Distribution Agreement

Eric Augustus Bai April 9, 2017

Delay Portfolio Disclosure and its Effect on Mutual Funds' Performance

by

Eric Augustus Bai

John Kim
Adviser

Economics Department

John Kim
Adviser

Junghoon Lee
Committee Member

David Zureick Brown
Committee Member
2017

Delay Portfolio Disclosure and its Effect on Mutual Funds' Performance

By

Eric Augustus Bai

John Kim
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics Department

2017

Abstract

Delay Portfolio Disclosure and its Effect on Mutual Funds' Performance

By Eric Augustus Bai

This paper investigates the effect of delaying mandatory portfolio disclosure on mutual fund's performance. Since 2004, mutual funds were required to file Form N-Q and Form N-CSR at the end of every quarter. However, funds have the option of delaying revealing their portfolio for up to sixty days. My results showed positive relationship between delay and fund return, but insignificant coefficients for the estimated model. I conclude that the market is efficient enough so as to make the delay variable negligible in determining fund return.

Delay Portfolio Disclosure and its Effect on Mutual Funds' Performance

By

Eric Augustus Bai

John Kim
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics Department

2017

Table of Contents

1. Introduction

Information in the investment industry has been the single most sought after and demanded signal among institutional investors and the public alike. For the investors, information about the companies' fundamentals and general macroeconomics trends gives a hint of where the stock price is heading in the future. For the institutional investors, capitalizing first on such information brings significant profits by exploiting mispriced securities. Likewise, the general public, who supplies a portion of their savings into various funds, demand information about their portfolio holdings and trading outlooks as a way of monitoring the funds' behavior. Under the U.S. Securities and Exchange Commission, management investment companies are required to include a "summary portfolio schedule of investments" in reports to shareholders ("Final Rule"). In a later amendment, companies would be required to "disclose its complete portfolio schedule on a quarterly basis in filings with the Commission that will be certified by the company's principal executive and financial officers and available on the Commission's Electronic Data Gathering, Analysis, and Retrieval System"("Final Rule"). In short, the requirement for investment companies to disclose complete portfolio holdings facilitates the information transfer between funds and its investors so as to make the process fluid and transparent.

The scramble for information about stock holdings, company news, and macroeconomics factors is closely tied to the Efficient Market Hypothesis, which states that all available relevant information regarding any security will be ultimately embedded into the price of that security. That is, because any new information that reflects the health of a company is quickly excavated through market research and processed through arbitrage by traders, an international economy, with thousands of banks and funds, will through its process of collecting and trading, price the security at fair value, given zero transactional costs. Mathematically, stock prices follow a random walk and the more efficient a market is, the more fluctuation the stock price is ("Efficient Market Hypothesis" 2). In this paper, we will examine the implications of efficient market hypothesis on mutual fund portfolio disclosure. Based on the efficient market hypothesis, the more frequent mutual funds show their stock holdings, the more efficient the market for those stocks should be, and as a result, the less profit the mutual funds can make, due to an increase randomness of the stock holdings.

Mutual Funds as an investment vehicle have been ever growing in popularity ever since its revitalization in the 1980s. One of the main reasons why it has become such of staple of the financial markets is due to its commitment to lower investor's risk through diversification of its stock holdings. Unlike a hedge fund, which survives and thrives on leveraged and other risky positions to capture large but sometimes, risky profits, mutual funds trades risk and large profits in for liquidity and slow but steady profits. But since most are actively managed, mutual funds, in terms of function, can be placed in between hedge funds and a traditional savings bank. Under the Investment Company Act of 1940, before the 2004 revision, mutual funds were required to disclose their portfolio holdings semiannually. After the revision in May of 2004, mutual funds had to increase the frequency of disclosure from semiannually to quarterly through N-CSR and N-Q forms ("Final Rule"). Within the revision, mutual funds are allowed up to sixty days from the end of the quarter to file their complete portfolio holdings as of the end of the that quarter. The SEC allows mutual funds to delay their reporting of their quarterly holdings for up to sixty days because the provision serves to curb "predatory trading practices" undertaken by some high frequency trading firms ("Final Rule").

Based on the Efficient Market Hypothesis, which stipulates that all available relevant information regarding any security will be ultimately embedded into the price of that security, I hypothesize that mutual funds with extra information about the market would be more likely to delay reporting after the quarter has passed. The discretion of whether to delay reporting portfolio holdings by mutual funds on the surface seems trivial, as many firms may need more time to reconcile and ultimately close trades. But from an efficient market hypothesis' standpoint, delaying portfolio disclosure report is in effect delaying information emission to all market participants. Thus, funds with special information with delay can capitalize on the mispricing of securities and make higher returns.

Our analysis is based on mutual funds dataset merged from Morningstar, CRSP and the SEC Edgar. I build upon the dataset created in Pastor et al. (2013) by merging it with mutual fund data extracted from SEC's Edgar database. Our final dataset contains 3,134 benchmark-adjusted returns for 354 distinct funds from 2004-2014.

For our empirical analysis, I first examine fund level returns measured in BAR, or benchmark adjusted return. I regress BAR on delay time using cross-sectional data and then adjust the regression model as to alleviate omitted variable biases. Next, I discretize delay variables into five quintiles and examine the extreme values to see if they are significantly different from each other. Lastly, I aggregate our returns dataset from fund level to portfolio level and run the same set of tests to give a more comprehensive picture.

This paper will be structured as follows. Section 2 provides a comprehensive survey of all literature related to this specific topic. Section 3 describes the contribution this paper makes to the field of empirical finance. Data work and study methodology make up the bulk of section 4.

Section 5 presents the results and subsequent analysis. Section 6 concludes the paper and explores its shortcomings and adds suggestions for future improvements.

2. Literature Review

Mandatory portfolio disclosure is an integral part of regulations in the securities market. The general public, many of whom have stakes in investment companies, obtain information on the performance of funds through different filings required by the SEC. In section 13(f) of the 1934 Securities Exchange Act, all investment companies who manage over 100 million assets must file Form 13F quarterly ("Securities Exchange"). Under certain circumstances, managers can file for confidential treatment of their portfolio holdings with the SEC, and if approved, will allow funds to delay their complete portfolio disclosure for up to 45 days. Starting in 2004, the SEC revised the Investment Company Act of 1940 and now also requires all investment management companies, regardless of size under management, to file quarter report of their fund holdings through Form N-Q and Form N-CSR.

Note that Form 13F is different from Form N-Q and N-CSR in a number of ways. First, Form 13F requires funds to disclose their portfolio holdings at the company level while Form N-Q and N-CSR disclose holdings at the individual fund level. As Agarwal et al. (2014) pointed out in their paper, N-Q and N-CSR give a more complete picture of an investment company's financial health, especially for mutual funds. Mutual Funds often times operate many different "sub-funds" inside the "family" fund. For example, Vanguard, a 4 trillion dollar family mutual fund, contains many sub-funds (Dividend Growth, Tax Exempt, etc.)[1]. It is generally more informative to analyze fund level data because different funds have different investing

[1] Every quarter, Vanguard Group would file Form 13F while its sub-funds would file N-CSR and N-Q

objectives, and thus considering the general picture of company level portfolio through Form 13F may overlook significant details at the fund level. Second, Form 13F is only required for investment companies with asset under management of above or at 100 million dollars, while Form N-Q and Form N-CSR are required for all mutual funds, regardless of size. In short, Form N-Q and Form N-CSR give a more comprehensive look at the mutual fund industry than Form 13F.

A significant amount of literature in empirical finance has focused on consequences of mandatory portfolio disclosure of investment companies on their performances. Agarwal et al. (2013) examined the performance of hedge funds that exercise their “confidential treatment” in regards to Form 13F and found that portfolios under “confidential treatment” outperform the market for up to twelve months. More specifically, Agarwal et al. (2013) compared the returns of hedge funds from “confidential treatment” to the returns of those funds that do not delay their portfolio holdings. The researchers found confidential holdings exhibited significant returns as compared to those without confidential holdings at 10% significance. In conclusion, the paper argues that private information, or information that spots market inefficiencies, is the primary motive for authorizing the delay of Form 13F. Agarwal et al. (2013) is built upon the works of Brown and Schwarz (2010), which investigated the impact of mandatory portfolio disclosure through Form 13F on market participants. The paper found abnormal trading activities right before and right after the disclosure of Form 13F. Although the study discovered that there were no long-term profits for traders trading on information from 13F, private information withheld by the hedge funds before the disclosure date still produced superior returns. Brown and Schwarz (2010) hinted at the possibility that information disclosed on Form 13F contain valuable information for those before actual disclosure while Agarwal et al. (2013) confirmed this theory using the delay time as the primary effect. Christenoffersen et al. (2014) and Aragon et al. (2010) further affirms the claim that fund managers withhold information from the public to not only prevent “copy-catter” traders, but also to mask current profitable strategies used by themselves during the withholding period.

The literature summarized above investigates hedge funds using Form 13F. As stated above, Form 13F may not be the most accurate measure for particular investment companies, such as mutual funds. Studies that explore the same problem in empirical finance using Form N-CSR and Form N-Q are sparse. Agarwal et al. (2014) examines the direct effect of increased mandatory portfolio disclosure through Form N-Q and Form N-CSR on mutual funds. From 2004, mutual funds went from a semi-annual portfolio disclosure to quarterly disclosure. The intuition is that given the requirement of more frequent disclosure of portfolio holdings, the liquidity of securities listed on those reports should increase and that the funds that disclosed those holdings should expect their profits to be negatively correlated to disclosure frequency. The paper found that the liquidity of securities listed on Form N-Q and N-CSR increased and that mutual fund performance decreased after the revision in 2004. More specifically, because the public learns more information with disclosure, more frequent filing causes the funds to lose significant profits.

In this paper, I seek to investigate Agarwal et al. (2013), Aragon et al. (2010) and Christenoffersen et al. (2014) claim that withholding information through delaying disclosure produces superior returns for mutual funds. But instead of perusing Form 13F, I will be analyzing Form N-CSR and Form N-Q. Certainly, I will be using data from 2004 onward, as the implementation of the SEC amendment began shortly after 2004.

3. Contributions

While prior literature has examined the effect of delay disclosure of portfolio holdings for hedge funds, there has been little to no studies investigating the effect of delay on mutual funds performance. Agarwal et al. (2014) looks at the effect of the 2004 amendment to mutual funds disclosure by comparing the pre 2004 to post 2004 fund performances at the fund level using Form N-Q and Form N-CSR. Instead of examining the effects of the 2004 rule change, this paper takes the rule as given and extends Agarwal et al.'s (2014) conclusion that mutual funds lose profits with increased frequency of disclosure. In my analysis, I hypothesize a direct byproduct of Agarwal et al.'s claim, that is, mutual funds with private information that do not want to lose quarterly profit will decide to exercise the "60 day delay" provision, an action that is very much similar to hedge funds using "confidentiality treatment" of their portfolios. Logically, if mutual funds with private information delay submitting their forms, they should see a significant increase in their quarterly returns, benchmarked against those that do not (or have little) delay. In short, this paper, using mutual funds return dataset from 2004 onward, will examine the direct link between delay time and benchmark-adjusted return, regardless of the assets under management.

The implication of this study on mutual funds is arguably more relevant than those done on hedge funds for several reasons. One, investing in mutual funds is increasing in popularity among the general public as the alternative and sometimes de facto way to direct savings. Due to the enduring low-interest-rates policies worldwide, it has become more and more attractive to invest in mutual funds. Second, not everyone has access to hedge funds. Many successful funds require certain capital limits and personal connections to join as a regular investor. On the other hand, mutual funds are, in essence, securities that can be bought or sold with little trouble.

4. Methodology

In this section, I develop the methods and data used for the study of mutual funds returns. The reader will find that the estimating the effect of delay on BAR (benchmark-adjusted return) involves more than a simple regression as endogenous variables inevitably present themselves in the process. The next subsection will explain the original model and the modifications undertaken to avoid biases. The following subsection discusses the data collection process, which proved to be the most arduous and labor-intensive operation for this study. Data collection involved web-scraping data from the SEC Edgar database, crossing mapping it with original dataset created by Pastor et al. (2013), and finally converting all relevant variables into quarterly data.

4.1. Model

4.1.1. Simple Model

Estimating the effect of delay on BAR (benchmark-adjusted returns) requires a time series model across funds as follows,

$$R_{it} = a + bDelay_{it} + e_{it} \ (1)$$

where R denotes benchmark-adjusted fund return a is intercept and e denotes idiosyncratic risk associated with that specific fund. R is measured in quarterly terms of a specific year. Thus subscript t can be 3rd quarter of 2004, which covers the specific fund's performance from July 1, 2004 to September 30, 2004. e refers to fund specific, unobservable risks or factors that can affect individual fund return. In addition, the error term is assumed to exhibit heteroskedasticity, but cases involving i.i.d. error terms will be considered in the analysis. The variable $Delay$ is measured in number of days[2] the fund delays reporting of either Form N-Q or N-CSR[3].

4.1.2 Fund-Fixed Effect Model

A significant problem that arises in the model is the omitted variable bias. Because returns are measured at the individual fund level, the coefficient *b* will biased if variations within each fund significantly affect the quarterly returns and is also correlated with the delay variable. More specifically, each individual fund has unique characteristics such as quality, skill of its managers, etc. that can affect fund returns in addition to delaying their portfolio. For example, manager A runs a successful fund, which happens to be large with 10 trillion under management. Because the fund is large, it may take more time filing N-Q or N-CSR forms than funds with less AUM. Thus, a greater delay, for a larger fund, does not necessarily imply private information and thus greater returns for that specific quarter. Larger funds are by nature more successful in producing superior returns and in turn need greater time to file N-Qs and N-CSRs. In short, if delay were independent of the size of the firm, then the simple regression given above would give an unbiased estimate for the effect of delay on returns. There are many other fund specific qualities that may cause the simple regression to be biased that are either unobservable or undiscovered.

To alleviate the potential problem of omitted variables, I inserted a fund fixed effect variable into the original model,

$$R_{it} = a_i + bDelay_{it} + e_{it} \ (2)$$

2 The value ranges from 0-60

3 Form N-Q is filed during the end of first and third quarter while Form N-CSR is filed during the second and fourth quarter

where a_i denotes the fund fixed affect variable.

4.1.3 Fund Category-Fixed Effect Model

In constructing the dataset in Pastor et al.'s (2013) paper, they included a variable provided by Morningstar named *Morningstar Category*, which classifies mutual funds into different investment categories to guide investors' decisions. Morningstar is an investment research fund that created the "Morningstar category" that maps portfolios to peer groups based on their holdings[4]. The Morningstar category currently contains nine groups, that is, U.S. equity, sector equity, allocation, international equity, alternative, commodities, taxable bond, municipal bond, and money market.

Considering the original simple regression model in (1) and the fund-fixed effect model in (2), the coefficient on *Delay* may still be biased due to variations within categories of funds. For example, one of Morningstar Category is money market. Portfolios within that category invest in either taxable or tax-free money market securities that provide a level of return that preserves original capital. Another fund in this paper's dataset is U.S. equity small growth, which comprises of portfolios that focuses on small capitalization companies that are up-and-coming. Since these companies are in their early growth stages, the stock prices underlying are the most volatile with very little information about future prospects. Clearly, since both funds are in the dataset, there would be significant bias in the original model without accounting for category-fixed effects. Funds that are classified as money market will produce lower quarterly return, but they would also not be incentivized to delay their holdings due to the transparency of money market securities. On the other hand, funds that are classified as U.S. equity small growth, in the

[4] "The Morningstar Category Classifications." (2014): n. pag. Web.

presence of private information about an up-and-coming company, may seek to delay their disclosure to capture bigger profits. Thus, just as variations within funds were accounted for using fund-fixed effects, so to variations within categories of funds must be adjusted for. The Morningstar category-fixed effect model is as follows,

$$R_{it} = c_i + bDelay_{it} + e_{it} \ (3)$$

where c_i denotes the categorical-fixed effect. Adjusting for variations within categories of funds that affect returns will give a more accurate measure of the effect of delay on fund performance.

4.1.4. Discretizing Delay Variables into Quintiles

So far in my methodology, *Delay* has been treated as a continuous variable across individual funds. However, a potential problem that may occur is noise. *Delay* as a continuous predictor may contain too much unobservable noise for the analysis to decompose, even if adjusted for fund-fixed and category-fixed effects. Thus, a potential remedy is to discretize the delay variables into quintiles, sorted by quarter and year. For each quarter of every year from 2004-2014, funds will be assigned to either the first, second, third, fourth or fifth quintile. The first quintile contains the bottom 20% of fund based on delay time; the second quintile contains the lower middle 20%, and so forth.

Discretizing the delay variables into quintiles creates yet another model for analysis, that is,

$$R_{it} = a + bDummy_2 + cDummy_3 + dDummy_4 + fDummy_5 + e_{it} \ (4)$$

where $Dummy_i$ assigns the value 1 if the particular fund is in the ith quintile and 0 otherwise for that particular quarter. Note that $Dummy_1$ is excluded from the analysis to allow for a benchmark comparison between the coefficients on the remaining dummy variables. In other words, this particular model is less concerned with variable $Dummy_i$ for i < 5. The main point of interest is the coefficient *f*, which measures the average difference between the extreme quintiles, that is, the top 20 percent of returns versus the bottom 20 percent of returns, indexed by delay time. The decision to leave out interpretations of coefficients *b, c*, and *d* is based on the assumption that delay time in the middle of the quintiles is too noisy a predictor of fund's quarterly returns.

4.1.5 Aggregation to Portfolio Level Using FundSize

Consider the original model (1),

$$R_{it} = a + bDelay_{it} + e_{it} \text{ ,}$$

In the previous sub-sections, I proposed several modifications to the model using fund-fixed and category-fixed effect to mitigate potential bias in the coefficient of interest. Nevertheless, idiosyncratic risk e_{it}, or risk associated with individual funds has not been examined and accounted for. Because idiosyncratic risk is both specific and unobservable, fixed-effect modeling as seen in (2) and (3) may be too intricate and difficult to implement in practice.

One remedy is to aggregate the return dataset from fund level to portfolio level using the value weighted approach. In other words, for each quarter's quintile, I calculated the weighted average BAR using *Fundsize*, or the current asset under management. Essentially, each quarter now contains five "funds" corresponding to the five quintiles that contain weighted average BAR for that specific quintile. Based on the modern portfolio theory, aggregating returns dataset into a large portfolio is described as diversification and it washes out idiosyncratic risk as the number of securities within the portfolio increases. In short, the process of aggregation polishes out fund-level noises, which also makes the estimation process less noisy.

In addition, aggregating to portfolio level data also in part simulates possible investing behavior. That is, an investment strategy can be created by going long (buying) on funds that are in the 5th quintile of the quarter and shorting (selling) funds that are in the 1st quintile of the quarter. Investigating the profitability of such an investment strategy would involve examining the average portfolio benchmark-adjusted return difference between the 5^{th} and 1^{st} quintiles. Thus, for the analysis part of this paper, the original model (1) and (4) will be estimated at the portfolio level and their results will be discussed.

4.2. Data

Data collection proved to be the most difficult and intricate process from this paper. I sought to obtain mutual funds data that included delay time, gross return, benchmark-adjusted return, and AUM (asset under management). The final dataset was created by means of cleaning and merging two separate data files, one with monthly fund returns obtained from Morningstar and CRSP, and the other with quarterly delay times for each fund from SEC Edgar. Merging the two dataset required converting all monthly variables into quarterly variables, and challenges involving missing data inevitably surface in the analysis. The next subsection will explain the data construction for both of the datasets that were used to create the final product, and the following subsection will explain the process of transforming monthly data points into quarterly data points.

4.2.1 Data Construction

The raw data points were extracted from three sources, Morningstar, CRSP, and SEC Edgar. The benefits of using mutual funds data from Morningstar and CRSP are described in detail in Pastor et al. (2013). One of them is the ability to check for accuracy of values by comparing data from Morningstar with that of CRSP. Second, as stated earlier, Morningstar classifies funds into categories based on their portfolio holdings, which clears up a lot of confusion with identification. Third, Morningstar provides a benchmark portfolio to each individual fund, which makes the calculation of benchmark-adjusted returns much easier. The SEC Edgar database was used to extract Form N-Q and Form N-CSR and mine the text for delay time.

The first dataset obtained monthly data points from Morningstar and CRSP. Key variables in the file include *FundId, year,month, fund_return, return_idx, and fund_size. FundId* gives a unique 10-letter string for fund identification. *fund_return* gives the monthly return in percentages of the fund of interest. *return_idx* gives the benchmark return for that specific fund during a month of a specified year. Finally, *fund_size* gives the previous month's calculation of the fund's asset under management[5]. Using Pastor et al.'s (2013) as a construction guide, the dataset dropped all bond funds, index funds, money market funds, and any other funds that were not equity related. In addition, the dataset followed the data cleaning process used by Berk and Binsbergen (2012) to increase the accuracy of the measure of return and fund size. Pastor et al. (2013), using Berk and Binsbergen's (2012) method, were able to decrease the error for fund return and fund size to 0.6% and 7.3%[6], respectively. The end result of the first dataset is a collection of funds with returns, benchmark returns, and fund size measured at the monthly level.

Since the only missing variable of interest is delay time in days, data points for the second dataset were scrapped from the SEC Edgar database, which contains information on the delay variable. Within Edgar, there exists an archive of all documents every mutual fund file under the SEC but I concentrated on selections involving only N-CSR and N-Q forms filed by mutual funds. Since the database range from 1993 to 2016, I extracted every single N-CSR or N-Q forms from each individual quarter of each year (2004-2014) and gather key information such as fund name, form type, file date, fund ID, and URL address to the actual document. Then, I

---

[5] Adjusted for inflation as of January 1, 2000. See Berk and Binsbergen (2012)

[6] The error rate is measured by the discrepancy between values from CRSP and those from Morningstar

went into each individual document (using the URL) to extract the file date and the date of report. An example of an N-Q form is provided for view in figure 1. The file date and date of report comprises the necessary numbers to calculate delay time which is simply file date minus date of report. Once all the information is extracted, we will utilize Stata to sort by company name, and each row will contain the associated variables—name, form type, file date, date of report, sub funds, ticker symbols, company ID, and URL. The end result is a collection of mutual funds from 2004-2014 with quarterly variable delay.

4.2.2 Data Merging

The next step is to merge the two dataset. One significant challenge I faced is the difference in time frequency of the data. The first dataset contain returns at the monthly level while the second dataset contain delay at the quarterly variable. The decision to convert monthly returns to quarterly return is intuitive, since quarterly return is a meaningful measure aggregated over three months, while converting quarterly delay into monthly delay makes the variable meaningless[7]. The process of converting a specific fund's monthly return to quarterly return is as follows,

$$R_q = \left(1 + \frac{R_{m(1)}}{100}\right) * \left(1 + \frac{R_{m(2)}}{100}\right) * \left(1 + \frac{R_{m(3)}}{100}\right)$$

That is, for each fund over three consecutive months, I compounded the monthly returns and also the benchmark returns (measured in percentage) over three periods. After adjusting the data for returns, I now have to match the fund names from the first dataset to the second dataset. Given a lack of natural language processing skills, I relied on Stata to given a complete string matching function to match the fund names. One big drawback using this method is that many firms' data were omitted because they did not have a complete string matching between the datasets, even if

[7] Recall that N-Q and N-CSR are quarterly filings

they captured similar funds. The completed dataset now contain all the necessary variables for analysis. That is, *Fund_Id*, *fund_name, delay*, *BAR_q*, *MorningstarCategory, FundSize, delay_qnt, and BAR_new*. *BAR_q* denotes the benchmark adjusted return for a fund given the quarter and the year. *Fund_Size* is the AUM measured in millions. delay*_qnt* assigns the quintile to the fund, given the delay time. Finally, *BAR_new* gives the fund-weighted benchmark-adjusted return for a particular quarter and quintile. Complete summary statistics are given in Table 1.

5. Empirical Results

In this section, results from analyzing the data will be discussed. Before diving into the actual estimations, it is necessary to visualize the distribution of the delay variable, which is shown in figure 2.

5.1 Distribution of delay time

My hypothesis is that the histogram of delay time will be bi-modal. The intuition behind the bi-modal distribution is that there are certain funds that have extra information about the market and where securities should be fairly priced, and they would withhold such information for as long as they can. Thus, a significantly large number of delay times would be centered on a particular value. On the other hand, there are companies who do not have extra information about the market and it would be unlikely that they delay the holding period.

The histogram of delay shows a slight skew to the right, with more than twenty percent of all funds delaying their filings for more than 45 days. This behavior presents a potential problem in my analysis. If the dataset does not contain many firms who choose to not delay and file their forms early, then my analysis will not accurately measure the effect of delay on returns.

However, it might be the case that most mutual funds do not file early. Thus, I proceed to regression analysis.

5.2 Simple Regression on Delay

The estimation of the coefficient on *Delay* for original model (1) yields .00000178 with a t statistics of 2.07, which is significant at the 5% level. Since the number is positive in direction, the relationship between delay and benchmark-adjusted return is as predicted. The more days a given fund chooses to delay, the greater the benchmark adjusted return for that fund. For every day the fund chooses to delay their holdings, the quarterly returns goes up by .00000178 or .000178 percent. The resulting coefficient is most likely not economically significant. Even with a 60-day increase in delay, the quarterly benchmark adjusted return will only go up by 1 basis point, which is negligible.

5.2 Regression with fund-fixed and category-fixed effects

Regressing the benchmark adjusted return on delay with the inclusion of fund-fixed effects (2) and category-fixed effects (3) gives coefficients of 0.00000115 and 0.00000171, respectively. Adjusting for variations within individual funds that may affect benchmark adjusted return does not produce significant coefficient on the delay variable and thus delay is not necessarily a good predictor for benchmark-adjusted returns of mutual funds after adjusting for fund-fixed effects. On the other hand, adjusting for categories (Morningstar category) fixed effects gives a slightly significant coefficient at the ten percent level. While both coefficients indicate a positive relationship between benchmark adjusted return and delay time, they are no longer significant after adjusting for fund and category fixed effects. The complete results are shown in table 2.

5.3 Quintiles Regression

The previous analysis has treated delay time as a continuous variable, and that treatment may make my analysis noisy. Thus, I now transform delay into a discrete variable through the process of quintiles generation. For each quarter of each year, funds will be classified into one of five quintiles. The first quintile contains the lowest twenty percent of funds based on delay time and the fifth quintile contains the highest twenty percent of funds based on delay time. Since this study is only concerned with differences between extreme values i.e. 5th quintile versus 1st quintile, the coefficient on the 5th quintile will only be considered. The estimators measuring the average difference between the extreme quintiles show a positive direction, that is, funds in the 5th quintile on average generate more benchmark-adjusted return than funds in the 1st quintile. Nevertheless, the coefficients are not significant, which means that I cannot conclude that firms that choose to delay will have a significant advantage over those that do not choose to delay their portfolio disclosure. The complete results are shown in table 3.

5.4 Results from aggregation to portfolio level

Idiosyncratic error terms that were prevalent in fund level data diminish significantly when benchmark-adjusted return is aggregated into portfolio level data. Now, for each quarter, quintile, benchmark-adjusted returns were recalculated by incorporating fund size. More specifically, every quintile (1-5) within each quarter of every year will have a new variable, which calculates the benchmark-adjusted return weighted by fund size. For example, for quarter 2 of 2004, the weighted portfolio level benchmark-adjusted returns are 0.00007221,-0.00004587, -0.00005119, -000006743, and 0.00024886 respectively for every quintile. The summary statistics for portfolio level returns is presented in table 4. Running the quintile regression as seen in the previous section yields -0.00000197 for the coefficient on $dummy_5$, which shows a negative average difference between the benchmark-adjusted return of the 5th quintile and the 1st quintile. However, given the magnitude of the coefficient, it is far from statistical or economical significance. Thus, aggregation from fund-level to portfolio level further strengthens the claim that portfolios with funds that significantly delay their disclosure would not outperform portfolios with funds that do not delay. The full results, including t-statistics, are shown in table 5.

6. Conclusion

In this paper, I examined the effect of delaying portfolio disclosure on mutual funds quarterly returns. The simple regression model shows a positive and significant relationship between delay and returns. In other words, the greater the delay, the greater the return. In addition, I utilized a number of models that sought to correct the omitted variable biases inherent in the simple regression model. The improved models, which adjusted for category and fund fixed effects, do not show a significant effect of delay on benchmark-adjusted return. Due to the noise presented in continuous variables, I then discretized delay time into quintiles and sought to find difference in values between the extreme quintiles. In the dummy regression, there was no evidence of significant difference between the extreme quintiles. Similarly, aggregating the data from fund-level to portfolio-level also showed no significant difference between funds that delay disclosure and funds that do not delay.

Given the results of this paper, it may seem that the market proves to be efficient enough that even with special information withheld from the public, prices still adjust quickly to the point that funds do not make significant profits. While there seems to be no statistical significance between delay and fund returns, most of the statistical test showed a positive relationship between delay and benchmark-adjusted return, even if the coefficients weren't significant. In addition, recall that the histogram of delay variable is skewed to the right, with a low number of funds that disclosure early (less delay). It may be the case that most mutual funds delay their holdings, but additional studies need to be undertaken to confirm this claim. During the data merging part of this paper, many funds were dropped because they did not exactly match for the two original dataset. I did not have extensive knowledge of text mining, and was unable to completely merge the two. Thus, if this study left out significant data points for funds that do not delay, then my analysis will not be representative of the whole mutual fund industry. Future studies would need to have a much larger database in order to carry out a bias-free investigation.

References

Agarwal, Vikas, Kevin A. Mullally, Yuehua Tang, and Baozhong Yang. "Mandatory Portfolio Disclosure, Stock Liquidity, and Mutual Fund Performance." *The Journal of Finance* 70.6 (2015): 2733-776. Web.

Agarwal, Vikas, Wei Jiang, Yuehua Tang, and Baozhong Yang. "Uncovering Hedge Fund Skill from the Portfolio Holdings They Hide." *SSRN Electronic Journal* (2013): n. pag. Web.

Aragon, George O., Michael G. Hertzel, and Zhen Shi. "Why Do Hedge Funds Avoid Disclosure? Evidence from Confidential 13F Filings." *SSRN Electronic Journal* (2010): n. pag. Web.

Berk, Jonathan, and Jules Van Binsbergen. "Measuring Managerial Skill in the Mutual Fund Industry." (20123): n. pag. Web.

Brown, Stephen J., and Christopher Schwarz. "The Impact of Mandatory Hedge Fund Portfolio Disclosure." *SSRN Electronic Journal* (2010): n. pag. Web.

Christoffersen, Susan Kerr, Erfan Danesh, and David K. Musto. "Why Do Institutions Delay Reporting Their Shareholdings? Evidence from Form 13F." *SSRN Electronic Journal* (2014): n. pag. Web.

"Final Rule: Shareholder Reports and Quarterly Portfolio Disclosure of Registered Management Investment Companies." *Final Rule: Shareholder Reports and Quarterly Portfolio Disclosure of Registered Management Investment Companies; Release Nos. 33-8393; 34-49333; IC-26372; File No. S7-51-02*. N.p., n.d. Web. 29 Mar. 2017.

Pastor, Lubos, Robert Stambaugh, and Lucian Taylor. "Scale and Skill in Active Management." (2014): n. pag. Web.

Securities Exchange Act of 1934. Ann Arbor, MI: Xerox U Microfilms, 1974. Web.

```
-----BEGIN PRIVACY-ENHANCED MESSAGE-----
Proc-Type: 2001,MIC-CLEAR
Originator-Name: webmaster@www.sec.gov
Originator-Key-Asymmetric:
 MFgwCgYEVQgBAQICAf8DSgAwRwJAW2sNKK9AVtBzYZmr6aGjlWyK3XmZv3dTINen
 TWSM7vrzLADbmYQaionwg5sDW3P6oaM5D3tdezXMm7z1T+B+twIDAQAB
MIC-Info: RSA-MD5,RSA,
 Q15t8mlzge9xrvt/OIoSXQc6TD9dvUkq9pwtaOdH0J+t/jnV0m0d676kFcMfmX/e
 ZnlIiyTesqTIQUc542NTOw==

<SEC-DOCUMENT>0000950135-08-006242.txt : 20080929
<SEC-HEADER>0000950135-08-006242.hdr.sgml : 20080929
<ACCEPTANCE-DATETIME>20080929125554
ACCESSION NUMBER:               0000950135-08-006242
CONFORMED SUBMISSION TYPE:      N-Q
PUBLIC DOCUMENT COUNT:          2
CONFORMED PERIOD OF REPORT:     20080731
FILED AS OF DATE:               20080929
DATE AS OF CHANGE:              20080929
EFFECTIVENESS DATE:             20080929

FILER:

        COMPANY DATA:
                COMPANY CONFORMED NAME:                 HARTFORD MUTUAL FUNDS INC/CT
                CENTRAL INDEX KEY:                      0001006415
                IRS NUMBER:                             000000000
                STATE OF INCORPORATION:                 MD
                FISCAL YEAR END:                        1231

        FILING VALUES:
                FORM TYPE:              N-Q
                SEC ACT:                1940 Act
                SEC FILE NUMBER:        811-07589
                FILM NUMBER:            081093257
```

Figure 1: In this example, the period of reporting or portfolio holdings was on July 31st of 2008 but the actual filed date was on August 29th of 2008, resulting in a 29-day delay.

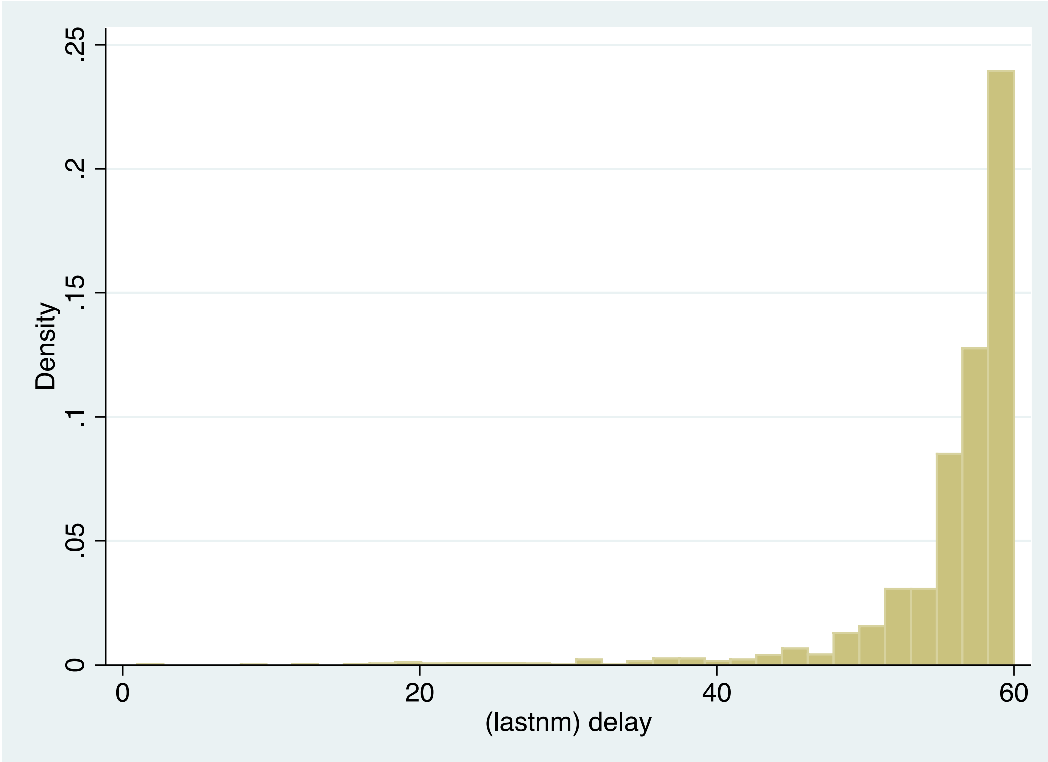

Figure 2: Histogram of the delay variable measured in days.

Table 1
Summary Statistics

	# of observations	Mean	Stdev.
Quarterly fund returns	3,134	0.000267	0.000852
Quarterly index returns	3,134	0.000295	0.0008187
Quarterly BAR	3,134	-0.0000278	0.0002847
Delay	3,134	56.02553	5.89638
Fund size	3,117	1499.059	4434.523

Table 2
Regression on Delay including FE Variables

The delay variable is measured in days. Coefficients marked with ***, **, and * is significant at the 1%, 5%, and 10% level. T-values are given inside the parentheses.

Delay	.00000178**	0.00000115	0.00000171*
	(-2.07)	(-1.53)	(-1.8)
Constant	(-)0.0001277***	(-)0.0001265***	(-)0.000116**
	(-2.63)	(-1.95)	(-2.23)
Model	OLS	OLS Fund FE	OLS Category FE

Table 3
Quintile Regression

In this study, I focused on coefficient on only $Dummy_5$, which gives the average difference of benchmark-adjusted return between the top 20% of funds versus the bottom 20% of funds based on delay time.

$Dummy_2$	(-)0.00000649	(-)0.00000792	(-)0.0000192
	(-0.49)	(-0.59)	(-1.32)
$Dummy_3$	0.0000175	0.0000155	0.0000156
	(1.21)	(1.06)	(0.97)
$Dummy_4$	0.00000604	(-)0.00000616	(-)0.0000047
	(-0.34)	(-0.35)	(-0.24)
$Dummy_5$	0.0000276	0.0000262	0.0000288
	(1.44)	(1.36)	(1.30)
Constant	(-)0.0000315***	(-)0.000026*	(-)0.0000278*
	(-3.45)	(-1.93)	(-2.79)
Methods	MLR	MLR Fund FE	MLR Category FE

Table 4
Summary Statistics at Portfolio Level

	Observations	Mean	Stdev.	Min	Max
Quintile 1	976	-0.000000249	0.0001513	-0.0004525	-0.0004525
Quintile 2	871	-0.000019	0.0001182	-0.0005016	0.0003506
Quintile 3	639	-0.0000128	0.0001045	-0.000321	0.0001744
Quintile 4	360	-0.0000505	0.0000935	-0.0004585	0.0003673
Quintile 5	288	0.00000222	0.0001794	-0.0003994	0.0006263

Table 5
Portfolio Level Quintile Regression

$Dummy_2$	(-)0.0000188**	(-) -0.000019**	(-)0.0000272***
	(-3.07)	(-3.09)	(-3.97)
$Dummy_3$	(-)0.0000125*	(-).0000128*	(-)0.0000171**
	(-1.87)	(-1.91)	(-2.26)
$Dummy_4$	(-)0.0000503***	(-)0.0000506***	(-)0.0000556***
	(-6.22)	(-6.23)	(-5.99)
$Dummy_5$	(-)0.00000197	0.0000219	(-)0.0000132
	(-0.22)	(-0.25)	(-1.26)
Constant	(-)0.000000249***	(-)0.00000324	(-)0.00000469
	(-0.06)	(-0.52)	-0.99
Methods	MLR	MLR Fund FE	MLR Category FE