Mutual Fund Investment
Flows in Sweden

Authors: Berggren Marcus & Sundström Karl-Peder
Supervisor: Alexius Annika
EC6902 Bachelor Thesis in Economics
Autumn 2023
Abstract

This thesis examines the behaviour of Swedish mutual fund savers concerning their investments in Swedish-domiciled mutual funds between 2013-2023. The analysis investigates the relationship between investor preferences and mutual fund characteristics, specifically focusing on management fees, Morningstar ratings, risk levels, and monthly returns. Drawing from contemporary literature and employing pooled OLS regression analysis, our findings highlight several noteworthy patterns in investor behaviour. Notably, our data underscores investor tendencies to gravitate toward high-performing mutual funds while demonstrating sensitivity to management fees. Our analysis reveals that past performance does not have a greater impact when it comes to investor decisions, however that Morningstar Rating and Fee’s do. These findings show insights into investor decision-making within the Swedish mutual fund market, contributing to the ongoing discourse on investor behaviour and mutual fund selection strategies.

Keywords
Mutual Funds, Pooled, Regression Analysis, Swedish Funds, Morningstar, Investment Behaviour
# Table of Contents

1. Introduction...................................................................................................................... 1

2. Previous Literature.......................................................................................................... 2

3. Data ................................................................................................................................. 5

4. Empirical Method ............................................................................................................ 8

5. Results ............................................................................................................................. 11
   Discussion ......................................................................................................................... 15

6. Conclusion ....................................................................................................................... 17

7. Special Thanks ................................................................................................................. 18

References ........................................................................................................................... 18

Appendix ............................................................................................................................... 21

A: Data .................................................................................................................................. 21
1. Introduction

During the 1980s there was a great increase in the number of mutual fund assets in Sweden (Waldenström, 2022, 3) and in 2019 it was measured that the total mutual fund assets of Sweden went up to more than 4.4 trillion Swedish Kronor (SEK) from 40 years prior being a mere 1 billion SEK (Waldenström, 2022). What has caused such a rapid acceleration has been a multitude of political factors such as laws being put in place to increase savings and financially sophisticated households, and in comparison, to other countries, there is a strong participation in the stock market (Pettersson, Sjöholm and Hård, 2019, 4-5). If you include Sweden’s pension system’s saving forms, the occupational pension, and the mandatory national public pension, almost all Swedish citizens are mutual fund savers (Fondbolagens Förening, Så blev Sverige världsbäst på fondsparande). With strong participation in the stock market and a large amount of Swedish citizens being mutual fund savers, questions can be asked regarding their investment behaviour. According to the Swedish Investment Fund Association, there is information to back how mutual fund savings are connected to rational thinking such as that people with a higher level of education will own mutual funds in almost twice the amount compared to those that have only gone to middle school (Waldenström, 2022, 41-42). In different studies of the American mutual fund market, there has been a strong positive relationship between the amount of net saving in funds and the mutual fund market return i.e. that the better the market performs the more money people will save in mutual funds (Andersen, 2006, 6). Research in behavioural economics shows that "people's judgements and decisions are often subject to systematic biases and heuristics and are strongly dependent on the context of the decision" (Reisch and Zhao, 2017, 190).

The essay aims to add value to the scientific area of financial and behavioural economics, by using new Swedish fund data acquired from the Morningstar Direct database to understand Swedish investors, with a focus on Swedish-domiciled share classes where the investors are non-professionals. The research question of this thesis will be then:
How does Swedish mutual fund savers invest towards Swedish domiciled mutual funds in relation to the management fee, Morningstar rating, risk level, and the monthly return studied during a ten-year time span between the years 2013-2023?

By answering this, insights can be made, into how investors behave in the Swedish fund market by looking at the results of this thesis and by concluding the output of the conducted multiple regression analysis. This might be intriguing for most modern fund managers of our time.

To achieve what this thesis has set out to do and to answer the research question, the thesis will be structured in the following way. Section 2 of this thesis will discuss the previous literature which will contribute to the discussion and conclusion of the thesis. Section 3 will describe the data points, and in section 4 the method of this thesis will be described and how the data will be analyzed. Section 5 will present the results and include a discussion of the results, where we aim to create clarity for our intention of adding value to our chosen subject and the last section, section 6, will be the conclusion.

2. Previous Literature

“I should have computed the historic co-variances of the asset classes and drawn an efficient frontier. Instead,...I split my contributions fifty-fifty between bonds and equity’s” (Zweig, 2007, 4)

Throughout academic literature, there have been numerous studies on how consumer behaviour interacts with mutual funds in an international context. In Sirri & Tufano’s paper Costly Search and Mutual Fund Flows (1998) they discover that consumers of mutual equity funds will gather disproportionately to high-performing mutual funds while at the same time failing to get away from lower-performing mutual funds at the same rate (Sirri and Tufano, 1998, 1590). In
addition to this, they discovered how mutual fund flows are fee-sensitive, but consumers’ response to these fees is asymmetric in that they, according to the authors, will respond differently to high and low fees, as well as to fee increases and decreases (Sirri and Tufano, 1998, 1590). Lastly, they argue that consumers may respond to the risk of their portfolios, which “may offset—but may not eliminate—managers’ incentives to increase fund volatility” (Sirri and Tufano, 1998, 1619). In the summary section of their paper the authors show that there is evidence that the aggregate pattern of consumer behaviour is rational, however, credit must be given to the marketing schemes of the different fund firms (Sirri and Tufano, 1998, 1619-1620). Lastly, the authors introduce theories of performance chasing behaviour, that investors will chase mutual funds with recent positive and strong behaviour (Sirri and Tufano, 1998, 1619).

Ferreira, Keswani, Miguel & Ramos study how fund flows depend strongly on past performance. They argue that fund flows respond to past performance, which is significant as it argues for fund persistence i.e. funds which have performed well in the past will be assumed to keep performing well. The flow-performance relationship will determine the degree to which fund size is affected by past performance and the conditions on how a fund performs in the future. (Ferreira et al., 2010, page 1).

When it comes to research more specifically about understanding consumer behaviour, in a Swedish context, Carlsson Hauff points out several key concepts in their paper, such as that behavioural finance and behaviour economics have had a similar evolutionary path and that is to argue that individuals are not purely rational decision-makers (Carlsson Hauff, 2014, 66). The economic man and what is commonly seen in real life according to the authors is not always the same, i.e. that there are discrepancies. In the book chapter “Attitude Toward Risk and Financial Literacy in Investment Planning”, Carlsson Hauff and others (Nicolini, Gärling, Carlander) focus on the relationship between financial literacy and investment decisions, using data from multiple countries such as Italy, Sweden, and Spain. The authors discuss how low financial literacy can lead one to suboptimal asset allocation (Nicolini et al., 2017, 310-311). In their paper, the authors reveal that people with lower financial literacy show more negative attitudes toward risk when
making their investments, which can cause them to accept negative returns and invest less in the stock market, arguing that they may not be rational (Nicolini et al., 2017, 313-317).

Several Swedish research papers have studied the connection between mutual fund performance and Morningstar performance metrics. One study, a bachelor's thesis by Englöv, Marktorp, Nilsson & Nimac studies the relationship between Morningstar’s rating, the return and then if it differed in economic good or bad times. This study was done by studying secondary data from daily newspapers and the internet (Nimac et al., 2008, 1). While this thesis studied multiple periods ranging from 2006-2008, which gave different results, however, the authors state that they could generally assess that equity funds that have a higher rating do not have a causal relationship with the return (Nimac et al., 2008, 37).

Another similar study called “Morgonstjärnornas krig: Morningstar-betyg & fonders framtida avkastning” by Fornander & Lindqvist studies if Morningstar’s fund rating system can be used as an indicator for future risk-adjusted return (Fornander and Lindquist, 2017, 1). The study makes use of regression analysis, and they discover some interesting points, such as that the regression shows that five-star funds do not outperform the four-starred ones and that overall, the study shows that there is no positive relationship between Morningstar ratings and future returns as the results do not show any statistically significant results. Lastly, while the results did not appear significant, the authors argue that the Morningstar rating is widely used in marketing and as a tool of comparison, it does work to affect the way investors invest. Nevertheless, according to the authors, it is perhaps more used as a tool to identify funds to avoid than funds to invest in (Fornander and Lindquist, 2017, 25).

We have also studied the previous literature to find appropriate empirical methods for our investigation. In Sirri and Tufano’s (1998) paper they use a multivariate regression model framework applied to twenty years of fund-level data, which makes use of a dependent variable FLOW and independent variables such as return, riskiness, and expenses, among others. These variables have subindexes of i and t to indicate the specific fund and specific time. In another paper “The Behavior of
Investor Flows in Corporate Bond Mutual Funds” authored by Chen & Qin regression analysis makes its use again when it comes to investigating the predictability of fund flows (Chen and Qin, 2015, 7). This thesis’ regression model also makes use of the variable $R$ which is return.

Previously introduced work by Fornander & Lindqvist (2017) make use of regression analysis in their work as well with the variable of the Morningstar Rating (Fornander and Lindquist, 2017, 15). Lastly, pooled OLS regression has been used to study the performance of the mutual fund industry, such as Rahman, Rahman and Subat’s article on the “Assessment of Performance of Mutual Funds Listed in Dhaka Stock Exchange, Bangladesh”. This paper used pooled regression on panel data between the years 2014 and 2019 where they could see a significant relationship between earnings per unit and return on assets (Rahman, Rahman and Subat, 2022, 1).

3. Data

The following graph represents an illustration of the aggregated yearly estimated net flows, for the mutual funds which we examine, between 2013-2023. This graph has two purposes, to first illustrate the performance of funds in Sweden as seen by the net flows from our data and to give context to the data on which this thesis will base its results and conclusions.

Graph 1  
*Taken from Morningstar Direct*
The data for this thesis is extracted from the Morningstar Direct database, which is a database that has an extensive collection of data on open-ended funds. We aim to study Swedish Large cap mutual funds that are domiciled in Sweden and have the Swedish Krona as the trading currency. This data covers 76 Swedish mutual funds between 2013-10 and 2023-09. Table 1 below is a summary of the data obtained:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENF</td>
<td>3393</td>
<td>6.379</td>
<td>463.798</td>
<td>-9009.105</td>
<td>8966.669</td>
</tr>
<tr>
<td>R</td>
<td>3393</td>
<td>.683</td>
<td>5.29</td>
<td>-18.074</td>
<td>17.251</td>
</tr>
<tr>
<td>Fee</td>
<td>3393</td>
<td>.865</td>
<td>.542</td>
<td>0</td>
<td>1.75</td>
</tr>
<tr>
<td>Rating</td>
<td>3393</td>
<td>3.17</td>
<td>1.054</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Risk</td>
<td>3393</td>
<td>18.991</td>
<td>1.988</td>
<td>14.192</td>
<td>24.188</td>
</tr>
</tbody>
</table>

Table 1

To clarify, out of the different filters we used to try and guarantee that the funds were domiciled to Sweden, that they used the Swedish krona as the trading currency we also only looked at Swedish Large Cap mutual funds. The panel data is monthly observations of the Estimated Fund Level Net Flow, Morningstar Rating, Monthly Return, Standard Deviation, and Management Fee taken over ten years. The different variables’ names have been shortened, and the new names following the previous order are ENF, Rating, R, Risk and Fee. The monthly observations are taken between the years 2013-2023. It should be noted that if we had a complete data set where monthly observations could be taken for all ten years, there would be 9120 observations, but there were missing months. Considering that this thesis makes use of regression analysis where the different variables are tested together, if one observation is missing then the rest will not be tested, ergo, only when there are observations for every variable that specific month can regression be done.

A potential problem that this thesis may face is survivorship bias. When it comes to the study of mutual funds one can argue that many databases use data from mutual funds that currently exist and do not tend to include information about
mutual funds that no longer exist. Mutual funds can appear and dissolve over ten years. This might cause omission for some of the monthly observations we extract. However, as argued by Sirri & Tufano in *Costly Search* (1998) it would be problematic if the data only included those mutual funds that did survive during the period we are measuring. Similarly, as for them, this thesis can fortunately use data from most mutual funds in the sample during the period because of monthly observations which makes the survivorship bias problem of this thesis not prevalent. This then includes observations from mutual funds that could have been liquidated or part of mergers during the 10 years. We argue that these observations will still play a part and contribute to our empirical model. In addition, this study has decided to include the years that COVID-19 took place, as the observations from that period are still useful for this thesis. Another issue that should be disclosed is regarding the Fee, Morningstar does not have it reported continuously, but only when it is updated from what it was from time to time, or when it is changed. We assume that it is constant during the months between, as a management fee rarely updates frequently. Apart from data being affected by limitation in the form of omission, we are dependent on the fund companies to report data to Morningstar. When fund companies do not report data, Morningstar conducts their estimations using their methodology.

In our empirical model, the dependent variable will be ENF, which is the differences in the flow of assets and liabilities (see Appendix A). The second variable is Fee, which is defined as the Management Fee at Morningstar Direct (see Appendix A). The third variable is Risk, which is described as the Standard Deviation of the monthly return, calculated on a 36-month basis. We aim to introduce this variable as a control variable as we assume that it will not have the same effect on investment decisions. There have been multiple changes in the mutual fund industry as to how risk is calculated and illustrated. To mention a few, we have the Value-at-Risk, Tracking Error, SRRI (Synthetic Risk and Reward Indicator) and now since the beginning of 2023, following the new PRIIP regulation, SRI (Summary Risk Indicator) has been infused and presented to the investors. All these risk measurements indicate that there is no single sufficient measurement of risk given the time studied for our thesis. According to the Institute of Business and Finance, "the most frequently used measurement of investment risk
is the standard deviation” (The Importance of Standard Deviation in Investment, 11/2023). The standard deviation with most investments is calculated using the monthly returns for the past 36 months, and therefore it is what shall be used in this thesis (The Importance of Standard Deviation in Investment, 11/2023).

The fourth variable Rating is a measurement that brings load adjustments, performance (returns) and risk together into one evaluation (see Appendix A). The last variable, $R$, is the Monthly Return of a fund. This is expressed in percentage terms (see Appendix A).

An adequate reason to extract monthly data in Sweden for our thesis is that most Swedes receive their salary, pension and other subsidies every month – which they, in turn, may use for investments in mutual funds. We assume that an investment decision can vary in time and that it is not realistic that all investors trade mutual fund shares as a reaction in the very same instance that the actual mutual fund characteristic is being reported. That is to say that investor A perhaps evaluates their investment decision based on mutual fund Xs’ e.g. performance depending on the previous months’ return or perhaps even the previous 3 years’ return. Consequently, we will lag $Fee$, $R$ and Rating by one month separately in our empirical model. We will also generate another control variable, an average monthly return variable, $R_{AVG}$, taking the average monthly return over the last 3 years, the very same period as the Risk variable. Average annual return is the more common type of metric which can be reported either for three, five- or ten-year periods, however, for the interests of the thesis it will be the average monthly return over three years because many of our other variables make use of monthly data.

### 4. Empirical Method

In this section the methodological framework will be described, as well as giving insight into how it will be used to analyze and examine the research question. As mentioned in section 2, previous work such as Sirri & Tufano’s in Costly Search (1998) made use of regression analysis when studying mutual funds and consumer
behaviour. They made use of 20 years of fund-level data, using variables such as flow, return, and riskiness among others. Other authors previously introduced, such as Rahman, Rahman and Subat (2022), used pooled regression to study the performance of mutual funds on panel data over a 5-year time period. Chen & Qin (2015) who were also previously introduced, make use of regression analysis to investigate the predictability of fund flows, whereby they have made use of the variable return return to understand the fund flows.

Taking inspiration from the previous literature, the thesis will make use of the following model:

\[ ENF_i = \beta_0 + \beta_1 R_i(t-1) + \beta_2 Fee_i(t-1) + \beta_3 Rating_i(t-1) + \beta_4 Risk_i + \beta_5 RAVG_i + \epsilon_i \] (1)

to study the proposed relationship of Swedish domiciled investors towards Swedish domiciled mutual funds. \( ENF_i \) will be the estimated net flows with a subindex of \( i \) and \( t \) whereby the former is the indicator for the fund and the latter is the time indicator. The following three variables \( \beta_1 R_i(t-1) \), \( \beta_2 Fee_i(t-1) \), \( \beta_3 Rating_i(t-1) \) will be the independent variables and \( \beta_4 Risk_i \), \( \beta_5 RAVG_i \) will be the control variables. By using equation (1) this thesis can create several models, whereby these models will test the different independent/control variables against the dependent variable until all the variables are accounted for in the different models. This will lead to five different models, which will allow us to see how the different variables will act towards each other.

These models will employ pooled multiple regression, given that this thesis makes use of panel data. Pooled time series regression analysis, also known as pooled regression, accommodates a greater variety of model covariates and sample sizes (Ward and Leigh, 1993, 646). Pooled regression is an extension of the ordinary least squares (OLS) regression and can be used for data whereby outcome variables are recorded for many individuals (cross-sections) at many time points (time-series), which is commonly known as panel data (Ward and Leigh, 1993, 646). Why this is useful for our thesis is that pooled regression is a technique that incorporates “the data of multiple time series from multiple individuals in a single regression model,
allowing the time course of the outcome to be modelled using the data for an entire group while accommodating the lack of independence of data contributed by given individuals in the group” (Ward and Leigh, 1993, 646). Considering that this thesis is making use of panel data and that we are studying a length of time that spans 10 years, this method was deemed appropriate. Furthermore, it is a useful model because it "assumes that there are no unobservable entity-specific effects, meaning that all entities in the data set are considered to have the same underlying characteristics" (The Pooled OLS Model - Tilburg Science Hub, accessed 16/11-2023), with this, we assume that all the funds we have looked at do not have any non-observable company-specific effects, with the same underlying character traits. This method is deemed appropriate as this thesis will empirically examine the relationship between generated net flows and the mutual funds’ characteristic variables.

As seen in equation (1), this thesis makes use of different variables as seen in previous literature but has introduced new ones as well. This thesis will make use of lagged variables, in the case of the return, rating and risk, which is also employed by Sirri & Tufano (1998). In their paper, the authors used the lagged variables "Return” & “Riskiness” because there is, according to them, uncertainty about the best period for measuring performance and risk. Questions can arise as to what investors care about, as in, do they care about short-term performance, or is there a greater focus on long-term performance (Sirri and Tufano, 1998, 1597). While the authors use it as a general model for their entire paper, where they change and tweak it for different models, we find the general model useful considering the scope of this thesis and we felt no need to change it greatly.

Furthermore, to make sure that potential data problems are resolved or at the very least displayed and discussed, this thesis will conduct multiple data tests. Firstly, this thesis will conduct a Breusch-Pagan test to see whether heteroscedasticity is present in the models. Secondly, the thesis will conduct a robust regression, so that the problems of heteroscedasticity will no longer be problematic. Robust standard errors are used when the assumption of uniformity of variance is violated, robust standard errors are a remedy for this as they are based on the square of the individual residual (Mansourenia et al., 2021, 347). Thirdly this thesis will use a Wooldridge
test to check for any first-order autocorrelation as the empirical model uses lagged variables (Riveros, 2022). Fourthly, the last test will be a Kernel Density test, which is also known as a kernel density estimation.

A Kernel Density test is a widely used approach to estimate the “underlying probability density function of a dataset (Chen, 2017, 2). Another potential issue is the issue of reverse causality, which will be further considered in the discussion section. Lastly regarding the method used for the data sample, considering that it is widely accepted in the literature that if you run a pooled multiple regression, there should be more than 100 items in the sample, which this thesis has achieved, bearing in mind that there are more than 3,000 observations for the data extracted. With our model, we do not risk overfitting as we deem that we have an adequate sample size. To finish this section, comments must be made regarding the hypothesis of this thesis. This thesis hypothesizes that there are significant relationships between the different variables, Fee, Risk, Rating and R with the ENF, i.e. the estimated net flows. More specifically, while there are possibly different variations of how strong the significance is, this paper believes that the monthly returns and the fee will have the most significant linear relationship with the dependent variable.

5. Results

The purpose of this section is to present the results of the different tests done on the data set and to present the results of our empirical model. Following the presentation of the results there will also be a discussion whereby the results are discussed with the contemporary literature. Firstly, using the Breusch-Pagan test, we could arrive at the that there was heteroskedasticity present in the data set with all the models. To deal with the issue of heteroskedasticity we did a robust regression to get robust standard errors, which are presented below. Another test that was conducted was a Kernel density estimate test to see the normal distribution of the residual, the tests for the different models showed that the residuals were not normally distributed for any of the models. In addition to the Breusch-Pagan/Kernel Density test, this thesis also conducted a Wooldridge test to detect autocorrelation within the data set, given
that this test is an appropriate one for panel data, and that this thesis made use of lagged variables. The test stated that with the lagged variables, there was no autocorrelation present in the data set for all the different models.

### Pooled OLS Regression

<table>
<thead>
<tr>
<th></th>
<th>(Model 1)</th>
<th>(Model 2)</th>
<th>(Model 3)</th>
<th>(Model 4)</th>
<th>(Model 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENF</td>
<td>ENF</td>
<td>ENF</td>
<td>ENF</td>
<td>ENF</td>
<td></td>
</tr>
<tr>
<td>R(_{c1})</td>
<td>-.112</td>
<td>-.104</td>
<td>-.21</td>
<td>-.179</td>
<td>-.279</td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(1.436)</td>
<td>(1.439)</td>
<td>(1.439)</td>
<td>(1.44)</td>
</tr>
<tr>
<td>Fec(_{c1})</td>
<td>-68.916***</td>
<td>-66.411***</td>
<td>-63.486***</td>
<td>-56.179***</td>
<td></td>
</tr>
<tr>
<td>Rating(_{c1})</td>
<td>21.052***</td>
<td>22.397***</td>
<td>19.632***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.071)</td>
<td>(5.12)</td>
<td>(4.835)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td></td>
<td></td>
<td>6.425***</td>
<td>17.122***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.192)</td>
<td>(3.839)</td>
<td></td>
</tr>
<tr>
<td>R(^{AVR})</td>
<td></td>
<td></td>
<td></td>
<td>152.577***</td>
<td>(37.548)</td>
</tr>
<tr>
<td>_cons</td>
<td>6.621</td>
<td>66.245***</td>
<td>-2.628</td>
<td>-131.442***</td>
<td>-434.16***</td>
</tr>
<tr>
<td></td>
<td>(7.963)</td>
<td>(18.348)</td>
<td>(26.954)</td>
<td>(44.335)</td>
<td>(99.666)</td>
</tr>
<tr>
<td>Observations</td>
<td>3331</td>
<td>3331</td>
<td>3331</td>
<td>3331</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0</td>
<td>.006</td>
<td>.009</td>
<td>.009</td>
<td>.014</td>
</tr>
</tbody>
</table>

*Robust standard errors are in parentheses*

*** p<.01, ** p<.05, * p<.1

*Table 2*

<table>
<thead>
<tr>
<th>Breusch-Pagan test</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable: Fitted values of ENF</td>
<td>chi2(1) = 36.11</td>
<td>chi2(1) = 632.79</td>
<td>chi2(1) = 299.09</td>
<td>chi2(1) = 364.98</td>
<td>chi2(1) = 149.11</td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Wooldridge test</td>
<td>F( 1, 61) = 2.920</td>
<td>F( 1, 61) = 2.879</td>
<td>F( 1, 61) = 2.879</td>
<td>F( 1, 61) = 2.879</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>Prob &gt; F =</td>
<td>0.0926</td>
<td>Prob &gt; F =</td>
<td>0.0949</td>
<td>Prob &gt; F =</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0925</td>
<td>0.0949</td>
<td>0.0949</td>
<td></td>
</tr>
<tr>
<td>F-test for joint significance</td>
<td>F(1, 3329) = 0.01</td>
<td>F(2, 3328) = 12.10</td>
<td>F(3, 3327) = 14.91</td>
<td>F(4, 3326) = 11.22</td>
<td></td>
</tr>
<tr>
<td>Prob &gt; F =</td>
<td>0.9382</td>
<td>Prob &gt; F =</td>
<td>0.0000</td>
<td>Prob &gt; F =</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Prob &gt; F =</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Table 3

Now that the different test results for the data have been done, the results of the empirical models can be presented. Table 2 shows the results of a pooled multiple regression according to our empirical model. The different columns in the table are presented to show different specifications whereby the different independent variables are tested towards the ENF (the Estimated Net Flow) and it is displayed in millions. The different tests are titled Model 1, 2, 3, 4 and 5 stating whether the linear relationships are statistically significant or not.

The first column with Model 1, displays the regression test between the lagged monthly return variable and estimated net flow variable, which shows a negative statistically insignificant relationship at the 1%, 5% and 10% levels. We cannot draw any inferential assumptions due to this insignificant parameter. In the second column with Model 2, the previous test is done but now in addition there is the lagged management fee variable added which shows a negative significant relationship at the 1% level. It can be interpreted that there is a negative relationship between the lagged management fee and the estimated net flows, whereby a 1% unit change in the management fee will result in 68.916 million SEK reduction in the estimated net flows. This is a large number, but it should be stated that as previously seen in the Descriptive Statistics from Table 1 the lowest and highest observations for Fee is between 0 and 1.75. This scaling of the variable implies that a standard increase in the management fee of 1% is of great essence for the investors. The lagged management fee also indicates that investors...
will look at the previous month's reported management fee when deciding how to act and invest.

The third column with Model 3 includes all the previous variables but adds the lagged Morningstar rating variable. With this variable, there is a statistically significant positive relationship at the 1% level. It can be interpreted that when the rating of a fund goes up by a 1-star rating, given that there are 5-star ratings available for a mutual fund, the estimated net flows will increase by 21.052 million SEK, according to this model. This is deemed as a large increase, which also holds as being reliable as it makes sense that the rating would play a strong role in the flows of a fund.

The fourth column which shows the results of Model 4 includes, both the lagged monthly return, the lagged management fee and Morningstar rating variables, but also includes the control variable of risk. With this control variable, all the previous p values are still the same, excluding the lagged return, and the control variable is statistically significant at the 1% level. As explained previously, the risk is calculated by looking at the standard deviation of the return for the last three years. The model suggests that an increase in the standard deviation of a mutual fund, computed for the past three years, will result in an increase in the estimated net flows by 6.425 million SEK.

Lastly, the fifth and final column shows the fifth model which adds the last control variable which is the average monthly return over the past three years. Its coefficient suggests that an increase in a mutual fund's average monthly return would lead to an increase of 152.577 million SEK in the estimated net flows. This is a large increase seen as if a mutual fund's monthly returns on average would increase by 1% per unit, more capital would be invested in comparison to how many shares would be sold on a monthly basis. All the tests except the lagged monthly return are statistically significant at the 1% level. With this introduced variable there have been significant changes in some of the previous variables, however most prominent is the risk coefficient which increases by almost 11 million SEK from the previous amount. In the bottom row of the table on page 16, you can see the R-squared, which is at its highest 1.4%. Thus, it is not a high value. In comparison to
Sirri and Tufano’s paper (Sirri and Tufano, 1998, 1599-1618), they reach R-squared values up to 24.1%. One factor that might contribute to this fact is that they introduce more variables to their different regression models. They also transform their variables by taking the logarithm of the values, which usually leads to more linearity since the scale is being brought to the same level.

**Discussion**

In this subsection of section 5, the results will be discussed with the contemporary literature on the subject. As Sirri and Tufano pointed out in *Costly Search* (1998) consumers of equity funds seek out high-performing funds, which is something that we can argue is seen in our results section. Mutual funds with an increase in return will gain eventually a higher estimated net flow. We can see that there is a suggested significant relationship between the Morningstar rating and how much capital will be invested in the specific mutual fund. Sirri and Tufano (1998) also pointed out that fund flows are fee-sensitive, which is something that can also be argued with the results we have, that is, management fees have a strong input into how much money an investor is willing to invest. Similarly, to what these authors have pointed out, it does seem that there is some evidence, on aggregate, that consumer behaviour is rational. While the scope of this thesis could not argue against previously introduced literature such as Carlsson Hauff (2014) that the investors studied are not rational, it would seem at least with the data this thesis is analyzing that investors are leaning more towards the rational side.

Following other literature previously introduced, which discusses how fund flows respond to past performance (Ferreira *et al.*, 2010), our results concur with this as the results show that funds with a strong Morningstar rating will have a strong relationship with the $ENF$. 
Furthermore, what should be discussed is how our data differs from other similar Swedish studies. The work done by Englöv, Marktorp, Nilsson & Nimac studied the relationship between the Morningstar rating and return and argued how equity funds with a higher rating do not have a causal relationship with the return. Our thesis however has tested and received different results, which point out that there is a significant relationship between the regressors and the estimated net flow of a mutual fund. Our data differs from the work done by Englöv, Marktorp, Nilsson & Nimac which could have led to different results. The data that is used by this thesis is different as we have access to a reliant and updated real-time database at Morningstar in comparison to the authors only having access to secondary data from daily newspapers and the internet. Morningstar is an independent source of fund information, and we argue that our data is more befitting the purpose for this subject. As could be deciphered from the results table there is no significant relationship between the lagged monthly return, i.e., the past performance and the estimated net flows, so it would seem that retail investors could evaluate different periods when making their investment decisions. If we were to delve even further into our research, other periods could be introduced to our empirical model. The star rating from Morningstar is a profound metric displayed on multiple mutual fund trading platforms around Sweden. According to our results, investors tend to favour increased development of a mutual funds rating. We deem it important to distinguish the methodology Morningstar uses to compute a mutual funds rating as there are several factors involved (Appendix A). Do all the investors comprehend its function to a full extent?

The issue of reverse causality, meaning that the regressors introduced to our models could be affected by the response variable is likely in theory. Unfortunately, we cannot assess an adequate specific test to rule out or materialize the issue of reverse causality. As a final part of the discussion, we would like to highlight the obvious fact regarding potential biases influencing our estimates. The reasons behind a mutual fund’s performance and the contributing factors for an investor’s preferences are presumably enormous in numbers and therefore too large for the scope of this thesis. We have not addressed what other externalities there might exist in a society that drive the net flows of a mutual fund every month. We deem
that this area of science needs more research than what we are capable of conducting for the purpose of this thesis. With this discussion, we have been able to address to some degree the research question which asked how does Swedish mutual fund savers invest towards Swedish domiciled mutual funds in relation to the management fee, Morningstar rating, risk level, and the monthly return studied during a ten-year time span between the years 2013-2023?

6. Conclusion

In conclusion, this thesis has attempted to understand the complicated scene of the Swedish fund market and Swedish investor behaviour towards it. This thesis started by first introducing the aim and research question, whereby the purpose of our thesis was to analyse Swedish retail investors' investment behaviour in a Swedish context. After this, the previous literature was delved into and discussed showing the overall academic context of the subject studied and how it will be used as inspiration for this thesis. The data section preceded this, whereby the data of the thesis was presented giving a descriptive table of the number of observations. Subsequently, the following section shows the empirical method, its structure and how it would attain the desired results. Successively through analysis and exploration, several insights were made in the results section. By using pooled OLS regression, with our empirical model, testing different mutual fund characteristics against the estimated monthly generated net flows we could interpret that the investors prefer low-cost mutual funds that take higher risks at the same time as they are performing well seen as relative risk-adjusted return performance metric conducted by Morningstar (Morningstar rating). Unfortunately, we could not assess whether past monthly returns affect the investor's investment decisions.

There could be other periods of interest for the investors when evaluating their subscriptions and redemptions of mutual fund shares. As first future research on this subject could be about extending the period of which this study is looking to
analyse changing/evolving trends and patterns when it comes to investor behaviour. Secondly, as the previously introduced literature *Costly Search* studies, it would be an interesting future research subject to discern the impact of marketing and media strategies done by the different firms and how they may affect the estimated net flows of mutual funds in Sweden, which brings up another future research topic whereby one could study the behaviour of investors in Sweden by doing a qualitative study instead by using surveys or interviews, to get a more personal grasp of consumer behaviour regarding mutual funds and why people invest in some funds instead of others.

**Special Thanks**

In the past 10 years, we have seen great changes in the mutual fund market in Sweden, and no doubt there will be even greater changes in the coming 10 years. We want to say a great thanks to all those who have made it possible to complete this thesis as well as a thank you to Morningstar for providing the data to make it possible. More specifically, we would like to say thanks to our supervisor Annika Alexius who has given great guidance and helpful notes during this term. Lastly, we also want to express our gratitude and thanks to Anders, Andreas and Anton whose unwavering support was invaluable throughout this research journey.

**References**


Appendix

A: Data

These data points are taken from the data definition section in the Morningstar Direct app.

- We get the Estimated Net Flow from Morningstar Direct, or rather the “Estimated fund-level net flow-comprehensive (monthly)” as its official name at Morningstar Direct.
- The Management Fee is defined as the amount the fund estimates will be charged for the costs of its management, which includes investment services but does not include distribution fees or transaction fees during the following fiscal year. The data is reported as a percentage of average net assets.
- Morningstar Rating is described with a rating system between 1 to 5 stars and to determine a fund's star rating for a given time, the fund's risk-adjusted return is plotted on a bell curve. If the score of the fund is into the 10 percent of its category, it will receive five stars which is the highest, if it falls to 22.5% it receives 4 stars which is above average, if it is at 35% it will be ranked at average, 3. If it is lower than that at 22.5% it receives 2 stars which is below average and lastly, if it
is at the bottom 10% it will only get 1 star which is the lowest.

- **Monthly Return** is the calculation of the total return decided each month by taking the change in the monthly net asset value, reinvesting all income and capital-gains distributions during that month and this is divided by the starting net asset value. Reinvestments are made using the actual reinvestment of the net asset values, and daily payoffs are reinvested monthly. Unless it is noted, Morningstar Direct does not adjust total returns for sales charges, and the total returns do account for administrative, 12b-1 fees, management and other costs taken out of fund assets.