Measuring US Mutual Fund Herding in Industries

Master Thesis

School of Business and Economics

Maastricht University

Submitted to:	Matteo Bonetti (Finance)
Submitted by:	Ichio Aoki
Student ID:	i6201190
Degree Programme:	Master of Financial Economics (M.Sc.)
Address:	Johann-Zimmermann-Weg 9, 53343 Wachtberg, Germany
Phone:	(+)49 176 7277 5068
E-Mail:	i.aoki@student.maastrichtuniversity.nl

Maastricht, July 14, 2019

Contents

1 Introduction

Herd mentality is defined as ”The tendency for people's behaviour or beliefs to conform to those of the group to which they belong” in the Oxford dictionary. Such behavior can be studied in many different contexts, including in financial markets. In particular, herd behavior may arise among institutional investors in the form of correlated investment decisions. Whether such behavior is efficient or not depends on the determinants of this behavior, as it may either arise spuriously due to similar strategies or information sets of managers (spurious herding) or occur due to an active decision by a manager to copy their peers (intentional herding). The former is usually in line with market efficiency and is in fact expected in a fully efficient market, while the latter may cause information blockage or lead to market instability. The implications of intentional herding are profound, as such actions may lead to inefficiently priced securities, market bubbles and excess volatility in the financial markets.

Past literature has focused primarily on detecting herd behavior among market participants, with the most popular methodologies employed relying on detecting herding as either contemporaneously correlated investment decisions among a subset of investors (Lakonishok et al., 1992) or as cross-sectional market consensus among all market participants in the form of return dispersion (Chang et al., 2000; Christie and Huang, 1995). Many posit that such behavior arises as a result of similar investment strategies, namely momentum investing by institutional actors (Grinblatt et al., 1995) and not through intentional decisions by investors to copy their peers, though they do not explicitly test for other drivers of herding. This is in part due to the difficulty of doing so, as it is only the actions of fund managers that are observable and not their decision making process (Spyrou, 2013).

In more recent research, herding is thought of as a dynamic process measured as the correlation between an investors current trades with the past trades of other market participants. This kind of herd behavior is more in line with theoretical literature on intentional herding, which tend to be settings in which investors act upon observing their counterparts actions (Banerjee, 1992; Bikhchandani and Sharma, 2001; Scharfstein and Stein, 1990) and was first introduced by a seminal paper by Sias (2004). Studies employing this measure confirm inter-temporal dependence among investor decisions and much of this behavior cannot be solely attributed to momentum following strategies (Sias, 2004). Recent work by Gavriilidis et al. (2013) and Holmes et al. (2013) are exceptions to this rather agnostic attitude towards the determinants of herding and find evidence supporting the hypothesis that fund managers do at times herd intentionally in Spanish and Portuguese markets. Jiang and Verardo (2018) also analyze potential determinants of herding behavior at the fund level, uncovering several relations and also find that there exists a negative relation between the herding tendency of a fund manager and and the future performance displayed.

The motivations behind the appearance of herd behavior at the security level are applicable to the industry level (Choi and Sias, 2009). Investors may invest in e.g. pharmaceuticals because everyone else is doing so, and thereby attempt to preserve their reputation by following their peers. As such, this thesis contributes to the smaller strand of research on industrial herding by institutional investors (Chen et al., 2013; Choi and Sias, 2009; Holmes et al., 2013; Voronkova and Bohl, 2005), rather than the more popular stock (Lakonishok et al., 1992; Sias, 2004) or asset class (Blake et al., 2017) levels of herding.

Using methodology based on Jiang and Verardo (2018), a measure to estimate the herding tendency of a fund is defined for the industry trading level. Based on this measure, this thesis extends the work of Gavriilidis et al. (2013), Holmes et al. (2013) and Jiang and Verardo (2018) by investigating the drivers of herd behavior based on theoretical models describing such phenomena. The variation in this herding measure across funds and across time is exploited to address the gap in the literature concerning the link between drivers of herding behavior and the phenomenon itself. The underlying thought is that variation in herding behavior may arise in conjunction with variation in various characteristics as predicted by the underlying theories of herding, such as noise in the information of funds, the skill of managers or reputationally linked variables. Confirming the existence of such relations empirically would provide support for the theories on herd behavior.

Industrial herding by US mutual funds is found to be economically significant and vary across time and in the cross-section. This is despite controlling for lagged returns (momentum characteristic), suggesting that some of the detected herding may not be spurious and be the result of active decisions by fund managers to do so. Market volatility, a proxy for the information to noise ratio of signals received by fund managers, exhibits a weak positive relationship with herding tendency, lending support to some predictions based on informational cascade models (Banerjee, 1992; Bikhchandani et al., 1992). Further analysis using proxies for reputation and skill such as management fees and returns show mixed results, leading to the conclusion that industry level herding is to some degree motivated by reputational aspects.

The remainder of this thesis is structured as follows: Section (2) briefly summarizes the theories on herd behavior and empirical results in this regard. Hypothesis motivated by previous findings are formulated and the contributions of this thesis are explained. Section (3) presents the data used in the analysis and outlines the construction process of the needed data-set. This is followed by a description of the methodology employed in section (4). Section (5) provides an overview of the results. Section (6) goes into detail and links these results to past future, while also discussing the limitations and adding suggestions for the path of future research into herd behavior. Section (7) concludes.

2 Literature

Herd behavior can be defined as a collective of individuals acting in a correlated manner without there necessarily being a central authority directing such action. Such behavior can be found in all settings, from animals to human decision making. The focus of this thesis is on herd behavior in financial markets, particularly among institutional investors and specifically in mutual fund managers' investment decisions.

Studying herd behavior in the context of financial markets and institutional investors is of interest due to the potential on part of these actors to impact market prices and stability. Many studies have found positive relationships between institutional trading and returns across the same period (Cai and Zheng, 2004; Grinblatt et al., 1995; Nofsinger and Sias, 1999; Sias et al., 2001) and some conclude that these effects are a result of the trading among institutional investors (Barclay and Warner, 1993; Chakravarty, 2001; Sias, 2004). In light of such evidence, it is important to understand whether these processes are of an efficient nature or whether the movements may be in excess of what the information at the time suggests. Herding, particularly of the intentional type which will be elaborated on below, may indeed be a cause of excess price movements and thereby facilitate instability in the markets.

The following subsections briefly summarize the various findings on institutional investor herding related to the topic of this thesis and introduces the theories explaining such behavior. More extensive reviews covering the entirety of herding literature in financial economics can be found in Bikhchandani and Sharma (2001), Hirshleifer and Hong Teoh (2003), Spyrou (2013) and Welch (1992). Chen et al.

(2017) provide a graphical overview of all findings prior their publication, with a focus placed on the empirical evidence of herding behavior, which serves as an overview of the evidence of the presence of herd behavior.

2.1 Theories on Herd Behavior

Theoretical work on herd behavior is primarily concerned with answering the following question: "Why would agents exhibit herd behavior?". In the setting of this thesis, agents refers specifically to mutual fund managers. Devenow and Welch (1996) provide an overview of the theories predicting such behavior in a rational setting, which this thesis will also employ. An important distinction can be made between intentional herding and spurious herding (Bikhchandani and Sharma, 2001). Spurious herding refers to similar actions on behalf of agents without necessarily targeting such an outcome. This means that when observing data, what may at first glance seem like herd behavior in the form of correlated trading is actually only the result of a reaction by all market participants towards the efficient market. In contrast, intentional herding arises from a direct intention on the part of an agent to imitate another, potentially abandoning their own information in order to do so.

Spurious herding is in line with market efficiency, as investors are making the same decisions regarding asset allocation precisely because it is the efficient allocation. In this case they are assisting in the process of incorporating fundamental information and investor beliefs into prices in an efficient manner. In other words, in an efficient market, one would expect to see contemporaneous correlation in the actions of investors. The information incorporated into prices is based on the asset in question, which is not the case for intentional herding. Here, information flows based on beliefs and fundamentals may be blocked and instead investors may be acting while completing ignoring beliefs pertaining to the securities or assets. This may lead to mispricings, bubbles and a violation of market efficiency as prices do not incorporate all information available. Of course, just because such behavior is inefficient does not mean that it is irrational. Intentional herding can be a rational decision by the investor, in the sense that it is the expected utility maximizing choice from the set of actions.

A finer differentiation between various drivers of herding behavior leads to several different theoretical reasons: the focus here is placed on informational cascades, reputational herding and characteristics herding (Graham, 1999; Sias, 2004; Welch, 1992). The former two can be categorized as being part of the intentional herding literature, while the latter is more suited to be characterized as spurious herding. Indārs et al. (2019) provide a detailed typology of herding behavior along this dimension.

Characteristics herding arises because investors pursue strategies for selecting stock with similar characteristics. A major characteristic that may be thought to contribute to herding is positive feedback or momentum trading (Grinblatt et al., 1995; Lakonishok et al., 1992; Sias, 2004), whereby investors follow the strategy of buying past winners and selling past losers. In a more general sense, characteristics herding should arise in an efficient market, as any changes in the market portfolio will register as herd behavior. Of course, herd behavior in this regard is fundamentally different from the kind associated with the intentional strand of literature, as herding is an outcome of the market and not a driver thereof. Moreover, the herding behavior exhibited due to a spurious nature should occur contemporaneously. This kind of herd behavior is uninteresting for the scope of this analysis, as the interest here lies in linking theories of intentional herding drivers with the empirical data.

In an informational cascade, agents ignore their own signals in favor of the actions of the agents who moved beforehand, leading to a blockage of information (Bikhchandani et al., 1992). Other agents actions may convey information such that future agents ignore private information in favor of other agents actions. Investors may herd towards a certain asset based only on the observation of others doing so in the past. Banerjee (1992) shows that such behavior may lead to an inefficient herding outcome. Related to this area of herding theory, costly information acquisition may increase likelihood of herd behavior arising, as agents may choose to free-ride by acting on the signals inferred from the actions of other agents. Banerjee (1992) asserts that adding such a cost dimension to the model would only exacerbate the herding problem.

The most relatable illustration of an informational cascade is the choice of a restaurant, and example also used by Banerjee (1992). The choice of a restaurant may depend not only on one's own private information but also on the number of customers already present in each establishment, as an empty restaurant is hard to choose over the lively neighbor regardless of what one heard about each establishment before arriving. This especially true if the information one had prior is from an unreliable source, which can be thought of as a noisy signal. This can easily be projected to financial decisions, as an investor may be willing to suppress a personal negative belief about a stock if many other investors seem to be particularly bullish on it, perhaps in the process coming to believe that ones belief was the result of faulty analysis or information. Based on such models, the prediction goes that the more inaccurate or noisy the personal signal, the more likely it is to be discarded by the agent in favor of herding onto the previous agents' actions.

Reputational herding can be seen as a form of principal-agent issue related to behavior. Investors with particularly high concerns about their relative reputation or standing may choose to simply follow what others are doing to not stand out should things go badly (Scharfstein and Stein, 1990). The prospect of success while others fail does not outweigh the possible cost of being the only one to fail. In the broader principal-agent context, herd behavior may act as an insurance policy for a manager in the case of under-performance (Rajan, 2006). With the financial industry being known as rather cut-throat and relative performance being the most easily observed evaluation criterion, agents may have personal incentives to follow others and such an outcome would be rational from the perspective of the agent.

In a model based on Scharfstein and Stein (1990), Graham (1999) calculates several testable hypothesis related to herding likelihood and reputation. Investors concerned with reputation may ultimately choose to mimic others, as the risk of being wrong alone is not outweighed by the chance of being correct. This stems from the fact that evaluators ultimately perceive results and form their beliefs based on these. If one would like to be reputed as being skilled, then herding is the smart choice, as all 'smart' investors are likely to possess similar signals regarding investments, whereas those who are not will not. Ex-post, having made the same decision as others increases the belief that one is skilled. This compounds with the level of skill one has. Agents with high levels of skills are less likely to be wrong when following their own information compared to low-skilled types. Therefore, a high-skilled individual may be better off by not herding, as the expected gain from being apart from the crowd in a positive situation outweighs any expected costs of being the sole loser. Low-skilled individuals may instead prefer to take the 'certain' payoff by herding on the consensus.

2.2 Empirical Evidence

The majority of empirical literature on herding places the focus on detecting herd behavior. The methodologies employed to detect herding and analyze markets can be subdivided in various ways. The first such distinction categorizes herding based on the type of data studied: detection of market wide herding based on return dispersion or detecting herding among a subset of market participants based on micro-economic data. Market wide herding looks for a specific type of herding of all market participants towards a common behavior, while the latter types of herding are based on funds or other investors mimicking each other but not necessarily all converging towards the same market wide behavior.

Herding towards the markets is usually analyzed by observing the dispersion of returns of securities within a market at a given point in time. By regressing these cross-sectional indicators of herd behavior onto various explanatory variables, specific forms of herd behavior can be uncovered. Christie and Huang (1995) and Chang et al. (2000) were among the first to analyze such behavior and fail to find evidence for herding behavior. Following research based on this strand of herding is rather mixed and the general conclusion is that herding is stronger in emerging markets (Indārs et al., 2019).

Other methodologies rely on micro-level data, that is data on the trading behavior of investors, to create a measure of herding. These methods compute statistics based on the aggregation of this micro-data to develop stock or period level indices of herd behavior. One of the most influential studies of contemporaneous herding in this regard is that conducted by Lakonishok et al. (1992). Herding is measured as the excess demand of a stock by institutional investors in a quarter. They find little to no evidence of herd behavior based on this measure in their data-set of US pension funds spanning from 1985 to 1989 .

Sias (2004) recognizes the inherently inter-temporal nature of herding models, as behavior tends to be sequential in the sense that agents act after observing their peers actions. In this seminal framework, herding is measured as the period-to-period correlation between the percentage of buyers of a stock and a statistic is calculated for every quarter, indicating whether trading activity that quarter could be described as herd behavior. Sequential herding is significant and cannot be fully explained characteristics, leading to the conclusion that institutional investors engage in intentional herding.

Following these firm-level data based methodologies, other researchers find evidence of herding in various markets and among various subsets of market participants. Wermers (1999) finds that US mutual funds exhibit more herding behavior in small stocks, while Blake et al. (2017) find strong evidence of herding by UK pension funds. Raddatz and Schmukler (2013) find similar results for pension funds in Chile, while Iihara et al. (2001) find evidence of herding in Japanese markets by both domestic and foreign institutional investors. The few studies focusing on herding at the industry level as opposed to the stock level, support previous findings that this behavior is present among institutional investors at the industry level in the United States (Choi and Sias, 2009), Taiwan (Chen et al., 2013) and Portugal (Holmes et al., 2013).

Beyond detecting herding, several of the aforementioned studies venture one step further and attempt to explain potential determinants of the behavior or the effects thereof. As mentioned above, several studies attribute herding mostly to characteristics, specifically momentum based strategies (Grinblatt et al., 1995). More recently, Sias (2004) finds that while momentum trading does explain a part of the correlated behavior, it is a relatively small amount. By separating the analysis by the size of the stocks considered, the author concludes that informational cascades play a strong role in herding behavior, relating the empirical results to the intentional strand of herding.

Work by Gavriilidis et al. (2013) and Holmes et al. (2013) address the intent to herd, establishing the link between previous empirical work and explanatory models for the observed behavior. At a stock level in Portuguese markets, herding is found to be the result of intent, with the primary explanation being window dressing by managers (Holmes et al., 2013). At the industry level, lagged under-performance, measured by negative returns in the market and/or industry, and uncertainty, measured by market volatility, are significantly related to herding behavior, consistent with the hypothesis derived from theories of principal agent and informational cascade based herd behavior (Bikhchandani et al., 1992; Scharfstein and Stein, 1990). Choi and Sias (2009) find mixed evidence for reputationally induced industry level herding, and consistent results with informational cascades, characteristics herding and investigative herding (Sias, 2004). All of these results, including both stock level and industry level herding determinants, apply to the Lakonishok et al. (1992) or Sias (2004) herding measures, which only allow for testing of various determinant hypothesis through market level variables or through specific subsets of mutual funds.

In more recent research, Jiang and Verardo (2018) develop a new measure resulting in a fund level measure of herding tendency. As part of their analysis linking fund manager skill to this newly developed measure, they study the relationship between several determinants of herding, by directly applying micro level data in the analysis. They find that less skilled and less active funds tend to herd more, supporting theories on herding based on skill (Graham, 1999) and to some extent on informational cascades (Banerjee, 1992; Bikhchandani et al., 1992), if one assumes signal noise is implicitly linked to the investors' skill to process information.

2.3 Hypothesis

Based on these theoretical drivers of herding, the next natural step is to test some hypothesis presented in the models. While some previous studies make inroads in this regard, the focus has still been very much on detecting herd behavior and analyzing the effects in a future-looking manner. Most work researching drivers of herding have not been able to leverage the rich data available on funds which can be used to proxy the theoretical drivers of herds. As such, this leaves a gap in the literature, namely the underlying link between herding tendency and the drivers of such behavior (Spyrou, 2013), which can be addressed by analyzing the decisions of mutual funds trading decisions at the individual level.

It is important to note that drivers of herding are not necessarily mutually exclusive and may all contribute to some degree to the herding tendencies of funds. Of course, this simultaneous influence also makes the decomposition of detected herding into the various components related to the drivers difficult, as these are ultimately non-observable quantities. Nonetheless, this thesis attempts to do so by proposing a method which takes advantage of the rich availability of data on mutual fund and market characteristics which can be linked to the various herding theories. The methods employed are strongly related to those developed by Jiang and Verardo (2018), but differ in that analysis is conducted on industry based herding and that potential determinants of herding are based on the theories on herd behavior, instead of previously established empirical relationships.

Based on the previous findings within herding literature, the remainder of the thesis is structured to test the following hypothesis regarding herd behavior:

- **H1**: Investors herd more in times of uncertainty
- **H2**: Skilled investors are less likely to herd
- **H3**: Investors with high reputation are more likely to exhibit herd behavior

Uncertainty regarding the future can be be similar to a situation in which the signal an investor receives regarding prospects is relatively noisy. Models of informational cascades predict that lower signal accuracy increases the likelihood of a cascade occurring (Banerjee, 1992). Thus, this train of thought can be extended and I hypothesize that noisier signals, proxying uncertainty, will increase the likelihood of herd behavior among investors (Kremer and Nautz, 2013). This can be tested by using a measure of market wide volatility as a proxy for uncertainty and analyzing the relationship between volatility and herding behavior.

The second and third hypothesis follow directly from Graham (1999), as his theoretical model predicts these outcomes in a simple closed form, as well as from Scharfstein and Stein (1990). Lagged and contemporaneous performance can be seen as proxies for skill and will be employed as one of the explanatory variables which determine fund herding. If a fund is skilled, the models predict a lower likelihood of them following the crowd. Fund flows can be used as a proxy for reputation (Choi and Sias, 2009; Sias, 2004), with flows likely to be experienced as a result of a change in reputation. The models of Graham (1999) and Scharfstein and Stein (1990) predict there to be an increased likelihood of herding with a higher reputation. As such, a negative relationship between the proxies for manager skill and herding tendency is expected. Assuming that a higher reputation can be proxied by a larger relative inflow, the expectation is a positive relationship between this proxy and the fund herding measure.

3 Data

3.1 Mutual Fund Data

The data used for the analysis conducted in this thesis is sourced from multiple databases. Information on quarterly mutual fund holdings is taken from the Thomson Reuters Mutual Fund and Investment Company Common Stock Holdings Databases (henceforth TR), available from the Wharton Research Data Services (WRDS) of the University of Pennsylvania. For every fund in every quarter, the database provides information on the holdings of the fund by identifying each holding by stock name and identifier (CUSIP), the number of split-adjusted shares held, the price of each share at the time of reporting and the change in the number of split-adjusted shares held compared to the previous reporting period as well as the industry to which the company held belongs. This data set is primarily used to identify the herding behavior of funds.

Mutual fund characteristics other than holdings, including the age of the fund in question, the turnover ratio of a fund, the fee structure of a fund, and other non-holdings information is derived from the Center for Research in Security Prices

Survivor-bias-free US Mutual Fund Database (henceforth CRSP), also available from WRDS. This database also provides identifiers on fund characteristic which will be employed in the subsequent data cleaning process.

The full universe of mutual funds available from TR is downloaded for the period between the first quarter of 2010 and the second quarter of 2018, inclusive. This yields a sample period of 34 adjacent quarter and initially, the universe consists of 40430 unique mutual funds. I proceed with the following methodology to eliminate funds not relevant to the analysis.

As in Jiang and Verardo (2018), the focus of this thesis is to capture herding among US equity funds. I follow and first eliminate all mutual funds classified by Thomson Reuters as "international", "balanced", "bond", "municipal bonds", "preferred" and "unclassified" mutual funds as well as funds registered outside of the US. Funds for which there is no data on investment objectives are also removed from the data-set. This initially leaves 1258 mutual funds remaining in the Thomson Reuters database.

For finer selection, note that later analysis requires data from the CRSP database. Therefore, any fund remaining which cannot be identified in the CRSP database is also dropped. TR and CRSP databases use different identifiers for mutual funds in their respective data sets, an issue which is typically bypassed by using the database MutualFundLinks (MFLinks), which provides a correspondence between the identifiers in the TR and CRSP sets. However, this data set is not available through the university and makes the linking process considerably more difficult. As an alternative solution I use the names of the funds in the TR set and match them to the corresponding fund in the CRSP database, noting down the respective identifiers in the process. The imperfect nature of this method leads to the non-identification of a number of funds, which must be dropped from the TR set, leaving 997 mutual funds with data available from both sources. Finally, as the methodology employed, which is explained in section (4.1) requires a fund to be present in at least two subsequent quarters, two funds are dropped as they only appear in the data set once, leaving a total of 995 funds.

Table (1) provides descriptive statistics of these mutual funds. Of the 995 funds, most funds are present in the data set for the majority of the quarters under consideration, with the average occurrence rate being at 30.5 out of 34 quarters and over half of the funds appearing in all quarters. Panel A of table (1) shows some statistics regarding this occurrence rate.

Table 1: Descriptive Statistics of the Mutual Funds
Panel A shows the number of quarters funds appear in the constructed dataset.
Panel B describes the holdings characteristics of the funds. Stocks Traded and Stocks Held refer to the number of Stocks traded and the number of different stocks held each quater. Industry Trades refers to the dollar value change in the holding of an industry during a quarter. Industries Held refers to the number of industries held at the end of a quarter.
Panel C describes the funds characteristics. log(TNA) refers to the natural logarithm of the funds assets at the end of quarter, computed as in equation (3). Fund age is the age of the fund during its last occurrence within the dataset. Flow refers to the percentage increase in fund assets after controlling for quarterly returns, computed as specified in equation (1). Expense refers to the expense fee of a fund. Turnover refers to the quarterly turnover ratio of assets. Return measures the quarterly return of a fund, computed fro the monthly returns. Any missing values have been omitted.

(a) Panel A: Fund Observations (34 quarters from 2010Q1 to 2018Q2)

	Mean	Min	Median	Max	n
All TR Funds	30.5	3	34	34	995

(b) Panel B: Fund Holdings and Trading Statistics

	Mean	Min	25^{th}%	Median	75^{th}%	Max
Stocks Traded	78.28	0	15	39	78	3390
Stocks Held	108.56	1	34	59	98	3517
Industry Traded	12.59	0	4	11	20	37
Industry Held	16.71	1	8	18	24	37
Industry Trades (Mio.$)	4.129	-5570.28	-0.60	0.13	2.76	5259.49

(c) Panel C: Fund Characteristics

	Mean	5^{th}%	25^{th}%	Median	75^{th}%	95^{th}%
log(TA)	20.00	16.59	18.72	20.112	21.35	23.05
Fund Age	29.37	14	18.5	22.5	29.0	56.5
Flow	-0.60%	-19.64%	-6.89%	-1.72%	2.81%	18.22%
Expense	2.87%	0.38%	0.75%	0.96%	1.28%	2.05%
Turnover	78.04%	7%	25%	48%	84%	174%
Return	1.16%	-11.92%	-0.37%	3.33%	7.19%	13.45%

Panel B shows the trading activity of the average fund. In an average fund quarter, 78.28 stocks and 12.59 industries are traded. Though the range is very large, the average trade of an industry is of positive value worth 4.129 million $, and the majority of these trades are in absolute value in the single digit million $ range. On average, not all stocks or industries that are held by a fund are traded in any given quarter and there are some funds which do not trade in a given quarter.

The final panel C in table (1) shows various fund characteristics derived from CRSP data. Assets, calculated as defined in equation (3) in section (4.1), are on average valued at $e^{20.00} = 485$ million $. The first row of panel C displays various percentiles of fund quarter total holdings as defined in this manner. Funds seem at first glance to be smaller than in other related studies, however this is due to the fact that assets, as defined in equation (3), only computes the total value of a funds quarterly stock holding portfolio and neglects any other assets it may carry. Fund age refers to the age of a fund at its last appearance in the data set, which for most funds is at the point in time of Q2 2018. This computed by subtracting the date of first offering from the date of last occurrence and is rounded to the nearest quarter and takes on an average value of 29.37 years.

Expense and Turnover refer to percentages of total net assets used for fees and that are turned over for trading respectively, relative to a 12 month average of total net assets. These are given directly by CRSP, and further details can be found by consulting the relevant documentation (CRSP, 2014). Fees are heavily right-skewed, with the majority being under 2%. Turnover is more symmetrically distributed and averages 78.04% over the fund quarters.

Return refers to the quarterly return of a mutual fund and is constructed from the monthly returns which are given by CRSP. The average return is positive and the distribution is relatively symmetrical around the positive range.

Flow refers to the in- or outflows of capital from a fund and is computed from CRSP data as it is not directly given. From the given data, this variable was computed as

$$flow_t^i = \frac{Assets_t^i - (1 + r_t^i) \times Assets_{t-1}^i}{Assets_{t-1}^i} \tag{1}$$

This yields an imperfect measure of flow, as only quarterly snapshots of holdings are considered and the value change of during-quarter flows are not accounted for. Nonetheless, it serves as a suitable approximation of the flow absent dedicated data to map this quantity. The values computed seem statistically similar to those in Jiang and Verardo (2018), with the largest difference being the sign of the mean value. The inner 90 percentiles span from roughly -20% to 20%, leaving the impression of a symmetrically distributed quantity.

3.2 Industry and Index Data

Data needed on various returns of industries and indices for the control variables as well the proxy for volatility are downloaded from Factset and From the Fama and French 49 industry portfolios (FF49) database available from Kenneth French's website. Returns from Factset are automatically downloaded as quarterly returns, while the monthly returns from the Kenneth French database are used to compute the quarterly returns.

4 Methodology

4.1 Fund Herding Measure

As introduced by Sias (2004), I measure herding as an inter-temporal correlation between changes funds holdings. However, the Sias (2004) measure aggregates all funds every period and cannot be used to develop a statistic measuring the herding tendency of an individual fund every period, despite the fact that it is intuitive to think that each fund may exhibit a different propensity to herd. Jiang and Verardo (2018) address this and develop an inter-temporal fund level measure by regressing the trades of a mutual fund on those of the other funds over each security in every period. The other funds in this context can be though of as the crowd which the fund being observed is thought to herd towards. From these regressions, they extract the partial regression coefficient between a funds trades and its counterparts lagged trades. This method develops a time series of herding measures for each fund, which can then be analyzed, creating in essence a panel of herding tendencies.

Both Sias (2004) and Jiang and Verardo (2018) are applied to individual security herding and must be modified to be applied in this industry herding setting, as Choi and Sias (2009) do with the Sias (2004) measure and I do with with Jiang and Verardo (2018) measure in the following section. Using information on the holdings of each mutual fund and the price of each security, I compute the necessary variables for the regression. The main observed variable employed in the analysis of herd behavior is the inter-temporal change in portfolio weights of a given industry j by fund i [1],

$$\Delta_t^{i,j} = \frac{\sum\limits_{k \in K_j^i} P_{t-1}^k \times (N_t^{k,i} - N_{t-1}^{k,i})}{Assets_{t-1}^i} \tag{2}$$

whereby the summation occurs over the set of individual stocks that fund i holds belonging to industry j, denoted as K_j^i. $N_t^{k,i}$ denotes the number of split adjusted shares held by i of stock k at time t. The difference between these two values at two points in times multiplied by the price denotes the amount by which fund i has actively changed its holding of this security in terms of dollars and corrects for any biases introduced by neglecting to include the passive change in holding due to return on the security (see Choi and Sias (2009)). By summing these changes in holding values over the securities of a given industry, the change in value of the industry j is obtained. Finally, dividing over the sum of all assets held at period t by fund i yields the relative change in the industry holding for each fund and serves as a form of scaling to make the values comparable across funds. Intuitively, the security portfolio weight changed is aggregated to an industry portfolio weight change by weighting each of the changes by the price of the stock. The total assets held by a fund at time t is computed by summing the product between the number of shares held and the price of the security at time t across the set of all securities held by fund i,

$$Assets_t^i = \sum_{k \in K^i} P_t^k \times N_t^{k,i} \tag{3}$$

As herding is analyzed here as an inter-temporal dependence of holding movements across funds, a variable capturing the previous periods changes in industry holdings of all other fund is employed as an explanatory variable. This means that the regression consists of a linear relationship between the observed change in industry weight of fund i and the average change in industry weight of all funds *excluding* fund i in the previous period. Specifically, denoting as F the set of mutual funds in the data set, let $F^i = F \backslash i$ be the crowd for fund i, or the set of all funds excluding

[1]Portfolio weights are also used by Jiang and Verardo (2018) in their internet appendix as a robustness measure, but their main results employ a measure based on the percentage change of the number of stocks held. This is unsuited for industry based herding, as in this case the percentage change in the number of stocks of an industry held does not accurately describe the change in an industry holding, especially if the prices of the stocks within an industry vary considerably.

fund i. The explanatory variable corresponding to $\Delta_t^{i,j}$ is defined to be

$$Change_{t-1}^{i,j} = \frac{\sum\limits_{s \in F^i} \Delta_{t-1}^{s,j} \times Assets_{t-1}^{s}}{\sum\limits_{s \in F^i} Assets_{t-1}^{s}} \tag{4}$$

This measures the asset weighted average change in portfolio weight of industry j of the crowd, with the weight being based on the assets held by a fund at the end of the previous quarter.

Finally, as established in previous works, momentum may play a factor in investment decisions and in fact explain a large factor of correlated trade behavior (Grinblatt et al., 1995). This may also hold true for industry herding and therefore a control term for this must be included as a conditioning variable, hence the calculation of a partial regression coefficient. This is done in order to separate the spurious herding due to momentum trading strategies from the residual correlation which may be due to intentional herding. In order to proxy the momentum of an industry, I use the three month cumulative lagged return of the industry. For each of the 37 industries $j = \{1, 2, ...37\}$ identified in the TR holdings database, a suitable proxy to capture momentum in the defined industry is selected. Data on industry returns used for this factor are derived from the Fama BS French 49 industry portfolio returns (FF49), if the industry in question matches that in FF49. For the 14 industries without a direct match, a suitable index tracking the relevant industry is selected and data on quarterly returns are downloaded from FactSet. An overview of the industries as well as the proxies used can be found in table (8) in the Appendix.

In order to extract the fund herding measure while controlling for the effect of momentum, the following regression is run over all industries for every fund quarter:

$$\Delta_t^{i,j} = \alpha_t^i + \beta_t^i Change_{t-1}^{i,j} + \gamma_t^i mom_{t-1}^{j} + \epsilon_t^i \tag{5}$$

The coefficient β_t^i captures the inter-temporal relationship between fund i's trading decisions in period t and those of the other funds in the set in the previous period, controlling for momentum investing strategies. This yields the partial effect between a funds trades and its peers' past trades, controlling for the momentum effect over the previous which will henceforth be interpreted as the herding measure for fund i in period t.

4.2 Variation in herding

After determining the partial regression coefficients of each funds trades with those of the other fund previous trades for every quarter, the analysis consists of linking the theoretical explanations of herding with the observed actions of mutual funds. To test hypothesis 1, for each period, the cross-sectional average of the fund herding measure is computed. This is then regressed onto the price of the volatility index, across all periods of time, while controlling for the market return following Holmes et al. (2013), as well as the lagged market return and the number of funds active in the data-set during the period. These control are included in part due to the results of Gavriilidis et al. (2013) who find herding to be significant in times of negative market returns, and as returns and volatility may be correlated (i.e. the leverage effect, Bae et al., 2007), excluding this control term may lead to biased estimates. Similarly, the number of firms trading may contribute to market volatility, which is in turn thought to be a constituent of herding as stated in hypothesis 1.

Due to the nature of the VIX index, its price level does not contain a unit root, which has been tested for in past works (Fernandes et al., 2014), subsiding the fear of a spurious regression. Using returns on this index would not truly represent a proxy for volatility and therefore, this is a desirable property given the hypothesis being tested. With all of this in mind, the following regression is run:

$$\overline{\beta_t} = \alpha_0 + \alpha_1 VIX_t + \alpha_2 MR_t + \alpha_3 MR_{t-1} + \alpha_4 N_t + \epsilon_t \tag{6}$$

In line with the hypothesis, the estimate of $\hat{\alpha}_1$ is expected to be positive, as less informative signals, proxied by a more volatile market in turn indicated by a high VIX price level, are expected to be associated with a higher likelihood of herding on average (Banerjee, 1992). As much of the information available in the panel is suppressed when using only the time series of cross-sectional averages, a pooled regression is also run over all betas estimated:

$$\beta_t^i = \alpha_0 + \alpha_1 VIX_t + \alpha_2 MR_t + \alpha_3 MR_{t-1} + \alpha_4 N_t + \epsilon_t^i \tag{7}$$

For hypothesis 2 and 3, fund herding measures β_t^i from equation (5) are collected for every fund in every period available and matched to the data frame of theoretically motivated explanatory variables. Explanatory variables include the age, the fund expenses, the turnover ratio, the size, the flow, the return and the lagged return of fund i at period t. The first five variables serve as a benchmark to compare the signs to the results obtained in Jiang and Verardo (2018), and expenses as well as flows are interpreted as a reputation proxy of the fund in question. Contemporaneous and lagged fund returns are taken as the proxy to measure the skill level of a fund. In order to replicate the results of Jiang and Verardo (2018) applied to the industry level herding measure, I run a Fama and MacBeth (1973) regression over the given variables while standardizing the regressors to have mean 0 and a standard deviation of 1 each period.

5 Results

5.1 Fund Herding Tendency

Following the steps outlined in section (4.1), the variables necessary for the regression in equation (5) are computed for each fund quarter. If a fund has in any quarter a $\Delta_t^{i,j}$ with an absolute value greater than 0.5, it is thought of as an outlier and removed for that quarter. This is due to the fact that such a value would imply that a fund traded at-least 50% of its assets into or out of a certain industry, which is very large and may lead to the estimated betas blowing up do to single observations outside of the norm. While the selection of 50% is rather arbitrary, it seems to a good compromise to balance loss of information and economically sensible analysis, as out of the 29353 fund quarters available, only 221 fund quarters are dropped.

The variables $\Delta_t^{i,j}$ and $\text{Change}_t^{i,j}$ are summarized in table (2), with panel A depicting the statistics across the entire panel of trades in all fund quarters and panel B showing only the trade values which are unequal to 0. Many of the industries are not traded within any given industry quarter, as can be seen in row 3 of panel B in table (1). Therefore, the statistics in panel A understate the scale of the trades that take place and looking at both tables yields a better picture of the data. The first noteworthy fact is that at least 50% of the observed $\Delta_t^{i,j}$s are 0% changes in portfolio weights. In effect, these actions of not trading may constitute passive herding as defined in (Spyrou, 2013, p. 190) and ignoring them in the analysis, as previous papers tend to do, will overstate trading activity and overestimate herding behavior. Looking at the statistics of trade behavior omitting the non-trades, the mean portfolio weight change increases almost threefold from 0.125% to 0.327%. The trades of the crowd, that is the $\text{Change}_t^{i,j}$s conditioned on $\Delta_t^{i,j} \neq 0$, also increase in scale, though not proportionally to the increase in the dependant variables. This

Table 2: Summary statistics of the computed variables $\Delta_t^{i,j}$ and $Change_t^{i,j}$, which are calculated as defined in equations (2) and (4) in sections (4.1) respectively. Statistics are computed across the panel, after having dropped the fund quarters during which atleast one of the variables had an absolute value greater than 0.5.
Panel A presents the statistics across all fund quarter industry holding trades. Panel B displays the same summary statistics conditioned on $\Delta_t^{i,j} \neq 0$ and thereby shows the characteristics of the variables conditioned on a trade having taken place by a fund during the quarter in an industry.

(a) Panel A: Summary Statistics including zero values

	Mean	Std. Dev.	5^{th} Pctl	25^{th}Pctl	Median	75^{th}Pctl	95^{th}Pctl
$\Delta_t^{i,j}$	0.125%	1.353%	-0.436%	0.00%	0.00%	0.00%	1.117%
$Change_t^{i,j}$	0.066%	0.277%	-0.021%	0.005%	0.026%	0.081%	0.388%

(b) Panel B: Summary Statistics excluding entries for which trading did not occur

	Mean	Std. Dev.	5^{th} Pctl	25^{th}Pctl	Median	75^{th}Pctl	95^{th}Pctl
$\Delta_t^{i,j}$	0.327%	2.170%	-1.150%	-0.132%	0.034%	0.476%	2.645%
$Change_t^{i,j}$	0.109%	0.355%	-0.028%	0.013%	0.052%	0.164%	0.532%

implies that an analysis using only the non-zero trades will likely overstate the degree of herding compared to the analysis incorporating the entire set of all trades and non-trades.

The procedure outlined in equation (5) results in 29132 regressions and consequently, estimated herding coefficients. Tables (3) and (4) summarizes the preliminary results regarding these estimates, across the panel and for each cross-section respectively. Figure (5.1) is the figure corresponding to table (4) and graphically shows the cross-sectional statistics of the results over time. Of the 29132 estimated coefficients, 11898 or roughly 40.8% are statistically significant at the 5% level. As Jiang and Verardo (2018) do not report such a statistic, not comparison can be drawn to past research in this regard.

Keeping the units used in this regression un-standardized, β_t^i can be interpreted as the %-change in portfolio weight by fund i for a 1% change in the average portfolio weight of all other funds in the previous period. On average across the panel, a 1% increase in portfolio weight by other funds in the previous quarter is associated with a 1.053% increase by a fund and the standard deviation of this measure is 4.137%. Over half of the estimated coefficients are positive and the right tail of the distribution of coefficients is larger, indicating that there is to some degree a larger

Table 3: This table presents the results of running the regression as defined in equation (5).
Panel A presents summary statistics of the estimated coefficients β_t^i as defined in equation (5) across the entire panel.
Panel B presents cross sectional statistics of the estimated coefficients β_t^i as defined in equation (5). Coefficients are averaged across all quarters for each fund, creating a fund average herding measure. These average herding measures are then used to compute the displayed statistics.

(a) Panel A: Summary Statistics of the panel

	Mean	Std. Dev.	5^{th} Pctl	25^{th}Pctl	Median	75^{th}Pctl	95^{th}Pctl
β_t^i	1.053	4.137	-2.146	-0.192	0.292	1.475	6.469

(b) Panel B: Summary Statistics for the cross-section (average over quarters)

	Mean	Std. Dev.	5^{th} Pctl	25^{th}Pctl	Median	75^{th}Pctl	95^{th}Pctl
$\overline{\beta^i}$	1.119	1.852	-0.478	0.289	0.807	1.52	3.545

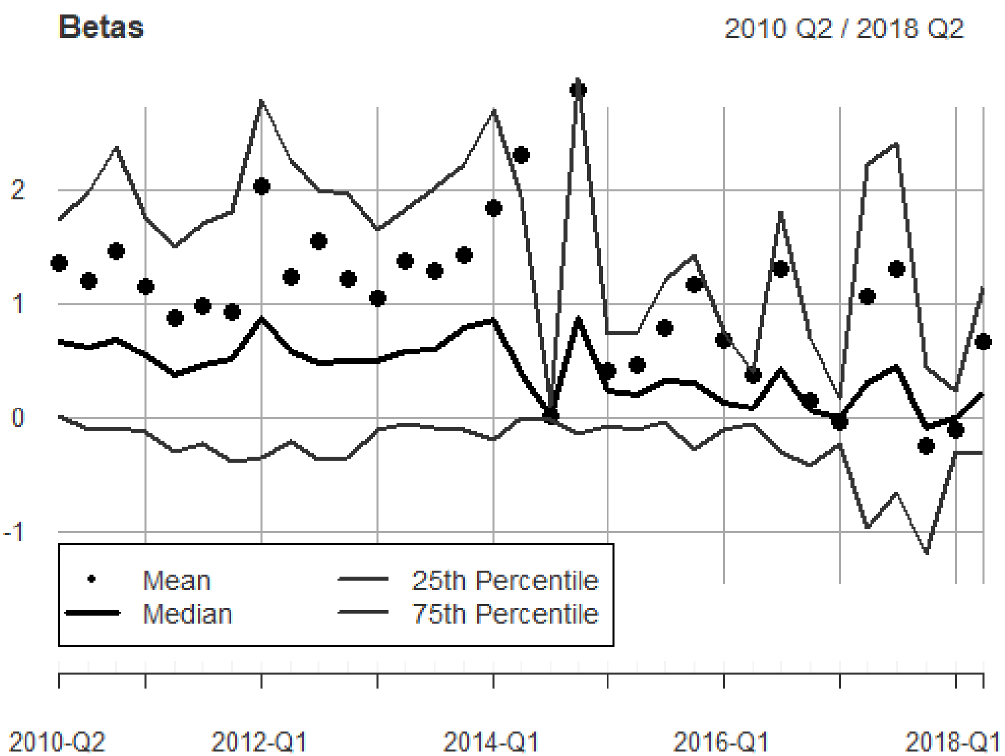

Figure 5.1: Development of betas in time. Each period, the mean, median, 25th and 75th percentiles of the individual fund herding tendencies β_t^i is computed and plotted against time. The values can be found in table (4)

Table 4: Cross Sectional statistics for every period for the estimated coefficients β_t^iEvery period, the mean, standard deviation and displayed percentiles for the distribution of fund herding tendencies β_t^i is computed and displayed.

	Mean	Std. Dev.	5^{th} Pctl	25^{th}Pctl	Median	75^{th}Pctl	95^{th}Pctl
2010 Q2	1.357	2.950	−1.103	0.015	0.668	1.744	5.823
2010 Q3	1.194	2.821	−1.445	−0.113	0.625	1.981	5.564
2010 Q4	1.465	3.608	−1.590	−0.097	0.678	2.376	7.324
2011 Q1	1.151	3.316	−1.560	−0.119	0.547	1.745	5.982
2011 Q2	0.878	2.814	−1.983	−0.292	0.379	1.486	5.533
2011 Q3	0.978	3.163	−1.946	−0.221	0.461	1.698	5.312
2011 Q4	0.918	3.937	−2.999	−0.385	0.519	1.811	6.237
2012 Q1	2.032	6.917	−3.089	−0.355	0.877	2.790	10.253
2012 Q2	1.232	5.070	−3.104	−0.202	0.584	2.247	7.344
2012 Q3	1.541	5.827	−2.245	−0.367	0.486	1.985	8.314
2012 Q4	1.216	4.894	−2.670	−0.346	0.493	1.960	7.600
2013 Q1	1.050	3.357	−1.790	−0.111	0.500	1.643	5.235
2013 Q2	1.372	3.938	−1.940	−0.058	0.582	1.823	6.785
2013 Q3	1.293	3.826	−1.883	−0.084	0.603	2.003	6.972
2013 Q4	1.426	3.888	−2.127	−0.107	0.781	2.213	7.402
2014 Q1	1.838	5.229	−2.688	−0.193	0.851	2.696	9.549
2014 Q2	2.299	4.553	−0.685	−0.0002	0.388	1.946	12.977
2014 Q3	0.023	0.381	−0.389	−0.023	0.023	0.099	0.433
2014 Q4	2.868	9.111	−2.672	−0.146	0.880	2.969	16.286
2015 Q1	0.416	1.305	−0.788	−0.079	0.246	0.745	2.109
2015 Q2	0.461	1.550	−0.993	−0.098	0.212	0.732	2.578
2015 Q3	0.795	2.279	−0.860	−0.032	0.323	1.196	3.846
2015 Q4	1.169	4.146	−2.067	−0.275	0.305	1.432	7.032
2016 Q1	0.686	2.370	−1.061	−0.107	0.129	0.756	4.987
2016 Q2	0.374	1.553	−0.634	−0.054	0.077	0.392	2.586
2016 Q3	1.301	4.390	−2.899	−0.293	0.428	1.796	8.371
2016 Q4	0.152	2.374	−2.546	−0.418	0.063	0.695	3.055
2017 Q1	−0.036	1.579	−1.836	−0.228	0.001	0.173	1.277
2017 Q2	1.068	6.398	−5.113	−0.962	0.309	2.209	9.572
2017 Q3	1.297	5.414	−3.969	−0.653	0.453	2.402	9.662
2017 Q4	−0.242	2.914	−4.426	−1.193	−0.084	0.445	3.261
2018 Q1	−0.105	0.935	−1.546	−0.302	0	0.235	0.866
2018 Q2	0.675	2.951	−1.900	−0.319	0.217	1.143	4.832

tendency for positively correlated trades.

As previous studies such as Sias (2004) or Jiang and Verardo (2018) standardize their variables for each regression, no direct comparison can be drawn between the results. The argument for not standardizing in each fund quarter is the loss of understanding of the meaning of both the variables as well as the coefficients. The intuitive understanding stated above is lost by standardization, though the benefit is better comparability of the coefficients across time and funds. As an informal form of comparison, consider that in Jiang and Verardo (2018) the mean fund herding tendency β_t^i of 2.30 implies that for a one standard deviation increase in the dependant variable depicting the crowds lagged trades, a fund follows that trade by 2.3 standard deviation. From table (2), it is easy to see that the dependant variable has a standard deviation of 0.277%. This mean that very roughly, for the specification employed in this thesis, a one standard deviation increase in the independent variable leads to roughly a $1.053 \times 0.277\% = 0.292\%$ increase in the dependant variable β_t^i. This translates to a $0.292\% \div 1.353\% = 0.215$ standard deviation increase in the dependant variable. Of course, this value should be viewed with a high level of skepticism due to the approximate nature of its computation, however it is also worth highlighting that it is much smaller in magnitude than the value reported by Jiang and Verardo (2018).

Panel B of table (3) indicate that the cross-sectional of the time-series average coefficients for each fund varies across time with a standard deviation of 1.852%. When looking at the average herding tendency of a fund over its appearances, the skew towards positive values is even more pronounced than in panel A. This may serve as an indication that a fund systematically tends to have a collection of positive (or negative) herding tendency coefficients, instead of a balanced mix of the two. Looking at the time series of cross-sections, there is a varying degree of variance across each cross-section, with Q3 2014 displaying the lowest variation, right before Q4 2014 with the highest variation. The cross-sectional average is in all but 4 quarters larger than the cross-sectional median value and the absolute distance between median and the higher percentiles tends to be larger than that between the median and smaller percentiles. This further serves as an indicator that there is a larger inclination towards positive herding tendencies. There appears to be a light downward trend over time across all percentiles displayed, as can be seen in figure (5.1), with the median herding value especially seemingly tending towards 0 near the end of the sample period.

5.2 Determinants of Herding

5.2.1 Volatility

The regression defined in equation (6) is run across all 33 periods, using the mean values for the beta for each period, found in table (4). According to the hypothesis, the slope coefficient for the volatility is expected to be positive. Holmes et al. (2013) find this to be the case using the Sias (2004) methodology on stock level herding. As explained in section (4), these controls are included to prevent omitted variables.

Table 5: Time Series Regressions of Cross-Sectional Average of β_t^i on explanatory variables. The dependant variable is the cross-sectional average of the fund herding measure β_t^i, which can be found in column 2 of table (4). log(VIX) refers to the natural logarithm of the price level of the CBOE VIX index and serves as a volatility proxy. Market refers to the market return, proxied by the S&P500 index during the period. MarketLag refers to the return of the same index. Funds refers to the number of funds trading during the quarter. Standard errors are reported in the brackets.

	Dependent variable: Average Fund Herding $\overline{\beta}_t$			
log(VIX)	0.0277	0.0518	0.0751	0.0276
	(0.0473)	(0.0568)	(0.0640)	(0.0748)
Market Return		0.0023	0.0034	0.0007
		(0.0029)	(0.0033)	(0.0039)
MarketLag Return			0.0023	0.0011
			(0.0028)	(0.0030)
Number of Funds				0.0003
				(0.0003)
Constant	0.0522	−0.0220	−0.0967	−0.2243
	(0.1336)	(0.1647)	(0.1899)	(0.2163)
Observations	33	33	33	33
Adjusted R^2	0.0027	0.0643	0.1040	0.1364
Note:			*p<0.1; **p<0.05; ***p<0.01	

The results of the regressions are summarized in table (5). In all four specifications of regressing the cross-sectional average beta onto explanatory variables, the coefficient for the volatility proxy is insignificant. As such, evidence supporting hypothesis 1

Table 6: Pooled least squares regression of β_t^i on explanatory variables. log(VIX) refers to the natural logarithm of the price level of the CBOE VIX index and serves as a volatility proxy. Market refers to the market return, proxied by the S&P500 index during the period. MarketLag refers to the return of the same index. Funds refers to the number of funds trading during the quarter. Standard errors are reported in the brackets.

	Dependent variable: Fund Herding β_t			
log(VIX)	0.0029	0.0235^{***}	0.0403^{***}	0.0239^{***}
	(0.0037)	(0.0047)	(0.0053)	(0.0064)
Market Return		0.0377^{***}	0.0548^{***}	0.0357^{***}
		(0.0051)	(0.0058)	(0.0071)
MarketLag Return			0.0298^{***}	0.0212^{***}
			(0.0047)	(0.0050)
Number of Funds				0.0020^{***}
				(0.0004)
Constant	1.0020^{***}	0.5303^{***}	0.1019	-1.3272^{***}
	(0.0698)	(0.0947)	(0.1159)	(0.3299)
Observations	29,132	29,132	29,132	29,132
Adjusted R^2	0.00002	0.0019	0.0033	0.0040
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

relating a higher tendency to herd to the level of volatility cannot be found when considering the cross-sectional average herding tendency.

As averaging all betas across the cross-section may be forfeiting some information, a pooled least squares regression is run over all betas onto the same explanatory variables, as described in equation (7). The results are presented in table (6). In this setting, the coefficients relating to volatility are significantly positively related to the herding, though the regressions leave much of the variation unexplained, with an adjusted R squared of under 1%. All other coefficients excluding the intercept are significantly positive, as opposed to the previous case equation (6).

5.2.2 Other Determinants

The 29132 regression coefficients are gathered and data on total assets, fund age, fund flow, fund expense, fund turnover and fund return are collected from the CRSP database as defined in section (3), for which statistics can be found in Panel C of table (1). As some values are not available for fund quarters, all observations for which at least one variable is not available is dropped from the data frame. This leaves 25455 observations belonging to 967 individual funds over a maximum of 33 quarters, yielding an unbalanced panel. To replicate the methodology of Jiang and Verardo (2018), a Fama and MacBeth (1973) cross-sectional regression is run for each period, with fund herding being regressed onto the explanatory variables for each cross-section. The cross-sectional estimates are then aggregated across time to compute coefficient and error estimates.

Table (7) summarizes the results of these regressions. The first column shows the base case, which is comparable to Jiang and Verardo (2018) without the alpha of a fund as an explanatory variable. Assets at the end of a quarter, turnover and flows during a quarter are significantly correlated with the fund herding measure. These three variables continue to remain significant upon adding lagged fund return, contemporaneous fund return and both returns into the specification. Contemporaneous return is also significant in regard to the relationship with fund herding measure.

The results provide supporting evidence for hypothesis 3, as the flow of assets into a fund is significantly positively related to the dependant fund herding measure. As a proxy for reputation, the assumption is that high flow is associated with greater reputation, as such, the relation between reputation and flow is thought to be positive.

Table 7: Fama and MacBeth (1973) regressions of the fund herding measure β_t^i on fund characteristics as described in table (1). Log(Assets) refers to the natural logarithm of the funds assets as calculated in equation (3). Log(Age) refers to the natural logarithm of the age of the fund. Expenses refers to the expense ratio of the fund and is provided for directly by CRSP. Turnover refers to the percentage of assets turned over during the quarter and is also provided by CRSP. Flow refers to the return adjusted change in assets of a fund during a quarter and is calculated as defined in (1). Return refers to the quarterly return of the fund during the quarter for which the fund herding measure was computed. Lagged return refers to the return in the previous quarter. As in Jiang and Verardo (2018), all independent variables are standardized to have mean 0 and standard deviation 1each period. Standard errors are reported in the brackets.

	Dependent variable: Fund Herding β_t^i			
Log(Assets)	-0.1833^{***}	-0.1929^{***}	-0.1889^{***}	-0.1957^{***}
	(0.0541)	(0.0550)	(0.0554)	(0.0554)
Log(Age)	0.0251	0.0307	0.0261	0.0298
	(0.0225)	(0.0221)	(0.0230)	(0.0226)
Expense	-0.2733	-0.2610	-0.2626	-0.2627
	(0.1708)	(0.1697)	(0.1662)	(0.1662)
Turnover	0.6783^{***}	0.6673^{***}	0.6631^{***}	0.6618^{***}
	(0.1549)	(0.1544)	(0.1508)	(0.1509)
Flow	0.7984^{***}	0.7964^{***}	0.8023^{***}	0.8022^{***}
	(0.1272)	(0.1269)	(0.1273)	(0.1276)
Return		0.1307^{***}		0.1355^{***}
		(0.0379)		(0.0470)
Lag Return			0.0087	-0.0259
			(0.0394)	(0.0484)
Observations	25,455	25,455	25,455	25,455
Adjusted R^2	0.0911	0.0938	0.0930	0.0963
Note:			*p<0.1; **p<0.05; ***p<0.01	

Expense, the second proxy for reputation, does not exhibit any significance.

Lagged returns, the proxy for skill which is investigated through hypothesis 2, is surprisingly not significantly related to fund herding. Contemporaneous return on the other hand, is significantly positively related to the fund herding measure. This goes counter the argument made, as the results indicate that higher skilled, proxied as contemporaneous returns, are related to a higher degree of herding.

6 Discussion

6.1 Fund Herding

According to the efficient market hypothesis, investors should buy and hold the market portfolio with a certain fraction of their wealth, subject to risk preferences. As assets as calculated in equation (3) counts only the stock portion of assets held by a fund, $\Delta_t^{i,j}$ captures the portfolio weight of the stock portfolio. Therefore, any changes in this measure should only occur due to a change in risk preferences or in asset characteristics. Any changes in asset characteristics affect all investors equally and any changes to portfolio weights should be made instantly, meaning that any changes in stock portfolio weights related to this should not exhibit any inter-temporal correlation unless the change in fundamentals do. Such correlated behavior would then be counted as spurious herding, as each individual would act only in an efficient manner in accordance with the market conditions. As patterns in the changes of fundamentals would imply a possibility to take advantage of these patterns, there shouldn't be any predictable pattern in the change in fundamentals in an efficient market. In addition, when assuming that mutual fund managers have fixed risk preferences over their lifetimes, then there should be no expected correlation between $\Delta_t^{i,j}$ and $\text{Change}_{t-1}^{i,j}$ over a large number of market participants and periods. This means that, on average, spurious herding should average to 0 when observing many periods. The fact that the average herding coefficient β_t^i is larger than 0 and that over 40% of the individual coefficients are statistically significantly different from 0 indicates that mutual fund managers exhibit some herding behavior into and out of industries.

As mentioned in the previous section, fund herding at the industry level seems smaller in scale compared to herding at the stock level as in Jiang and Verardo (2018). This discrepancy can be due to two reasons, aside from sampling and slight methodological differences. The first is related to what Spyrou (2013) names passive herding, which was briefly mentioned in section (5.1). The methodology employed in this thesis corrects for potential bias in previously reported herding measures by taking into account the fact that a fund *not* trading when their peers have done so in the past can be seen as a decision to not follow the crowd. Upward bias can occur when one removes all non trades and these non-trades are associated with positive lagged trades by the crowd of other funds, while downward bias can occur if non-trades associated with lagged negative trades are not accounted for.

Due to the structure of most data sets and the nature in which previous statistics were computed (Sias, 2004), only trades themselves are considered. Considering all actions, including the non-trades of securities by funds, is nearly impossible at the security level due to the vast amount of securities traded. However, in practice, not trading a security despite the crowds decision to do so in the past may in fact represent a decision by a fund to actively not herd on its peers decisions. For herding at an industry level, it becomes much easier to take into account such non trading actions, as the set of all tradable assets is limited to the number of industries defined. In this case, there are an unchanged number of 37 industries for each fund quarter and non-trading can be represented by a 0 value for $\Delta_t^{i,j}$. This allows for a more complete picture of a funds trading activity and removes bias induced by not accounting for non-trading. The results obtained here may support the existence of an upward bias in the measure of Jiang and Verardo (2018).

The second reason for the smaller results are connected with results from Choi and Sias (2009). They find that stock level herding explains a significant portion of industry level herding. As such, there may be some substitution occurring among stocks within industries arising from the stock level herding. However, they note that "[...] although institutional investors herding into individual stocks contribute to institutional industry herding, industry herding is unique from stock herding"(Choi and Sias, 2009, p. 478), supporting the notion that a portion of industry level herding is independent of stock level herding.

6.2 Informational Cascades

In the initial regression, the analysis fails to find evidence relating informational uncertainty, proxied through market volatility and average herding behavior. The results of these section may at first glance seem comparable to those based on the study of herding towards the market return (Christie and Huang, 1995; Chang et al., 2000). Both of these studies employ methodology to test the hypothesis that herding is more likely to be present in times of extreme price movements or high volatility. They fail to find such a relationship and conclude that market participants in the US do not exhibit herd behaviour. While these studies define herding as all market participants herding towards the market, herding is thought to be the inter-temporal mimicking of investment decisions and there is no necessary market towards which is herded. Furthermore, being more likely to herd in times of extreme price movements does not imply that there is no herding in times of more subdued market movements. Therefore, it would be mistaken to believe that these results support the findings of Christie and Huang (1995) and Chang et al. (2000) or other market based herding literature, as the research questions themselves are very different.

When analyzing the entire panel of fund herding measures, some weak evidence supporting the hypothesis stated emerges. This is linked to results from Gavriilidis et al. (2013), who find that herding into industries is indeed more prevalent during quarters of high market volatility in Spain. They attribute this finding to the link between volatility and complexity in signals. Choi and Sias (2009) also find weak evidence supporting the notion of informational cascades leading to industry level herding, although they rely on a different proxy to measure this link.

6.3 Skill and Reputation

The analysis between fund herding behavior and skill as well as reputation indicate that flow, a proxy for reputation, as well as fund return, a proxy for funds skill, are significantly related to a tendency to herd. The significantly positive coefficient relating flow to a funds herding tendency presents evidence in favor of hypothesis 3 and in turn for the predictions made by reputational herding models (Scharfstein and Stein, 1990). Taken together with the findings of Choi and Sias (2009), it is justified to conclude that mutual fund managers in the US do seem to exhibit herd behavior in conjunction with concerns regarding their reputation. This also conforms with results from Graham (1999) that higher reputations increase herding likelihood among US analysts. Finally, in conjunction with findings of a positive link between reputation and herding tendency among fund managers in Portugal Holmes et al. (2013), these results provide evidence for the robustness of herd behavior arising as a result of reputational concerns in a wide variety of market settings.

Contrary to the results on reputation, those on skill do not conform with the hypothesis stated, instead resulting in an outright rejection of the stated hypothesis.

The positive coefficient between herding tendency and fund return indicate that, a fund thought to have a higher skill level as proxied by return is more likely to exhibit herd behavior. The fact that contemporaneous return is positively related to herding tendency brings an additional layer of complexity to the analysis, as it is not quite clear whether this implies that skilled managers herd or herding leads to good performance. Regardless of the interpretation, these results are at odds with previous findings. Using the same universe of funds, Jiang and Verardo (2018) find extensive evidence in favor of hypothesis 2, while among Spanish fund managers, Gavriilidis et al. (2013) also find evidence supporting the notion that bad managers are more likely to herd on the actions of the crowd. Given the fact that differing methodologies reach the same conclusions, it may be possible that the proxy chosen for the skill level of a fund in this analysis is not a suitable one. Indeed, while it is justified to believe return and skill to be correlated, it is perfectly possible for unskilled managers to exhibit a high return and vice-versa, purely due to chance.

6.4 Limitations and Suggestions

As with any quantitative analysis, the methods applied in this thesis face a great deal of limitations, which can be improved upon in following research. In any analysis, it is highly unlikely to correctly specify a model. The inclusion of various factors to explain variations in herding behavior is however, limited both by theory and by data availability. The latter aspect is also a major decider on the selection of proxy variables. As alluded to before, the selection of proxies may have large implications on the results of the analysis. In order to better analyze the different drivers of herd behavior, a larger selection of variables, such as those concerning manager pay structures or better measures of skill could be merged with the existent data to allow for a more comprehensive study.

The main data used employs quarterly holdings information from the Thomson Reuters database, which is itself derived by sorting the 13F filing by the institutions. This in itself limits both the scope and accuracy of the analysis. First, Anderson and Brockman (2016) report that 13F filings are prone to significant reporting errors. This limits the scope to which research relying on data retrieved from this source can accurately depict the actions of filing institutions. Second, the frequency of the data leads to the ability to only observe quarterly snapshots of mutual funds actions, without the ability to observe what occurs in between. In the context of the analysis into herd behavior, this implicitly limits the analysis to a time-frame of quarters.

Herding may in fact take place at e.g. a monthly frequency, that is institutions view the trades of other funds in the month prior and act accordingly. One month of purchasing a security based on the lagged actions of other funds, followed by a month of selling the same security also due to other funds doing so, will not show up in such a quarterly view of the actions of mutual funds. More advanced data-sets, sourced from the funds themselves, should help alleviate some of these concerns, while also addressing the first limitation concerning a wider variety of proxies.

One final aspect concerning herd behavior worth mentioning is the concept of the crowd which is being followed. In section (4.1), the crowd is defined to be the set of all mutual funds in the data-set trading in the previous quarter. This need not necessarily be true and there is a large degree of ambiguity as to who may constitute the crowd. Indeed, looking at the entire market, there cannot be herd behavior whenever the market clears, as for every buyer, there is a seller (Bikhchandani and Sharma, 2001). This should also hold true in expectation for any random subset of market participants. While mutual funds cannot be thought of as a random subset, it is possible quite possible that herding takes place in much smaller and more homogeneous subsets of mutual funds and that analyzing all mutual funds at once understates or misidentifies the nature of herding occurring. An intuitive example can be thought of by example of two funds, A and B. Assume that fund B always follows the trades of fund A, and further that all asset purchases by A and B are sourced from all other mutual funds. Then, when looking only at funds A and B, the results would indicate a strong presence of herd behavior by fund B following the crowd consisting only of mutual fund A. However, defining the crowd to be the set of all other mutual funds, including fund A, the results would indicate that fund B does not herd onto the trades of the crowd, even though by construction, fund B will always herd onto the actions of another party. In this regard, future research may benefit from limiting the analysis to a smaller set of funds. Examples for smaller subsets may be funds based in the same city (Hong et al., 2005) or funds with managers from the same university (Cohen et al., 2008), as these fund managers have been found to trade similarly, possibly due to an easier degree of information transmission. Focusing on a smaller number of funds also implies that less data is required for the analysis, which can be advantageous in addressing the previous two concerns, allowing for a more limited but accurate form of analysis.

7 Conclusion

This thesis sought to extend the literature on herd behavior in financial markets by analyzing the trading behavior of mutual fund managers in the United States. Methodology developed by (Jiang and Verardo, 2018) is used and adapted to work in an industry herding setting. During the period from 2010 to mid 2018, US mutual funds are found to exhibit some herd behavior, with the average herding coefficient, calculated as the inter-temporal regression coefficient between the change in portfolio weights of a funds holding in an industry with the average change in portfolio weights of all other funds (the crowd) for the same industry in the previous quarter, being approximately 1. This value implies that mutual funds seem to copy each others allocation decisions on average.

The analysis continues by attempting to explain the variation in herding behavior across funds and across times based on the theories of principal-agent and reputational herding (Scharfstein and Stein, 1990; Graham, 1999) as well as informational cascades (Banerjee, 1992; Bikhchandani et al., 1992). Volatility, a proxy for signal noise received by the managers, is positively correlated with the panel of herding measure, supporting the hypothesis that herding may be attributed to informational cascades. Furthermore, fund flows, which are assumed to proxy the reputation of a fund, are significantly positively related to the herding measures, which provides support for the theory of reputation based herding behavior. These two findings are in line with past literature and support the notions that industrial herding is present among US fund managers and partially driven by informational cascades and reputational concerns.

The proxies used for skill, lagged and contemporaneous fund return, indicate that the hypothesis relating higher skill to a lower herding tendency on part of a manager is incorrect. This refutes prior literature, leading to the belief that the chosen proxies may not necessarily be the best ones and stresses the need for larger, more comprehensive datasets from which one can compute better proxy variables for all three measures tested.

8 References

Anderson, A. M. and Brockman, P. (2016). Form 13f (mis) filings. *Available at SSRN 2809128*.

Bae, J., Kim, C.-J., and Nelson, C. R. (2007). Why are stock returns and volatility negatively correlated? *Journal of Empirical Finance*, 14(1):41–58.

Banerjee, A. V. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, 107(3):797–817.

Barclay, M. J. and Warner, J. B. (1993). Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics*, 34(3):281–305.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026.

Bikhchandani, S. and Sharma, S. (2001). Herd Behavior in Financial Markets. *IMF Staff Papers*, 47(3):279–310.

Blake, D., Sarno, L., and Zinna, G. (2017). The market for lemmings: The herding behavior of pension funds. *Journal of Financial Markets*, 36:17–39.

Cai, F. and Zheng, L. (2004). Institutional trading and stock returns. *Finance Research Letters*, 1(3):178–189.

Chakravarty, S. (2001). Stealth-trading: Which traders trades move stock prices? *Journal of Financial Economics*, 61(2):289–307.

Chang, E. C., Cheng, J. W., and Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, 24(10):1651–1679.

Chen, Y.-C., Wu, H.-C., and Huang, J.-J. (2017). Herd Behavior and Rational Expectations: A Test of China's Market Using Quantile Regression. *International Journal of Economics and Financial Issues*, 7(2):649–663.

Chen, Y.-F., Yang, S.-Y., and Lin, F.-L. (2013). Foreign institutional industrial herding in Taiwan stock markets . *International Journal of Emerging Markets*, 5(3):474–490.

Choi, N. and Sias, R. W. (2009). Institutional industry herding. *Journal of Financial Economics*, 94(3):469–491.

Christie, W. G. and Huang, R. D. (1995). Following the Pied Piper: Do Individual Returns Herd around the Market? *Financial Analysts Journal*, 51(4):31–37.

Cohen, L., Frazzini, A., and Malloy, C. (2008). The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy*, 116(5):951–979.

CRSP (2014). Survivor-bias-free us mutual fund guide. *Center for Research in Security Prices, Graduate School of Business, The University of Chicago.*

Devenow, A. and Welch, I. (1996). Rational herding in financial economics. *European Economic Review*, 40:603–615.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Fernandes, M., Medeiros, M. C., and Scharth, M. (2014). Modeling and predicting the cboe market volatility index. *Journal of Banking & Finance*, 40:1–10.

Gavriilidis, K., Kallinterakis, V., Pedro, M., and Ferreira, L. (2013). Institutional industry herding: Intentional or spurious? *"Journal of International Financial Markets, Institutions & Money"*, 26:192–214.

Graham, J. R. (1999). Herding among Investment Newsletters: Theory and Evidence; Herding among Investment Newsletters: Theory and Evidence. *The Journal of Finance*, LIV(1):237–268.

Grinblatt, M., Titman, S., and Wermers, R. (1995). Momentum Investment Strategies, Portfolio Performance, and Herding: A Study of Mutual Fund Behavior. *The American Economic Review*, 85(5):1088–1105.

Hirshleifer, D. and Hong Teoh, S. (2003). Herd Behaviour and Cascading in Capital Markets: a Review and Synthesis. *European Financial Management*, 9(1):25–66.

Holmes, P., Kallinterakis, V., and Ferreira, M. P. (2013). Herding in a concentrated market: A question of intent. *European Financial Management*, 19(3):497–520.

Hong, H., Kubik, J. D., and Stein, J. C. (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance*, 60(6):2801–2824.

Iihara, Y., Kato, H. K., and Tokunaga, T. (2001). Investors herding on the tokyo stock exchange. *International Review of Finance*, 2(1-2):71–98.

Indārs, E. R., Savin, A., and Lublóy, Á. (2019). Herding behaviour in an emerging market: Evidence from the moscow exchange. *Emerging Markets Review*, 38:468–487.

Jiang, H. and Verardo, M. (2018). Does Herding Behavior Reveal Skill? An Analysis of Mutual Fund Performance. *Journal of Finance*, 73(5):2229–2269.

Kremer, S. and Nautz, D. (2013). Causes and consequences of short-term institutional herding. *Journal of Banking & Finance*, 37(5):1676–1686.

Lakonishok, J., Shleifer, A., and Vishny, R. W. (1992). The impact of institutional trading on stock prices. *Journal of Financial Economics*, 32(1):23–43.

Nofsinger, J. R. and Sias, R. W. (1999). Herding and Feedback Trading by Institutional and Individual Investors. *The Journal of Finance*, 54(6):2263–2295.

Raddatz, C. and Schmukler, S. L. (2013). Deconstructing Herding: Evidence from Pension Fund Investment Behavior. *Journal of Financial Services Research*, 1.

Rajan, R. G. (2006). Has finance made the world riskier? *European Financial Management*, 12(4):499–533.

Scharfstein, D. S. and Stein, J. C. (1990). Herd Behavior and Investment. *The American Economic Review*, 80(3):465–479.

Sias, R. W. (2004). Institutional Herding. *Review of Financial Studies*, 17(1):165–206.

Sias, R. W., Starks, L. T., and Titman, S. (2001). The price impact of institutional trading. *Available at SSRN 283779*.

Spyrou, S. (2013). Herding in financial markets: a review of the literature. *Review of Behavioural Finance*, 5(2):175–194.

Voronkova, S. and Bohl, M. T. (2005). Institutional traders behavior in an emerging stock market: Empirical evidence on polish pension fund investors. *Journal of Business Finance & Accounting*, 32(7-8):1537–1560.

Welch, I. (1992). Sequential Sales, Learning, and Cascades. *The Journal of Finance*, XLVII(2).

Wermers, R. (1999). Mutual Fund Herding and the Impact on Stock Prices. *The Journal of Finance*, LIV(2):581–622.

Appendices

A Industry Mometum

Table 8: List of all industry classifications provided by Thomson Reuters in the Mutual Fund Database as well as a corresponding proxy to use for returns. If available, an industry portfolio from Fama and French's 49 industry portfolios is employed. If the choice is ambiguous or not available, a suitable index is used.

Industry Code	Thomson Reuters Industry Classification	Proxy
100	Unknown	S&P 500
101	Aerospace	Fama French 49 Industry Portfolios Aircraft (Aero)
102	Agriculture	Fama French 49 Industry Portfolios Agriculture (Agric)
103	Airlines	S&P Transportation Select Industry Index
104	Automobiles	Fama French 49 Industry Portfolios Automobiles and Trucks (Autos)
105	Banks & Savings Institutions	Fama French 49 Industry Portfolios Banking (Banks)
106	Beverages	Dynamic Food & Beverage Intellidex Index
107	Chemicals	Fama French 49 Industry Portfolios Chemicals (Chems)
108	Computer Hardware, Software & Services	Fama French 49 Industry Portfolios Computer Software (Softw)
109	Construction & Engineering	Fama French 49 Industry Portfolios Construction (Cnstr)
110	Consumer Services	Fama French 49 Industry Portfolios Personal Services (PerSv)
111	Electrical & Electronics	Fama French 49 Industry Portfolios Electrical Equipment (ElcEq)
112	Miscellaneous	S&P 500
113	Energy and Fuels	Energy Select Sector Index
114	Financial Services	MSCI US IMI Financials 25/50
115	Food & Restaurants	Fama French 49 Industry Portfolios Restaurants, Hotels, Motels (Meals)

116	Healthcare	Fama French 49 Industry Portfolios Healthcare (Hlth)
117	House Wares & Household Items	Fama French 49 Industry Portfolios Consumer Goods (Hshld)
118	Industrial Manufacturing	Industrial Select Sector SPDR Fund
119	Insurance	Fama French 49 Industry Portfolios Insurance (Insur)
120	Investment Services	Fama French 49 Industry Portfolios Trading (Fin)
121	Leisure Travel & Lodging	Dynamic Leisure & Entertainment IntellidexSM Index
122	Machinery & Equipment	Fama French 49 Industry Portfolios Machinery (Mach)
123	Media	Dynamic Media Intellidex
124	Metals & Mining	Fama French 49 Industry Portfolios Non-Metallic and Industrial Metal Mining (Mines)
125	Packaging	MSCI USA IMI Materials 25/50
126	Paper & Forest Products	S&P Global Timber & Forestry Index
127	Publishing & Printing	Dynamic Media Intellidex
128	Real Estate	Fama French 49 Industry Portfolios Real Estate (RlEst)
129	Retail & Consumer Goods	Fama French 49 Industry Portfolios Retails (Rtail)
130	Semiconductors	Fama French 49 Industry Portfolios Electronic Equipment (Chips)
131	Telecommunications	Fama French 49 Industry Portfolios Communication (Telcm)
132	Textiles & Apparel	Fama French 49 Industry Portfolios Apparel (Clths)
133	Tobacco	Fama French 49 Industry Portfolios Tobacco Products (Smoke)
134	Transportation	Fama French 49 Industry Portfolios Transportation (Trans)
135	Water-Electric-Gas Utilities	Dow Jones Utility Average

136	Waste & Environment Management	NYSE Arca Environmental Services Index

B Consent Form

Appendix C: Official Statement of Original Thesis *

By signing this statement, I hereby acknowledge the submitted thesis (hereafter mentioned as "product"), titled:

Measuring US Mutual Fund Herding in Industries

to be produced independently by me, without external help.

Wherever I paraphrase or cite literally, a reference to the original source (journal, book, report, internet, etc.) is given.

By signing this statement, I explicitly declare that I am aware of the fraud sanctions as stated in the Education and Examination Regulations (EERs) of the SBE.

Place: Bonn, Germany

Date: 14. July 2019

First and last name: Ichio Aoki

Study programme: M.Sc. Financial Economics

Course/skill: 2018-001-EMTH0001 Master's Thesis

ID number: i6201190

Signature:

* Please complete this page and add it as an appendix to your MSc Thesis.