1	γ_2	γ_3	Pseudo R^2
Index	Quantile					
S&P 500 dynamic	τ=10%	0,007 ***	0,126 ***	1,641 ***	6,248 ***	0,120
		(133,474)	(16,630)	(10,160)	(2,804)	
	τ=25%	0,008 ***	0,160 ***	1,676 ***	6,490 ***	0,278
		(148,476)	(20,027)	(10,098)	(3,724)	
	τ=50%	0,010 ***	0,192 ***	2,560 ***	7,366 ***	0,752
		(154,034)	(22,232)	(16,050)	(4,067)	
	τ=75%	0,012 ***	0,334 ***	1,703 ***	3,767	1,407
		(122,541)	(27,801)	(9,477)	(1,526)	
	τ=90%	0,014 ***	0,555 ***	-0,808 **	-8,275	1,724
		(68,215)	(21,921)	(-2,234)	(-1,421)	

The T-statistics are in parentheses.
* Statistically significant at the 10% level
** Statistically significant at the 5% level
*** Statistically significant at the 1% level

According to the result table above, the results for non-bubble and bubble periods deviate significantly. At the lower quantiles the γ_2 coefficient of non-bubble period is distinctively lower than the γ_3 coefficient of bubble periods which suggests strong anti-herding at the 1% statistical significance level. This relationship applies conversely to upper quantiles as the γ_2 has a lower negative value, whereas the γ_3 has much higher negative value. However, it is important to notice that only the non-bubble period γ_2 coefficient for quantile 90% is negative (-0,808) and statistically significant at the 5% level, and thus suggesting herding. Even though the γ_3 herding coefficient for bubble period is highly negative, it is not statistically significant and therefore does not imply herding.

Based on the previous results, the third null hypothesis H_0^3 which states that based on the methodology used, there is no difference in investor herding results for non-bubble and bubble periods, is being tested. As can be seen in the previous table and as described in the paragraph above, there are distinct differences in the herding results for coefficient γ_2 of non-bubble, and γ_3 coefficient of bubble periods, and thus the null hypothesis H_0^3 is rejected. The third alternative hypothesis H_a^3 which states that based on the methodology used, there is a difference in herding coefficients for non-bubble and bubble periods, is accepted. Accepting the alternative hypothesis suggests that a relationship exists between investor herding and stock market bubbles.

8 Conclusions

This thesis has examined investor herding in S&P 500 index, and the relationship of herding and stock market bubbles. The herding results of this study are inconclusive to a certain degree. This is probably caused by the assumptions made, and the limitations of the methods used. The complexity of herding as a phenomenon additionally causes limitations to measure it. The academic herding methods available do not seem to measure it accurately, as they may be biased against finding herding or not well oriented to detect it.

The fixed and dynamic index composition were used in this study to investigate whether there are differences in herding between different index compositions. Separate compositions were also used to see if index survivorship bias would affect the herding behavior. Both indexes were affected by the survivorship bias to some extent, as in the fixed index, the stocks stayed in the index for whole period or the ones that got deleted during the period, eventually got added back before the end. The dynamic index is affected by the survivorship bias because stocks with weak performance get deleted and stocks with good performance get added to the index.

The results still suggest that the CSAD method does not capture the variation in herding in a similar way as the QR method does. An important finding regarding the QR is the irregularity in herding results between different quantiles. The QR sub-period analysis unveils the significant differences in herding between separate sub-periods. The QR analysis with the bubble dummy indicates that there is prominent divergence in herding results between non-bubble and bubble periods. This suggests that bubble periods or excessive returns affect investor behavior and herding.

Regardless of the limited evidence for investor herding in this study, herding supposedly is an important factor in stocks markets. Such incapability to reliably measure herding does not imply that herding does not exist, but it indicates the apparent need for more accurate herding measures.

Stock market bubbles have gained distinctively limited attention in the literature which implies that there would be wide possibilities for future research on them. New measures on bubbles should be developed, and the existing and new measures could be incorporated into other financial models. Possibilities for future research also include studying the relationship between herding and bubbles more profoundly, and in different indexes and markets. Additionally, the causality of herding and bubbles could be investigated. The index survivorship bias should be increasingly accounted for in future research by measuring its effect, and by creating indexes which aim to minimize the survivorship bias. By combining all these topics in future research, new revelations could be made from the effect of survivorship bias on herding results, and possible relationship with stock market bubbles.

References

Andrews, D. W. K. (2003). End-of-Sample Instability Tests. *Econometrica, 71*(6), 1661-1694. https://doi.org/10.1111/1468-0262.00466

Andrews, D. W. K., & Kim, J. (2006). Tests for Cointegration Breakdown Over a Short Time Period. *Journal of business & economic statistics, 24*(4), 379-394. https://doi.org/10.1198/073500106000000297

Astill, S., Harvey, D. I., Leybourne, S. J., & Taylor, A. M. R. (2017). Tests for an end-of-sample bubble in financial time series. *Econometric reviews, 36*(6-9), 651-666. https://doi.org/10.1080/07474938.2017.1307490

Astill, S., Harvey, D. I., Leybourne, S. J., Sollis, R., & Robert Taylor, A. M. (2018). Real-Time Monitoring for Explosive Financial Bubbles. *Journal of time series analysis, 39*(6), 863-891. https://doi.org/10.1111/jtsa.12409

Barro, R. J., & Ursúa, J. F. (2017). Stock-market crashes and depressions. *Research in economics, 71*(3), 384-398. https://doi.org/10.1016/j.rie.2017.04.001

Bekiros, S., Jlassi, M., Lucey, B., Naoui, K., & Uddin, G. S. (2017). Herding behavior, market sentiment and volatility: Will the bubble resume? *The North American journal of economics and finance, 42*, 107-131. https://doi.org/10.1016/j.najef.2017.07.005

BenSaïda, A. (2017). Herding effect on idiosyncratic volatility in U.S. industries. *Finance research letters, 23*, 121-132. https://doi.org/10.1016/j.frl.2017.03.001

Bikhchandani, S., & Sharma, S. (2000). Herd behavior in financial markets. *IMF staff papers, 47*(3), 279-310. https://doi.org/10.2307/3867650

Black, F. (1986). Noise. *The Journal of finance (New York), 41*(3), 529-543. https://doi.org/10.2307/2328481

Bohl, M. T., Branger, N., & Trede, M. (2017). The case for herding is stronger than you think. *Journal of banking & finance, 85*, 30-40. https://doi.org/10.1016/j.jbankfin.2017.08.006

Brav, A., Heaton, J., & Li, S. (2010). The Limits of the Limits of Arbitrage. *REVIEW OF FINANCE, 14*(1), 157-187. https://doi.org/10.1093/rof/rfp018

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of finance (New York), 52*(1), 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

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. https://doi.org/10.1016/S0378-4266(99)00096-5

Chiang, T. C., & Zheng, D. (2010). An empirical analysis of herd behavior in global stock markets. *Journal of banking & finance, 34*(8), 1911-1921. https://doi.org/10.1016/j.jbankfin.2009.12.014

Chiang, T. C., Li, J., & Tan, L. (2010). Empirical investigation of herding behavior in Chinese stock markets: Evidence from quantile regression analysis. *Global finance journal, 21*(1), 111-124. https://doi.org/10.1016/j.gfj.2010.03.005

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. https://doi.org/10.2469/faj.v51.n4.1918

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *The Journal of political economy, 98*(4), 703-738. https://doi.org/10.1086/261703

Diba, B. T., & Grossman, H. I. (1988). Explosive Rational Bubbles in Stock Prices? *The American economic review, 78*(3), 520-530. https://www.jstor.org/stable/1809149

Dickey, D. A., & Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica, 49*(4), 1057-1072. https://doi.org/10.2307/1912517

Duygun, M., Tunaru, R., & Vioto, D. (2021). Herding by corporates in the US and the Eurozone through different market conditions. *Journal of international money and finance, 110*, 102311. https://doi.org/10.1016/j.jimonfin.2020.102311

Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial analysts journal, 21*(5), 55-59. https://doi.org/10.2469/faj.v21.n5.55

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of finance (New York), 25*(2), 383. https://doi.org/10.2307/2325486

FAMA, E. F., & FRENCH, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of finance (New York), 47*(2), 427-465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (2004). The Capital Asset Pricing Model: Theory and Evidence. *The Journal of economic perspectives, 18*(3), 25-46. https://doi.org/10.1257/0895330042162430

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics, 116*(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of financial economics, 128*(2), 234-252. https://doi.org/10.1016/j.jfineco.2018.02.012

Garcia, C., & Gould, F. (1993). Survivorship Bias. *Journal of portfolio management, 19*(3), 52-56. https://doi.org/10.3905/jpm.1993.409445

Harvey, D. I., Leybourne, S. J., Sollis, R., & Taylor, A. M. (2018). Detecting regimes of predictability in the US equity premium. Retrieved 7.11.2024 from https://scholar.google.com/scholar?as_q=%22Detecting+regimes+of+predictability+in+the+U.S.+equity+premium%22&ie=utf8&oe=utf8

Homm, U., & Breitung, J. (2012). Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods. *Journal of financial econometrics, 10*(1), 198-231. https://doi.org/10.1093/jjfinec/nbr009

JEGADEESH, N., & TITMAN, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of finance (New York), 48*(1), 65-91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291. https://doi.org/10.2307/1914185

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). ostrich effect: Selective attention to information. *Journal of risk and uncertainty, 38*(2), 95-115. https://doi.org/10.1007/s11166-009-9060-6

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica, 46*(1), 33-50. https://doi.org/10.2307/1913643

Koenker, R., & Machado, J. A. F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association, 94*(448), 1296-1310. https://doi.org/10.1080/01621459.1999.10473882

Li, Z., Diao, X., & Wu, C. (2022). The influence of mobile trading on return dispersion and herding behavior. *Pacific-Basin finance journal, 73*, 101767. https://doi.org/10.1016/j.pacfin.2022.101767

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The review of economics and statistics, 47*(1), 13-37. https://doi.org/10.2307/1924119

Litimi, H., BenSaïda, 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. https://doi.org/10.1016/j.ribaf.2016.03.008

Ljungqvist, A., & Qian, W. (2016). How Constraining Are Limits to Arbitrage? *The Review of financial studies, 29*(8), 1975-2028. https://doi.org/10.1093/rfs/hhw028

Lux, T. (1995). Herd Behaviour, Bubbles and Crashes. *The Economic journal (London), 105*(431), 881-896. https://doi.org/10.2307/2235156

Markowitz, H. (1952). PORTFOLIO SELECTION. *The Journal of finance (New York), 7*(1), 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

Nofsinger, J. R., & Sias, R. W. (1999). Herding and Feedback Trading by Institutional and Individual Investors. *The Journal of finance (New York), 54*(6), 2263-2295. https://doi.org/10.1111/0022-1082.00188Park, A., & Sabourian, H. (2011). Herding and Contrarian Behavior in Financial Markets. *Econometrica, 79*(4), 973-1026. https://doi.org/10.3982/ECTA8602

Phillips, P. C. B., Shi, S., & Yu, J. (2015). TESTING FOR MULTIPLE BUBBLES: HISTORICAL EPISODES OF EXUBERANCE AND COLLAPSE IN THE S&P 500. *International economic review (Philadelphia), 56*(4), 1043-1078. https://doi.org/10.1111/iere.12132

Phillips, P. C. B., Wu, Y., & Yu, J. (2011). EXPLOSIVE BEHAVIOR IN THE 1990s NASDAQ: WHEN DID EXUBERANCE ESCALATE ASSET VALUES? *International economic review (Philadelphia), 52*(1), 201-226. https://doi.org/10.1111/j.1468-2354.2010.00625.x

Pochea, M., Filip, A., & Pece, A. (2017). Herding Behavior in CEE Stock Markets Under Asymmetric Conditions: A Quantile Regression Analysis. *The journal of behavioral finance, 18*(4), 400-416. https://doi.org/10.1080/15427560.2017.1344677

Ramiah, V., Xu, X., & Moosa, I. A. (2015). Neoclassical finance, behavioral finance and noise traders: A review and assessment of the literature. *International review of financial analysis, 41*, 89-100. https://doi.org/10.1016/j.irfa.2015.05.021

ROLL, R., & ROSS, S. A. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. *The Journal of finance (New York), 35*(5), 1073-1103. https://doi.org/10.1111/j.1540-6261.1980.tb02197.x

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of economic theory, 13*(3), 341-360. https://doi.org/10.1016/0022-0531(76)90046-6

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of finance (New York), 19*(3), 425. https://doi.org/10.2307/2977928

Shefrin, H., & Statman, M. (2000). Behavioral Portfolio Theory. *Journal of financial and quantitative analysis, 35*(2), 127-151. https://doi.org/10.2307/2676187

Shiller, R. J. (2003). From Efficient Markets Theory to Behavioral Finance. *The Journal of economic perspectives, 17*(1), 83-104. https://doi.org/10.1257/089533003321164967

Shleifer, A., & Vishny, R. W. (1997). The Limits of Arbitrage. *The Journal of finance (New York), 52*(1), 35-55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Statman, M. (2018). A Unified Behavioral Finance. *Journal of portfolio management, 44*(7), 124-134. https://doi.org/10.3905/jpm.2018.44.7.124

Ukpong, I., Tan, H., & Yarovaya, L. (2021). Determinants of industry herding in the US stock market. *Finance research letters, 43*, 101953. https://doi.org/10.1016/j.frl.2021.101953

Venezia, I., Nashikkar, A., & Shapira, Z. (2011). Firm specific and macro herding by professional and amateur investors and their effects on market volatility. *Journal of banking & finance, 35*(7), 1599-1609. https://doi.org/10.1016/j.jbankfin.2010.11.015

Vogel, H. L., & Werner, R. A. (2015). An analytical review of volatility metrics for bubbles and crashes. *International review of financial analysis, 38*, 15-28. https://doi.org/10.1016/j.irfa.2014.11.003

Wang, Y., Yu, C., & Zhao, X. (2024). The asymmetric effect of investor herding on the asset price bubble - evidence from the Chinese stock market. *Applied economics*. https://doi.org/10.1080/00036846.2024.2319651

Wanidwaranan, P., & Padungsaksawasdi, C. (2022). Unintentional herd behavior via the Google search volume index in international equity markets. *Journal of international financial markets, institutions & money, 77*, 101503. https://doi.org/10.1016/j.intfin.2021.101503

Whitehouse, E. J., Harvey, D. I., & Leybourne, S. J. (2023). Real-Time Monitoring of Bubbles and Crashes. *Oxford bulletin of economics and statistics, 85*(3), 482-513. https://doi.org/10.1111/obes.12540