The Law of One Price? Price Dispersion & Persistence in the Danish Mutual Fund Industry

An Undergraduate Thesis in Pricing Management

by Patrick Natarajan Larsen*

May, 2019

Keywords: Mutual Fund, Price Dispersion, Price Persistence, Heterogeneity, Search Costs

*Please do not refer to the paper without formal acceptance from the author. The thesis is written with the Department of Opeartions Management at Copenhagen Business School. A sincere thank you to Morningstar for providing me with data for the empirical analysis and my colleagues at Morningstar and Copenhagen Business School for competent sparring. Furthermore, thank you to my supervisor for help with model specification and the overall structure of the paper. Any mistakes are of the sole responsibility of the author.

Contents

List of Figures

List of Tables

1 Executive Summary

In recent years, mutual fund fees has been scrutinised by researchers and media alike[i]. There is one anomaly however, which has not received extensive attention - price dispersion. The mutual fund fees are extremely dispersed, despite the apparent homogeneity in returns. So are investors paying for another feature of these funds, or is the dispersion caused by other underlying characteristic of the industry? Combining theoretical approaches of statistics, pricing management and industrial organisation, this paper aims to develop some of the first steps in analysing this complex anomaly in a Danish setting.

I develop a model which is gradually complicated throughout the paper in order to understand the pricing decision from a managerial standpoint and the prices effect on consumers (investors) in the industry. The analysis finds that price dispersion exists and persists with the assumption of homogeneity in fund characteristics and that fixed-income portfolios has the lowest dispersion in this regard. By loosening the assumption of homogeneity in fund characteristics however, it is identified that this assumption is too restrictive on the industry as a regression of prices on appropriate fund characteristics explain approximately 43 percent of the price variation. Counter-intuitively, the analysis finds that larger risk adjusted returns, historically has led to lower prices, which questions the investors investment decision. Furthermore, it finds that larger, older and funds with monetary switching costs, has a mark-up on prices.

Lastly, I analyse the unexplained variation in prices from the statistical model by acquiring the residuals and utilising them as a heterogeneity controlled price. The analysis finds that fund managers have historically made pricing decisions predicted by a spatial price dispersion model. More specifically, funds that utilise a pricing policy with unjustifiably large (or small) prices, tend to continue to price in an unjustifiable manner. These findings suggest that investors are not able to differentiate between these funds without incurring significant costs of search, despite the existence of several clearinghouses and the efforts of regulators to increase investor transparency.

[i]Cost, fees and prices are used interchangeably throughout this paper. Please note that all of it refers to the annual price that investors pay to have their money managed by a mutual fund

2 Introduction

Over the past decades, the market for mutual funds has experienced fundamental changes. In 1975, the late John Bogle introduced the index funds in order to challenge the leading investment vehicles, hedge funds, with a low-cost alternative (Nocera, 2019). Technological developments has made it easier to track returns on a daily basis while both the Securities and Exchange Committee (SEC) and the European Parliament have passed regulations on increasing investor transparency (White, 2013), (Regulation (EU) No 1286/2014). The changes have all led to increased inflows to mutual funds of all kinds over the same period (Anadu, Kruttli, McCabe, Osambela and Shin, 2018, p. 2). In a Danish setting, the situation is no different. The tendency to invest in mutual funds as an alternative to the classic approach of bond and equity investing is more prevalent than ever and the Danish mutual fund industry has over 850 billion assets under management at the time of writing (Morningstar Direct, authors calculations). With increased inflows and marketing to the private investor, it is more important than ever to understand the industry and its implications on investor welfare. In the light of the index funds, the price of annual management has naturally been scrutinised, but few have analysed prices explicitly. Instead, most established literature has been developed around the comparison of risk-adjusted returns, finding that the majority of fund managers fail to outperform their respective passive peers consistently. (Anadu et al, 2018, p. 8), (Schaefer, 2018).

These findings however, provide more questions than answers: How can such a large industry, where no managers consistently outperform the others, exist in equilibrium? And even more importantly: How can prices differ across funds, if no funds consistently outperform the others? As the research on the area, to the best of my knowledge, has been rather scarce, this paper aims to develop some of the first thoughts on this apparent pricing anomaly, in order to identify the underlying conditions of a pricing equilibrium in the Danish mutual fund markets. Furthermore, I seek to establish which implications such an equilibrium has on pricing managers, investors and financial regulators. The paper works around the following research question:

- **RQ**: Which conditions justifies price dispersion in the Danish mutual fund industry and what are its implications on stakeholder welfare?

The analysis is structured in order to gradually complicate the model for pricing dispersion and allow for variation in the underlying market conditions. The first part of the paper, section [4.2-4.3], analyses the, until now, undocumented price dispersion in the Danish mutual fund market in order to understand the extent of the dispersion. The model is as simple as possible assuming homogeneous products and then analyses the extent and persistence of price dispersions over time and within investment strategies. The second part of the paper, section [5] allows for product heterogeneity, and thus, a statistical model is developed in order to capture the cause of pricing variation. The last part of the paper, section [6] serves as a thorough analysis of the unexplained variation of this statistical model. This is done in order to capture the welfare implications on stakeholders in the industry.

3 Literature Review & Theory

The research on price dispersions has been a topic of interest for decades, and was motivated by the consistent empirical falsification of the law of one price in seemingly homogeneous markets (Belleflamme and Peitz, 2010, p. 158). Stigler (1961) was one of the first authors to develop a mathematical framework for this phenomenon and suggests two things relevant for this paper: (1) Unless a market is completely centralized, no one person, will know all the prices of all sellers, (2) Even though marginal product heterogeneity can lead to price dispersions, it would be ”meta-physical, and fruitless, to assert that all dispersion is due to heterogeneity” (Stigler, 1961, p. 214). There is a general consensus for Stiglers arguments in the industry and therefore, it shall serve as a theoretical framework for this paper. Ad (1), I assume that, at least some part of the consumer base, is uninformed about prices in the industry. Ad (2), an establishment of product heterogeneity in the second part of this paper, is not expected to serve as an adequate and final explanation of price dispersion. More specifically, unless the product heterogeneity explains 100 percent of the price variation over time, it is necessary to study the nature of the remaining price dispersion. I do this, consistent with, among others, Lach (2002), Belleflamme and Peitz (2010), Varian (1980) and Baye, Morgan and Scholten (2005) by matching the remaining dispersion with the two most prevalent price dispersion models: spatial and and temporal price dispersion (Varian, 1980).

3.1 Spatial Price Dispersion

Spatial price dispersion was given its name by Varian (1980), however the theoretical framework for such dispersion was developed years before. Salop and Stiglitz (1977) provide such a framework. In their model, consumers are different when it comes to the cost of acquiring information in the market. In such a model, some stores are able to persistently sell their products at a price above the market average, even for perfectly homogeneous products. In a regular perfectly competitive framework, such a supplier should be unprofitable and therefore lower prices or shut down rather quickly. In a framework of spatial price dispersion however, such suppliers are able to remain profitable, as consumers have fundamentally different information and learn at different speeds. Therefore, high-cost firms can remain high-cost, and will in an equilibrium supply to the uninformed consumers exclusively (Belleflamme and Peitz, 2010, pp. 158-163)[ii]. In other words, investors search costs are different but all have the ability to acquire information.

3.2 Temporal Price Dispersion

The spatial price dispersion model is rather intuitive, as it is evident that the cost of information is empirically different for consumers. Varian (1980) challenges this approach however, arguing that in a long run equilibrium, all consumers should be able to differentiate between high and low-cost firms. Instead, he provides the framework for temporal price dispersions, where firms consistently randomise over prices. Randomisation of prices, deems it impossible for both informed and uninformed consumers to identify the high and low-cost suppliers in each period (Belleflamme and Peitz, 2010, pp. 158-163), Varian (1980).[iii]

3.3 Heterogeneity

The models above fundamentally assumes the homogeneity of products. As motivated by Stigler (1980) however, it is a rather simplifying assumption that products are homogeneous. Therefore, in order to apply the theoretical framework described, it is necessary to account for product heterogeneity. Heterogeneity among seemingly homogeneous products have not only been studied

[ii] See Stiglitz (1977) or Belleflamme and Peitz (2010, pp. 158-160) for formal proof of such an equilibrium
[iii] See Varian (1980) or Belleflamme and Peitz (2010, pp. 160-163) for formal proof of such an equilibrium

within the field of industrial organisation, but within operations and marketing management as well. Hinterhuber (2016) recently studied the homogeneity assumption in the automobile market, suggesting that such an assumption is a self-fulfilling prophecy leading to actual non-differentiated products and inferior profits.

Nagle and Müller (2017), propose the Economic Value Model as a way to quantify the positive and negative characteristics of a product. I utilise this framework in a slightly altered version, computing differential values as econometric estimates. Therefore, assuming no reverse correlation in the statistical model, a positive parameter estimate shall refer to a positive historic differential value, while a negative estimate shall refer to a negative value.

Lach (2002) is one of the first authors who control directly for product heterogeneity through a statistical model and study the post-estimate price dispersion. The methodology works around explaining prices in a panel data estimation:

$$p_{it} = \mu + \alpha_i + \delta_t + \gamma_c + \lambda_T + \epsilon_{it}$$

Where the explanatory variables are fixed effects variables. Lach's study is computed on consumer goods, namely chicken, coffee, flour and refrigerators. Therefore, the data structure and thus the statistical model differ from a similar approach on the mutual fund market. More specifically, Lach can utilise a fixed effects model as effects are generally time-invariant in his data set. Although the mutual fund industry requires another statistical model, I do utilise his approach to study the heterogeneity controlled price dispersion. By acquiring $\hat{\epsilon_{it}}$, Lach idenitifes whether the market exhibits spatial or temporal price dispersion by studying the residuals over time. As residuals are defined as the discrepancy between the realisation and expectation from the model, it is possible to identify whether the store charges prices which are larger than the model justifies. More specifically, firms with large (numeric) residuals utilised a pricing policy which was unjustified by the statistical model. Standardising these residuals over time allows for the study of intra-distribution movement, i.e. whether firms that have upper-quantile residuals (high-cost firms), remain in the upper-quantile over time.[iv]

As mentioned in the introduction, the research on price dispersion for mutual funds has been rather scarce. Horatscu and Syverson (2004) are the first to study the phenomenon, controlling for

[iv] See section [6.2] for a thorough explanation of this methodology

heterogeneity and studying dispersions on the American market for index funds. Both the statistical methodology and the latter dispersion analysis differs quite substantially from my approach. They do find that the price dispersions are a function of search costs, product and investor heterogeneity.

Ianotta and Navone (2008) and Cooper, Halling and Yang (2016) both control for heterogeneity in a similar fashion to this paper while the latter paper explicitly studies the standardised residuals like Lach (2002). I shall apply a similar statistical model to both papers, namely the Fama-Macbeth procedure, which is described in detail in section [5]. Both papers do find consistent price dispersions, even after controlling for heterogeneity and Cooper et al (2016), document the development of intra-distribution dynamics.

4 Data & Methodology

4.1 Data Description

The paper adopts a rather positivistic approach, consistent with the established literature within this field (Lach, 2002), (Ianotta & Navone, 2008), (Cooper et al, 2016). Therefore, the data will primarily be of a quantitative nature (Holm, 2016, pp. 22-39). Since the goal of this paper is to analyse price dispersions in a historic setting, I acquire data on all changes of price and relevant control variables for Danish mutual funds since 2010. The data is by courtesy of the authors workplace, Morningstar, which is one of the largest providers on mutual fund, equity and fixed income data and research in the world. The data is provided to Morningstar by the mutual funds themselves, or through an intermediary, and are under regulation from the European Union which makes the data an internal secondary data source and naturally ensures its validity (Stavnsager, Østergaard and Beckmann 2006). The Danish mutual fund industry is not nearly as large as the industry on a global scale and consists of 1.384 at the time of writing compared to approximately 29.000 funds in the US alone. Therefore, it is a natural concern that the sample is not large enough to produce reliable results (Stavnsager et al, 2006) and satisfaction of statistical limit theorems. I work around this by adopting a panel data on a quarterly basis which significantly increases the amount of data points and ensures the statistical results. The panel data is unbalanced as several funds are opened and closed throughout the eight year period. This complicates the econometric approach and is formally addressed in section [5].

The industry is compiled of many different kinds of funds. As the purpose of the paper is primarily to study the effects on private investors, I solely consider open-end funds which means that institutional funds, money-market funds and the like are omitted. Therefore, the amount of funds is decreased to approximately 1.000. The preliminary estimates in section [4.2] consider all these funds, but for the statistical analysis in section [5], i require funds to have three year history. This is done to ensure that the results are consistent and are not a result of a market condition or randomness.

Since the main goal of the paper is to study the fee structures of Danish mutual funds, the choice of variable is of obvious importance. There is no established consensus in the literature in regards to the exact variable used which may be the cause of different findings. The majority of the established literature uses the Total Expense Ratio, which is defined by Morningstar as: ”the percentage of assets deducted each fiscal year for fund expenses, including 12b-1 fees, management fees, administrative fees, operating costs, and all other asset-based costs incurred by the fund” (Morningstar Glossary, 2019). The European Union recently released regulation on the area of mutual funds with regulation (EU) No 1286/2014 of the European Parliament and of the Council. Here it is specified that mutual funds are obligated to enclose ongoing charges and performance fees separately. As Denmark falls under the regulation of EU, I use the **ongoing charge** data point when estimating price dispersions. The ongoing charges is thought to be the most accurate representation of the costs investors actually incur when investing in a mutual fund and is defined: ”The Ongoing Charge represents the costs you can reasonably expect to pay as an investor from one year to the next, under normal circumstances. Many investors will be used to looking most closely at the Annual Management Charge, but neither this charge nor the Ongoing Charge includes the performance fees incurred so neither is perfect. However, the Ongoing Charge does represent a more accurate cost of fund ownership as it encompasses the fund's professional fees, management fees, audit fees and custody fees.” (Morningstar Glossary, 2019). In a Danish setting, the ongoing charges generally consist of three types of costs (Henriksen, 2010):

- **Holding commission**. Fees paid to banks for selling and advising in regards to selling the shares of the fund

- **Portfolio commission**. Fees paid directly to the portfolio managers of the fund. Please note that this is the 'flat commission' and is thus not performance dependent confer the regulation discussed above.

- **Administrative costs**. Fees to the administration of the fund company which reflects everything from rent to auditing fees.

It should be noted that some funds change prices within quarters as a result of spontaneous price changes or a lagged report to Morningstar. I acquire data on all price changes that the sample has ever made, and define the four critical cut-off points corresponding to quarter closes. Therefore, if a fund changes a price mid-quarter, it won't be reflected in the data before the end of the quarter. This should not be a problem, due to the rather infrequent price changes and due to the annualised nature of ongoing charges in general. Furthermore, some funds provide annual reports and thus ongoing charges with a substantial lag. I have disregarded those lags, making it possible for a fund to have a fee which slightly deviates from actual fees.

4.2 Preliminary Estimates

As previously mentioned, the analysis takes an approach of gradual complication. The simplest model possible, is done by simply analysing the cross-sectional price dispersion at some point in time. In the preliminary analysis I rule out any time variation, hence the kernel density depicted in figure one is a cross sectional distribution from the last trading day of 2018. I define cross sectional price dispersions consistent with Lach (2002) among others, as the discrepancy from the average price of the product i.e.:

$$\text{Price Dispersion}_t = (P_t - \bar{P}) * 100 \qquad \text{(i)}$$

Where $\bar{P} = T^{-1} \sum_{i=1}^{T}$,which is a simple arithmetic average. This standardises the variable in order to make accurate density representations. The multiplication factor expresses the dispersions in *basis points* as is standard in the financial industry. For the law of one price to uphold, the distribution is expected to be tightly distributed around the mean which has been standardised to zero. From figure 1 it is obvious that dispersions are too large to assume that the law of one price is upheld. The skewness is significant and the tails are long especially towards the upper part of the density.

Even though that the law of one price is rejected for the mutual fund industry in general, it is necessary to control for investment strategies (Horatscu and Syverson, 2004) (Ianotta and Navone, 2008). Mutual funds that pursue strategies that invest in riskier assets will most likely have larger fees. As the sample includes both equity and fixed income strategies in both developed and emerging markets, it is to be expected that prices significantly differ.

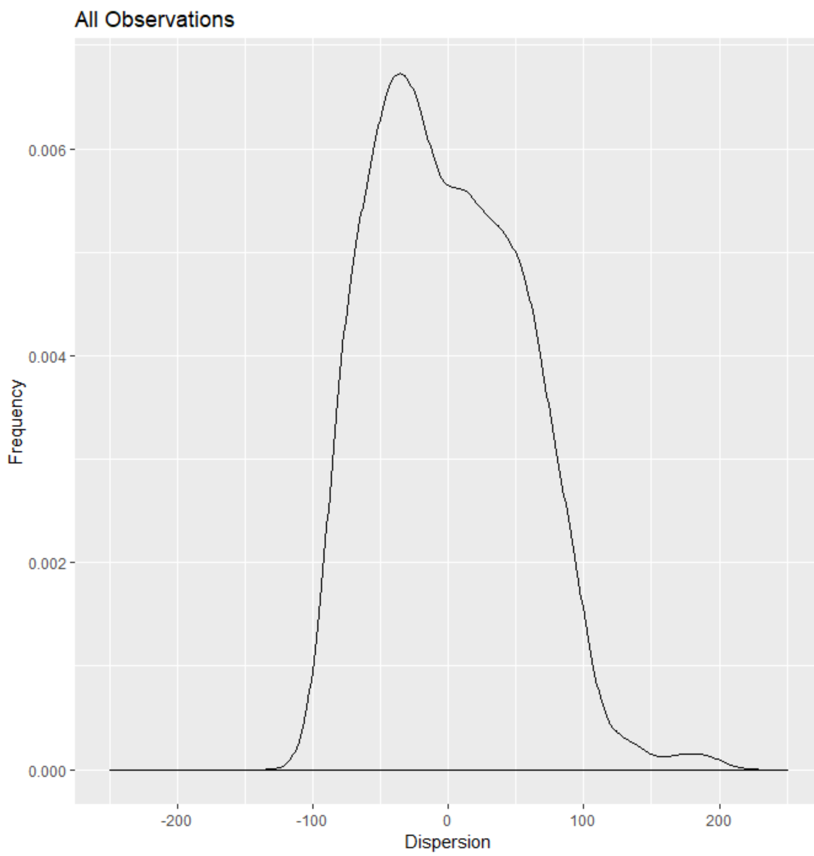

Figure 1: Preliminary Kernel Densitiy

One may argue that this is an argument for a strict rejection of the thesis of homogeneous products. I consider the issue by simply creating three groups of mutual funds: equities, fixed income and mixed allocation strategies. All european mutual funds which fall under the regulation of the European Parliament, must produce central investor information documents, commonly knowns as KIIDs. These documents hold information in regards to the funds investment strategy, hence the grouping is rather simple. In general, equity funds primarily invest in equities, fixed income funds primarily invest in fixed income while the mixed allocations are allocations in between. The standardisation approach has been changed in order to account for this difference, and thus a funds price dispersion is naturally calculated as the deviation from the average price of the investment strategy in which it belongs.

Table 1: Summary Statistics by Investment Strategy, 2018-12-31

Strategy	N	Mean	St. Dev.	Min	Max
Mixed	186	0	48,48	-121,86	168,14
Fixed Income	339	0	34,63	-58,53	85,46
Equity	456	0	56,42	-104,382	180,61

Note: The table provides descriptive statistics for the three categories on the last trading day of 2018.

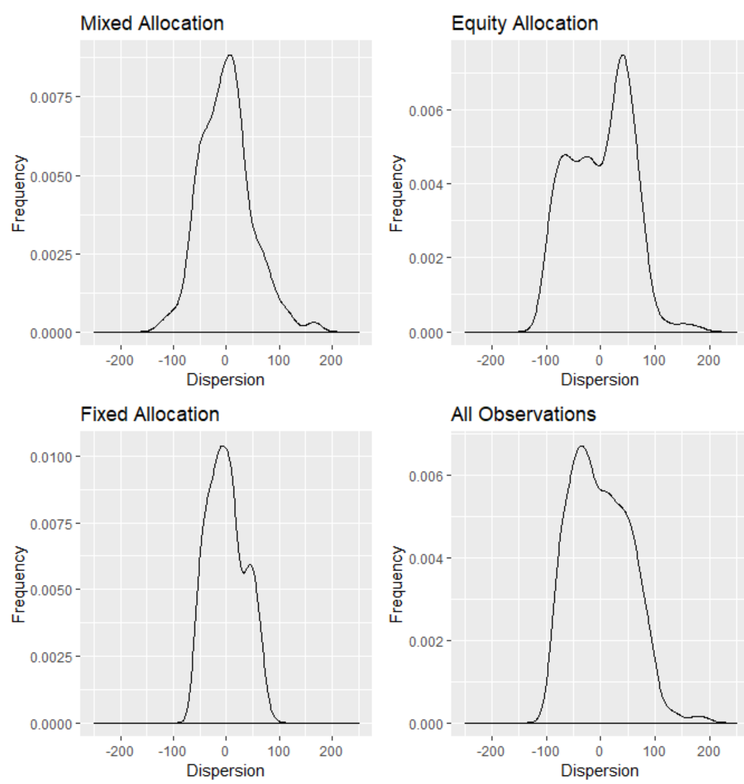

Figure 2: Densities side by side 2018Q4

From the table and figure above, it is evident that there is still a considerate price dispersion among funds that pursue the same investment strategy, but the deviations are considerably smaller than for the full sample. The difference between the minimum and maximum points are largest for the mixed strategy, which may seem puzzling when it is related to the smaller standard deviation. Most interesting, both the standard deviation and the range between the minimum and maximum points are lowest for the fixed income strategy. This is consistent with Lach (2002) and Pratt, Wise and Zeckhauser (1979) who find that price dispersions are lower for goods with a lower mean price as is the case with the fixed income fees. The fact that fixed income funds have the lowest mean price is rather intuitive, considering the nature of the investment vehicle. Fixed income portfolios with high credit ratings are often referred to as risk-free. Conclusively, it is evident that Danish mutual fund market exhibits deviations from the law of one price even after controlling for investment strategies.

4.3 The Time Variation of Price Dispersion

Complicating the model once again, I study the phenomenon over time. After all, it is possible that the industry is in a transitional phase towards an equilibrium price. Therefore, I study the development of the price dispersions within the three categories over the eight year period, 2011-2018. As funds are frequently merged, closed or liquidated, it is important to stress that the amount of observations differ from quarter to quarter. This is not an issue as the amount of observations for each quarter makes the development statistically viable and rejects the case of randomness. Once again the variable is altered to exhibit comparability traits over time. Baye et al (2006) suggests that price dispersion over time, is to be measured in variance or standard deviation, trivially:

$$\sigma_p^2 = var(p) = \frac{1}{T-1}\sum_{i=1}^{n}(x-\bar{x})^2 \tag{ii}$$

$$\sigma_p = SD(p) = \sigma^{2\frac{1}{2}} \tag{iii}$$

I adopt the latter approach of standard deviation. The computations are summarised in table 2 below.

From table 2 the story of an industry in steep progression is confirmed as the number of funds is increasing for all strategies over the eight year period. Both the equity and fixed income strategies

Table 2: Summary Statistics by Investment Strategy, $2011_{Q4} - 2018_{Q4}$

	Mixed		Fixed Income		Equity	
	N	Std. Dev	N	Std. Dev	N	Std. Dev
2011Q4	123	0,5691	171	0,4417	274	0,4933
2012Q4	116	0,5815	174	0,4309	281	0,5148
2013Q4	117	0,5452	201	0,3997	289	0,4906
2014Q4	107	0,4892	213	0,3997	299	0,4924
2015Q4	127	0,4297	250	0,4321	349	0,6004
2016Q4	210	0,5968	288	0,3964	414	0,5613
2017Q4	178	0,4632	336	0,3541	435	0,5552
2018Q4	186	0,4848	339	0,3463	456	0,5642

Note: The table provides descriptive statistics for the degree of price dispersion in the fourth quarter of each year.

experience an approximate 100 percent increase in the number of funds while the mixed strategy is lower. The standard deviations however, provides an ambiguous result. For the *mixed* and *fixed income* strategies, the standard deviations decrease significantly which may suggest that dispersions are simply in a transitional phase. The equity funds on the other hand, experience an increase in standard deviation and thus dispersion. The results are puzzling, especially when considering the fact that the equity strategy is almost as large as the mixed and fixed income strategies combined. When that is said, the larger dispersion may simply suggest that funds pursuing the investment in equities has seen a development of more heterogeneous products over the eight year period while the mixed and fixed income has experienced the opposite. I shall not test this empirically. Instead, I utilise it as motivation for further analysis: Price dispersion is definitely present in the Danish mutual fund industry, even after controlling for time and investment strategies.

5 Identification and Estimation Strategy

As it has been argued by Stigler (1961) and Hinterhuber (2016), it would be delusional to expect products to be perfectly homogeneous. Therefore, I allow product heterogeneity by conducting a formal statistical analysis in order to explain price variation within the mutual fund industry. In the following, the strategy for estimation and variable specification is outlined. Furthermore, I touch upon the statistical assumptions required for the utilisation of such a strategy.

As discussed in [3.1] the data is a panel of 32 quarters (T) and a varying number of funds (N). Determining the correct estimation strategy is highly dependent on the observed data characteristics. Due to obsoletion and inception of funds, the data takes the form of an uneven panel. Despite the complication of the statistical strategy that this data incurs, the analysis shall adopt an uneven panel due to two reasons (1) Balancing the panel out, i.e. removing the funds that were opened or closed during the period, results in a significantly smaller sample. Although the severity of the reduction is not enough to compromise the satisfaction of limit theorems, the larger sample is obviously preferable. (2) The process of removing such funds would result in a significant survivorship bias as only funds that have managed to stay open over the past eight years would be included in the sample.

5.1 Standard Error Bias

Petersen (2006) suggests that the tendency to produce biased estimates is prevalent in financial literature, documenting that 42 percent does not directly correct for their estimates. Therefore, a foreword on this is required. There are two general forms of dependence which are most common in finance applications Petersen (2006, p. 436): time series dependence and cross-sectional dependence. Time series dependence is a case of autocorrelation, i.e. that residuals, for a given fund, show persistence over time. Cross-sectional dependence is the case of correlation of residuals across different firms in a given year.

5.1.1 Time Series Dependence

The tendency to correct for time series dependence is very sporadic in most established financial literature (Petersen, 2006, pp. 436-437). The tendency to disregard such dependence for certain areas of finance may not be such a bad idea. For a panel to exhibit time series dependence, we require the residuals for a given firm, i, to show dependence through time. Therefore, when studying variables such as asset returns which historically has shown no dependence over time, disregarding autocorrelation may result in a simpler model (Cochrane, 2005, p. 248). Adopting this assumption for mutual fund fees however, seems less plausible. This would lead to the assumption that the pricing decision for a fund, i, in a given quarter and hence the residuals, would be independent with the pricing and thus the residuals in any other quarter. Therefore, the time series dependence requires attention.

5.1.2 Cross-Sectional Dependence

Cross-sectional dependence, compared to time series dependence, is less often neglected for financial data as the motivation for its existence is easier. As business cycles are generally present, it is easy to see that in any given quarter, the expected price, return, investment or acquisition for a firm is expected to be correlated with other firms, especially when considering rather homogeneous firms.

5.2 Model Alternatives

In order to address the above biases, and the structure of the data, a broad range of approaches could be adopted. Petersen (2006) finds that the most common method of correct estimation is the Fama Macbeth Estimation. Estimating by regular ordinary least squares to a panel:

$$y_{it} = \beta' x_{it} + \epsilon_{it}$$

Produces biased residuals unless post-estimations are corrected. The Fama-Macbeth approach is so common, because it produces unbiased estimates as long as the data solely suffers from cross-sectional dependence, and thus is applicable to most financial data (Cochrane, 2005, p. 247). The literature which studies mutual fund pricing is no different and similarly utilises this approach

(Ianotta and Navone, 2012) , (Cooper et al, 2015). Therefore, I adopt this estimation strategy in order to strengthen the direct comparability. Both Petersen (2008) and Cochrane (2005) find that the Fama-Macbeth procedure is preferable due to its simplicity and applicability on financial data, and that it produces unbiased estimators as long as the data does not exhibit time series dependence[v]. Although I argue for the existence of time-series dependence, I utilise the Fama-Macbeth approach, and explicitly correct for time-series dependence by computing Newey-West standard errors (Newey and West, 1987).

5.2.1 Fama-Macbeth Procedure

The Fama-Macbeth technique is a two-step regression procedure firstly presented by Eugene Fama and James Macbeth (1973). In the first step, cross sectional parameter estimates are computed in each time period as so:

$$Y_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + ...\beta_k X_{ki,t} + \epsilon_{i,t} \tag{iv}$$

Where X_j is a list of dependent variables, $Y_{i,t}$ is the independent variable, ongoing charges and $\epsilon_{i,t}$ is the unexplained variation. Estimating a yearly cross section yields a series of equations:

$$Y_{i,2011_{Q1}} = \beta_0 + \beta_1 X_{1i,2011_{Q1}} + \beta_2 X_{2i,2011_{Q1}} + ...\beta_k X_{ki,2011_{Q1}} + \epsilon_{i,2011_{Q1}} \tag{2011Q1}$$

$$Y_{i,2011_{Q2}} = \beta_0 + \beta_1 X_{1i,2011_{Q2}} + \beta_2 X_{2i,2011_{Q2}} + ...\beta_k X_{ki,2011_{Q2}} + \epsilon_{i,2011_{Q2}} \tag{2011Q2}$$

.
.
.

$$Y_{i,2018_{Q4}} = \beta_0 + \beta_1 X_{1i,2018_{Q4}} + \beta_2 X_{2i,2018_{Q4}} + ...\beta_k X_{ki,2018_{Q4}} + \epsilon_{i,2018_{Q4}} \tag{2018Q4}$$

In the second step, time averaged parameter estimates are computed by a simple arithmetic mean:

$$\hat{\beta}_{FM} = T^{-1} \sum_{t=1}^{T} \beta_t = T^{-1} \sum_{t=1}^{T} \left(\frac{\sum_{i=1}^{N} X_{i,t} Y_{i,t}}{X_{i,t}^2} \right) \tag{v}$$

[v]See Cochrane (2005, pp. 248-251) for a formal proof. For a simulation based proof, see Petersen (2008).

Computing standard errors and deriving significance estimates is then simply computed by utilising the sum of the differences between Fama-Macbeth estimates and the yearly OLS estimates (Petersen, 2008)

$$\sigma^2(\hat{\beta}_{FM}) = T^{-1} \sum_{t=1}^{T} \frac{(\hat{\beta}_t - \hat{\beta}_{FM})^2}{T-1} \qquad \text{(vi)}$$

Which are finally corrected by computing Newey-West standard errors.

5.3 Model Specification

In the following, a specification of the independent variables and their correlation is considered. I propose three different models: (1) Solely addresses monetary gains in terms of performance (2) adds fund company structure and portfolio composition while (3) adds switching costs.

5.3.1 Performance

First of all, the model should correct for direct monetary return in accordance with the EVE model (Nagle and Müller, 2017, pp 29-35). From a mutual fund standpoint, this is included in returns. It is rather implausible to suggest that fund managers manage prices in a given quarter as a function of the returns in that quarter, especially since returns themselves are stochastic in nature. Therefore, I lag the returns three years in order to allow for funds to alter prices as a function of past return rather than of current returns. Furthermore, the lags are consistent with the established literature on the topic such as Ianotta and Navone (2008), Cooper et al (2016) and Hortacsu and Syverson (2004). I express monetary returns in terms of **alpha** which measures the excess return of the fund, compared to a benchmark. Alpha is commonly referred to as being the end goal for most actively managed funds and thus is unavoidable in this study. I define alpha consistent with Morningstar as: "The difference between a portfolio's actual returns and it's expected returns, given its level of risk as measured by beta". Let $\bar{R}$ be defined as a funds monetary return over a given period t:

$$\bar{R} = \frac{\text{Price}_n + \text{Dividends}_{n-0} - \text{Price}_0}{\text{Price}_0} \qquad \text{(vii)}$$

Then a funds excess return, $\bar{R}^e$ is computed by measuring the excess return relative to a government bond of the same length:

$$\bar{R}_i^e = \bar{R}_\text{i} - R_\text{i} \qquad \text{(viii)}$$

Where i represents the length of the period. Similarly, let the excess return of a benchmark, $\bar{B}^e$ be defined:

$$\bar{B}^e = R_{Bt} - R_F \tag{iix}$$

Finally, define β as the systematic risk of an investment, as standard in financial literature:

$$\beta = \frac{Cov[R_j, R_m]}{Var(R_m)} \tag{ix}$$

Where j is the return of the portfolio and m is the return of the market. Then define alpha:

$$\text{Alpha}_Q = \bar{R}^e - \beta\bar{\beta}^e \tag{x}$$

Quarterly alphas are lagged just as returns confer the explanation above which means that every observation is lagged 12 periods, corresponding to three years. Large alphas indicate better performance than their respective benchmark. The expected sign of the Alpha parameter is intuitively positive. The willingness to pay for investors is expected to be a positive function of alpha. It is thus expected that fund companies realise this, and add a mark-up to their product. Conversely, funds companies with high alpha may reward the investors with lower prices when performance is good.

5.3.2 Fund Structure

General portfolio composition is considered through the amount of **holdings** at fund level. I include the number of holdings for a fund to address the diversification of the portfolio which is generally considered to be a desired feature. The holdings are measured as the total amount of holdings, both long and short, for a given fund in a given quarter. The expected sign of this is trivially positive.

Define the variable **fund size** which measures the size of the fund as assets under management in Danish kroner. The size of the fund naturally implies a larger market share which tends to indicate price setter behaviour. Therefore, an argument for a positive impact on price setting behaviour would be for the funds to strategically set mark-ups on products due to their sheer power in the market. Conversely, the mutual fund industry possesses a considerate potential for exhibiting economies of scale. The marginal costs of ”producing” another unit are theoretically low, if not non-existent. Managing more money does not require any additional labour power nor capital investments which are the two most common components of costs in a production. Therefore, investment managers are in a position of potentially large economies of scale. This can be challenged in two ways: (1) Mutual funds with large assets under management may feel obligated to make a more thorough analysis of quantitative and qualitative investment strategies which may result in additional employees and thus higher labour costs (2) Investing more money has larger transaction costs. The effect is low however, considering the diminishing costs which is in place on most brokerage platforms. If economies of scale are as high as their theoretic potential, fund size may have a negative impact on price setting as large funds may seek to simply undercut competition in order to gain the additional revenue. It is not the goal of this paper to analyse economies of scale hence I will lean on the general consensus that there are considerate economies of scale in the mutual fund industry with a very large saturation point (Haslem, 2017).

Lastly, the fund structure consists of **fund age** which is measured as months since the inception date of the funds *oldest share class*. More months in the market is expected to have a positive differential effect as they have had a longer time to prove track records and to gain investor recognition. Conversely, once again an argument can be made in regards to the economics of scale as described above. This produces a potential issue, as older funds intuitively shoulder have more assets under management. I address the correlation explicitly in section [5.3.4].

5.3.3 Switching Costs

Switching costs is an often overlooked variable for one reason: it is quantitatively complex. The measure is naturally important, and is mentioned as one of the more obvious explanations of price dispersion by Belleflamme and Peitz (2010, p. 167). I proxy switching costs with the measure *deferred load* which measures a funds: "back-end sales charge (which) is imposed when an investor redeems shares" (Morningstar Glossary, 2019). Therefore, it is the cost that investors incur when selling their shares in a fund. Despite the proxy's relation to switching costs, it notably only considers the transactional switching costs (Belleflamme and Peitz, 2010, p 167). I address the limitations of this measure explicitly in section [7]. For now, it should be adequate to understand that this is the most relevant quantitative proxy for search costs.

5.3.4 Correlation

An issue in regards to including such a large range of control variables in a statistical model is the matter of correlation between the covariates - multicollinearity. This is an obvious concern, especially for the fund structure variables where, e.g., fund age and market share are expected to have rather large correlations. Due to the Fama-Macbeth estimation strategy, I consider the correlations in a cross-sectional sense and simply compute the correlation coefficients utilising definition (iii):

$$Corr_{x,y} = \rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \sigma_y} \tag{xi}$$

Where the covariance in the numerator is: $Cov_{x,y} = \frac{\sum_{i=1}^{n}(x-\bar{x})(y-\bar{y})}{n-1}$. I then compute the cross-sectional correlation matrix, Σ:

$$\Sigma = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \cdot & \cdot & \rho_{1,n} \\ \rho_{2,1} & & & & \\ \cdot & & & & \\ \cdot & & & & \\ \rho_{n,1} & & & & \end{bmatrix}$$

The computation yields:

Table 3: Correlation Matrix for Control Variables

	3Y - Alpha	# Holdings	Fund Size	Fund Age	Deferred Load
3Y - Alpha	1				
# Holdings	0.0178	1			
Fund Size	0.0554	0.2493	1		
Fund Age	0.0448	-0.0482	0.0248	1	
Deferred Load	0.0102	0.0377	0.0461	-0.0057	1

Note: The table depicts the coefficient of correlation of the independent numeric variables

From table 3, it is clearly seen that the correlation of the included, numeric independent variables is not significant. Assets under management and number of holdings exhibit the strongest correlation. A rather interesting observation, as it indicates some co-variation between a fund company's degree of diversification and its assets under management. A preliminary answer to this, could be that larger fund companies feel obligated to diversify their portfolios and thus reduce the investors risk. Although the degree of correlation is 0,24 it is only an issue for the statistical model if it creates hurtful multicorrelation. Established literature requires correlation coefficients as large as 0.9 in order be certain of hurtful multicollinearity which significantly alters the results (Dohoo, Ducrot, Fourichon, Donald, and Hurni, 1997). As we shall see in the next section, the slight correlation is no where near hurtful to the model when applying the Variance Inflation Factor (VIF) and therefore, I do not consider instrument variables or instrument variable equivalents.

5.4 Empirical Results

I establish three different model specifications in accordance with the model specification section above in order to emphasise the robustness of the model. The first model (1) solely considers performance. Model (2) adds fund structure variables while model (3) adds switching costs. Therefore, I effectively compute 32 x 3 = 96 first step regressions and then compute the second-stage regressions three times. The three models are specified below. Please note that all models control for investment strategy confer the argumentation in section [4.3]:

$$\text{Ongoing Charge} = \beta_0 + \beta_1 * \text{Alpha} + \beta_2 * \text{\# Holdings} \tag{(1)}$$

$$\text{Ongoing Charge} = \beta_0 + \beta_1 * \text{Alpha} + \beta_2 * \text{\# Holdings} + \beta_3 * \text{Fund Size} + \beta_4 * \text{Fund Age} \tag{(2)}$$

$$\text{Ongoing Charge} = \beta_0 + \beta_1 * \text{Alpha} + \beta_2 * \text{\# Holdings} + \beta_3 * \text{Fund Size} + \beta_4 * \text{Fund Age} + \beta_5 * \text{Deferred Load} \tag{(3)}$$

As mentioned above, I intend to rule out multicollinearity, utilising the Variance Inflation Factor. As defined by Bowerman O'Connel and Koehler (2004, pp. 223-225):

$$VIF = \frac{1}{1 - R_j^2}$$

Where R_j^2 is the coefficient of determination for the regression model which related the j'th independent variable to all the other independent variables. Bowerman et al (2004) shows that a VIF which never exceeds 10 does not cause hurtful multicollinearity. The VIF never exceeds 1,2 in the year-by-year cross-sectional estimates which allows for a margin of $\frac{10}{1.2} = 8,3$ before the correlation becomes an issue.

Table 4: Fama-Macbeth Heterogeneity Esimators

	Y Variable: Ongoing Charge		
	(1)	(2)	(3)
Constant	68,71***	72,20***	52,74***
	(1,886)	(2,906)	(1,731)
3Y-Alpha	-1,28***	-1,34***	-1,139***
	(0,279)	(0.307)	(0.163)
Holdings		-0,049***	-0,049***
		(0,003)	(0,004)
Fund Size		1.55e-09***	1.83e-09***
		(0,000)	(0,000)
Fund Age		0,002***	0,001***
		(0.000)	(0.000)
Deferred Load			8.932***
			(2,137)
Investment Strategy Control	Yes	Yes	Yes
Adjusted R^2	0,3259	0,3577	0,433
N	14925	14925	14925

Note: *p<0.1; **p<0.05; ***p<0.01. The table provides the Fama-Macbeth Estimators (See equation (v)). and their corrected standard errors. All of the three Fama-Macbeth Models are depicted. (1) Addresses performance, (2) adds fund structure while (3) adds switching costs

The estimates from the second step of the model is provided in table 4. The **alpha** parameter provides strictly negative estimates for all three models. With reference to the discussion in section [5.3.1], this is a rather interesting result as it strongly suggests that funds with larger alphas have lower prices. The results in the established literature is rather ambigious as well: Cooper et al (2015) find a similar result while Ianotta and Navone (2008) find positive parameter estimates. There might exist three reasons for this parameter estimate. (1) Investors simply disregard looking at alpha parameters and perhaps only consider returns confer definition (vii). (2) Funds with large alphas lower prices in order to attract more investors or to accommodate existing investors. (3) Alphas might be negative due to reverse causality, i.e.: funds with lower prices simply have larger alphas. No matter the reason, the variable shows both statistical and economic significance. A parameter estimate of -1,139 implies that an increase in alpha by one (percent) should cause a reduction in prices by 1,139 basis points.

The number of **holdings** has a strictly negative effect on prices which is rather surprising. On one hand, investors should be willing to pay for a larger degree of diversification and thus drive the estimate up. Conversely, funds that tend to hold a larger number of assets at one time usually pursue a strategy in which they rebalance less and have lower costs. I ruled out the theoretical possibility of the latter effect, by controlling for investment strategies. The estimates may suggest that the control variable is not specific enough and thus does not allow for a proper control of investment strategy. Bear in mind, the control variable allows for a rather large degree of flexibility within the three assets classes as described in section [4.3]. For now it is interesting to note that the estimate suggests that investors historically have paid more for funds with less assets.

Fund size has strictly positive parameter estimates. The effect seems rather small, but for interpretation, assets are denominated in millions of kroner and the dependent variable is denominated in basis points. Therefore, the model predicts an increase by 0.00183 basis points for an increase of one million kroner - or alternatively, 0,183 in case of an additional 100 million kroner. After standardisation, the effect may still seem rather small, but putting it into perspective the industry had 850 billion assets under management at the final trading day of 2018 suggesting a rather large economic significance. The model suggests that investors have historically paid more for funds which had more assets under management. It could be argued that this is a result of visibility or perhaps due to the idea that investors believe that size and ability is correlated. Haslem (2017) actually documents the reverse effect, making the visibility argument more plausible.

Fund age has strictly positive estimates, at 0.001 basis points per *day* corresponding to 0.001 * 365 = 0.365 basis points per year. This means that investors historically have paid more for more experienced funds. This results is once again as expected, and can be thought of in the same way as the fund size interpretation above. When that is said, the economic significance is extremely small at an estimate of 0.365 basis points.

Finally, the proxy for **switching costs** has a positive parameter estimate, suggesting that funds with significant monetary switching costs, have similar large ongoing charges. This result is perhaps the most interesting, as it suggests that investors are willing to pay more for funds that also charge them when they leave the fund. The estimate is both statistically and economically significant, as funds with one (percent) larger switching cost on average have ongoing charges 8.9 basis points larger. The result suggests one of two things: (1) Funds with switching costs have some kind of positive differential value which is not included in the statistical model, or (2), investors do not consider switching costs when they invest in Danish mutual funds.

5.5 Model Robustness

In the previous section, I argue that the sign consistency of the three model specifications secure the robustness of the model. In this section, I conduct specific model variation tests and consider alternative variable specifications and samples.

5.5.1 Time Consistency

In order to establish the time consistency of the model, I test the final model for two time subsamples: $2011_{Q1} - 2014_{Q4}$ and $2015_{Q1} - 2018_{Q4}$. The sample size is naturally dramatically decreased. This is mostly a concern for the first part of the time sub-sample, $2011_{Q1} - 2014_{Q4}$, as the amount of funds and thus the sample size is lower for this period. The estimates are presented in table 5 below. Most importantly, the sign of the variables stays consistent, fund size, age and deferred load is positive while the holdings and alpha parameter is strictly negative. Fund age exhibits a small decrease in statistical significance, however it is still significant at a level of one percent. The explanatory power remains approximately the same as for the entire sample, which is reassuring, as the number of observations is naturally smaller for this sub-sample. For the latter sub-sample, the results are once again reassuring, as the sign of the parameter estimates are generally consistent. Once again, the p-value of age and assets is slightly compromised, but not to a degree that requires attention. Explanatory power is naturally a bit larger for this sub-sample, which is a result of the larger number of observations. Generally, the model remains consistent with time.

5.5.2 Variable Consistency

It is no surprise that a wrong variable specification can be detrimental to the results. One variable in particular, **alpha** should be considered in this section.

Table 5: Fama-Macbeth Heterogeneity Esimators, Robustness Test

	Y Variable: Ongoing Charge		
	2011-2014	2015-2018	Variable Alternative
Constant	55,626***	50,534***	45,885***
	(1,633)	(3,044)	(1,617)
Holdings	-0,578***	-0.004***	-0,499***
	(0,279)	(0.000)	(0.163)
Fund Size	1.56e-11***	2.04e-11***	2,24e-09***
	(0,279)	(0.000)	(0.163)
Fund Age	0,0001***	1.99e-06***	0,0003***
	(0,279)	(0.000)	(0.163)
Deferred Load	8,865***	8,342***	7,261***
	(2,544)	(2.358)	(0.163)
3Y-Alpha	-0,9708***	-1,269***	
	(0,238)	(0.218)	
Return			0.802***
			(0.020)
Std Dev			3.253***
			(0.332)
Investment Strategy Control	Yes	Yes	Yes
Adjusted R^2	0,428	0,4419	0,411
N	5224	9701	14925

Note: *p<0.1; **p<0.05; ***p<0.01. The table provides Fama-Macbeth Parameter Estimates for the Robustness Tests

I chose alpha as the primary performance variable as most literature leans on such a variable specification. Furthermore, as the alpha variable includes both relative return and risk, this seems like the obvious choice. Alternatively, I could have included separate return and standard deviation measures. Returns and standard deviations are lagged three years to remain consistent, and are defined:

$$\text{Return} = \frac{\text{Price}_n + \text{Dividends}_{n-0} - \text{Price}_0}{\text{Price}_0} \qquad \text{Standard Deviation } = \sqrt{T^{-1}\sum_{i=1}^{n}[R_i - \bar{R}]^2}$$

Replacing alpha with return and standard deviation, model (3) is ran once again. The results are summarised in the third column of table 5.

Once again, the robustness of the model is confirmed. The sign of the estimates for fund size, age, holdings and deferred load remains the same. Interestingly, both return and standard deviation estimates are positive. The return estimate is intuitive confer the discussion regarding alpha above. The estimate for standard deviation is positive as well, suggesting that funds with larger standard deviations are more expensive. This is naturally the case as funds that invest in more risky assets, as measured by standard deviation, requires more money for their efforts. Furthermore, the strictly positive parameter estimates of returns proposes an interesting question: why are prices increasing in returns but not in alpha? Once again, this might just be the case of reverse causality. Alternatively, this result might suggest that investors fail to include alpha in their choice of fund, and simply invests for returns.

5.5.3 Product Homogeneity Rejection

Conclusively, the model strongly rejects the hypothesis of homogeneous products and thus favours an industry with heterogeneity as was predicted by Hinterhuber. The full model explains approximately 43 percent of the variation in prices for the eight year period, suggesting that prices do in fact reflect performance, fund structure and switching cost measures. Furthermore, the model remains consistent with model variation tests. Fund size and fund age both have strictly positive parameter estimates, with fund size having rather large economic significance. This is an interesting observation, especially as Haslem (2017) finds that larger funds perform worse. This might be due to the consumer assumption that size and ability is positively correlated. The alternative hypothesis is that fund size creates visibility and thus, larger funds can have mark-ups on their products due to lower consumer elasticity.

The negative alpha estimate and positive return estimate, suggests that investors might fail to adjust returns for the risk in a fund and instead considers returns.

Finally, the switching cost proxy is strictly positive which is perhaps the most interesting result. This implies that investors fail to adjust for switching costs when they invest in a fund. This effect may partly be justified, as most funds advise investors to avoid reallocating resources from the fund within three or five years (Morningstar Direct, authors own calculations). Another reason, could be that investors simply do not realise that funds have switching costs before investing in them. Although funds are generally required to disclose switching costs confer the discussion in section [5.3.3.] regarding KIIDs, the average investor may fail to read KIIDs before investing.

Despite of the models explanatory power, there still remains a rather large unexplained price variation and thus dispersion in the market, even after controlling for heterogeneity. Assuming that this rather simple model is a perfect[vi] test for heterogeneity, there must be another structural variable in the industry which describes the remaining dispersion. Most literature, confer section [3] suggest that this is a result of uninformed investors, who incur high search costs allowing high-cost firms to stay in the market.

6 A Price Dispersion Equilibrium

Having already documented the extent of the price dispersion and the role of heterogeneity, this section considers the nature of this dispersion. More specifically, I aim to document whether it is plausible to say that search costs are the reason behind the remaining dispersion, section [6.1] and to identify whether the industry shows characteristics of spatial or temporal price dispersion, section [6.2]. As motivated in section [3] and suggested by Belleflamme and Peitz (2010) and Baye et al (2005) I do this by analysing the degree of price dispersion for each fund company at each point of time. As I have computed heterogeneity controlled estimates, a mere price comparison as was essentially utilised in section [4.3] is no longer adequate. Instead, by computing standardised residuals for each fund company at the cross sectional level, I compare the post-heterogeneity

[vi] Suggesting that prices are not explained by any other quantitative variables

controlled prices over time, as suggested by Lach (2002).

6.1 Existence of Search Costs

Before considering the distribution of residuals over time, I consider the root cause of this inefficiency. Much literature suggests that the heterogeneity controlled price dispersion is due to search costs. However, search costs are quantitatively complex and generally unobservable (Belleflamme and Peitz, 2010, p. 166). Belleflamme and Peitz (2010, p. 166) suggest that the most relevant proxy for search costs is by comparing different economic states, where search costs are intuitively different. The most utilised procedure is by comparing online and offline price dispersion, where offline dispersions are intuitively larger, due to shoeleather costs. An application of this on the mutual fund industry would intuitively produce a similar result. Unfortunately, the data on offline versus online demand is inadequate in this regard. Therefore, I consider two different proxies suggested by Stigler (1961), Ianotta and Navone (2008) and Horatscu and Syverson (2004)

6.1.1 Education Effects

Stigler (1961) presents the argument of education, suggesting that more experienced (educated) buyers naturally ”pay” less search costs. Educated buyers accumulate knowledge of asking prices and thus of the variance of prices in the market. Conversely, inexperienced buyers (tourists), enter a new market, in which they know little about the variance of prices and thus have no idea of the optimal degree of search (Stigler, 1960, pp. 218-219). This is essentially the same argument that Salop and Stigliz (1977) propose, by simply defining the investors with less information about prices as ”less educated”. A relevant proxy for education in the mutual fund market is acquired by considering the role of institutional funds. As was mentioned in section [4.1], institutional funds were not included in the statistical model as these naturally are heterogeneous compared to open-end funds. Institutional funds, are funds that are only sold to institutions and thus not to private investors. As institutions hire highly educated portfolio managers, the cost of search naturally fulfils the above description. I follow Morningstars definition of such funds, defined as funds which exhibit at least one of the following characteristics:

- The fund states explicitly in its Prospectus that the fund is only suited for institutional investors

- The fund has a minimum investment requirement
- The fund has institutional in its name

In order to compare the dispersions for open-end and institutional funds, the cross-sectional fee dispersion for institutional funds and open-end funds for the last trading day of 2018 is computed. The comparisons are depicted in figure three below. It is quickly seen that the price dispersion is significantly larger for the open-end funds than for the institutional funds, suggesting that the price dispersion and thus the search costs, are smaller for institutional funds. This informally confirms the existence of education effects within the mutual fund industry, as was expected.

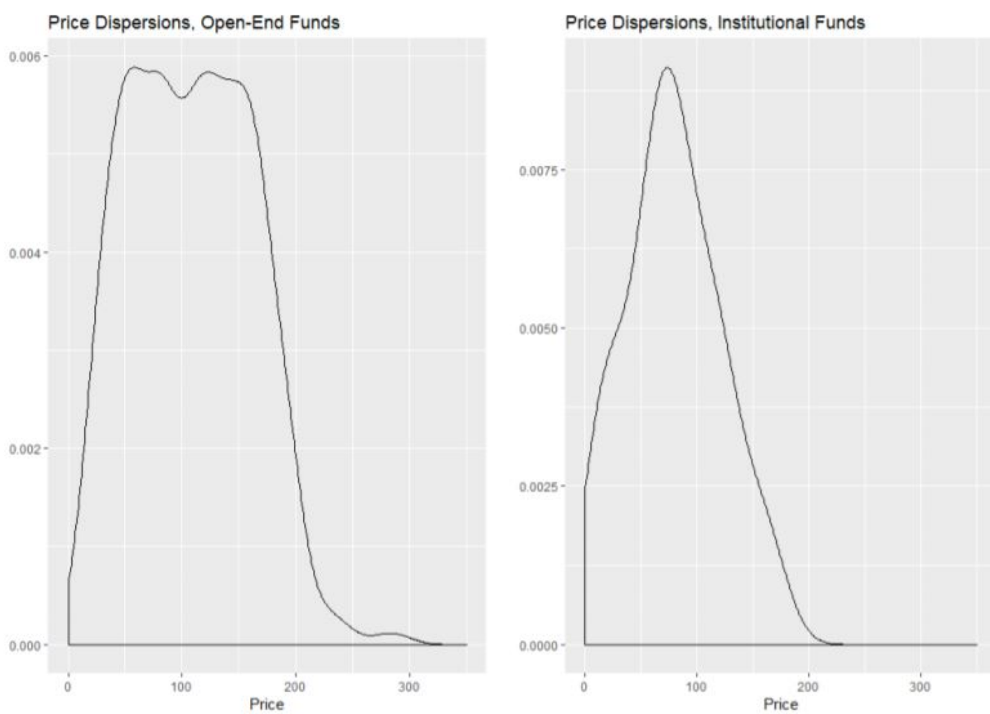

Figure 3: Price distribution of open-end and institutional funds

6.1.2 Control Variable Comparison

An alternative, and perhaps more intuitive approach, is to compare the (unexplained) variation of prices through the residuals (Ianotta and Navone, 2008, p. 852). If the remaining dispersion of the model is really due to search costs, the residuals should vary less for a sub-sample of funds with intuitively less search costs. For example, the remaining dispersion of the statistical model for institutional funds, if they were included in the model, would intuitively vary substantially less than for the open-end funds.

Firstly, older firms intuitively have smaller search costs as they have had a longer period of time to provide track records and to inform their prospective customers about their value proposition. Therefore, the unexplained variation in prices should vary substantially less for older funds than for new ones. Define two samples. S_1, funds with intuitively low search costs, which are at least five years old. S_2, funds with high search costs, which are 5 years old or younger. By utilising definition (ii), I compare the variance of these residuals for the two samples, S_1 S_2. The results confirm the expectation of search cost being larger for smaller funds, as the variance of S_1, low cost firms, is 1523 compared to 1941 for the high cost firms. Testing the significance of this difference, I compute an F-test for equality of variances (Newbold, Carlson, Thorne, 2013, pp. 403-404):

$$H_0 : \sigma_{S_1} = \sigma_{S_2} \qquad H_A : \sigma_{S_1} < \sigma_{S_2}$$

The test statistic, F, is computed:

$$F = \frac{Variance_{S_2}}{Variance_{S_1}} = \frac{1941.131}{1523.100} = 1,274$$

Comparing the test statistic to an appropriate F-distribution, the null hypothesis, is barely not rejected, with a p-value of 0.0562. As it is insignificant at the limit, acceptance of significantly different variances is not inappropriate. Therefore, the test of state comparison once again yields the expected result: (Heterogeneity controlled) prices are less dispersed for older funds confirming that the remaining dispersion might be due to search costs.

Similarly, larger funds as measured by assets under management, should have intuitively lower search costs. Once again, I compute two different samples, the upper and lower quantile of assets. The hypothesis approach is similar to the above, and the test statistic yields:

$$F = \frac{Variance_{SmallFunds}}{Variance_{LargeFunds}} = \frac{1762.385}{1347,916} = 1,342$$

The hypothesis is once again rejected, with a p-value of 0.0347 which suggests that funds with larger assets under management have significantly lower variance of price dispersion. It is naturally difficult to establish a conclusive result in regards to the existence and extent of search costs as the primary inefficiency in the market. The brief analysis however, documented three relevant proxies for search costs which all strongly favours the existence of this phenomenon.

6.2 The Persistence of Price Dispersion

Having established that the remaining dispersion is due to search costs, I now return to consider the nature of this dispersion. I compute standardised residuals of the model at the cross sectional level for each fund to ensure comparability over time (Belleflamme and Peitz, 2010) (Baye et al, 2005), (Lach 2002). At the cross-sectional level, each standardised residual is allocated to a relevant quantile: Q_0, Q_1, Q_2, Q_3 where Q_0 is for funds with residuals ranked in the 25 lowest percent, Q_1 from 25 to 50 and so forth.

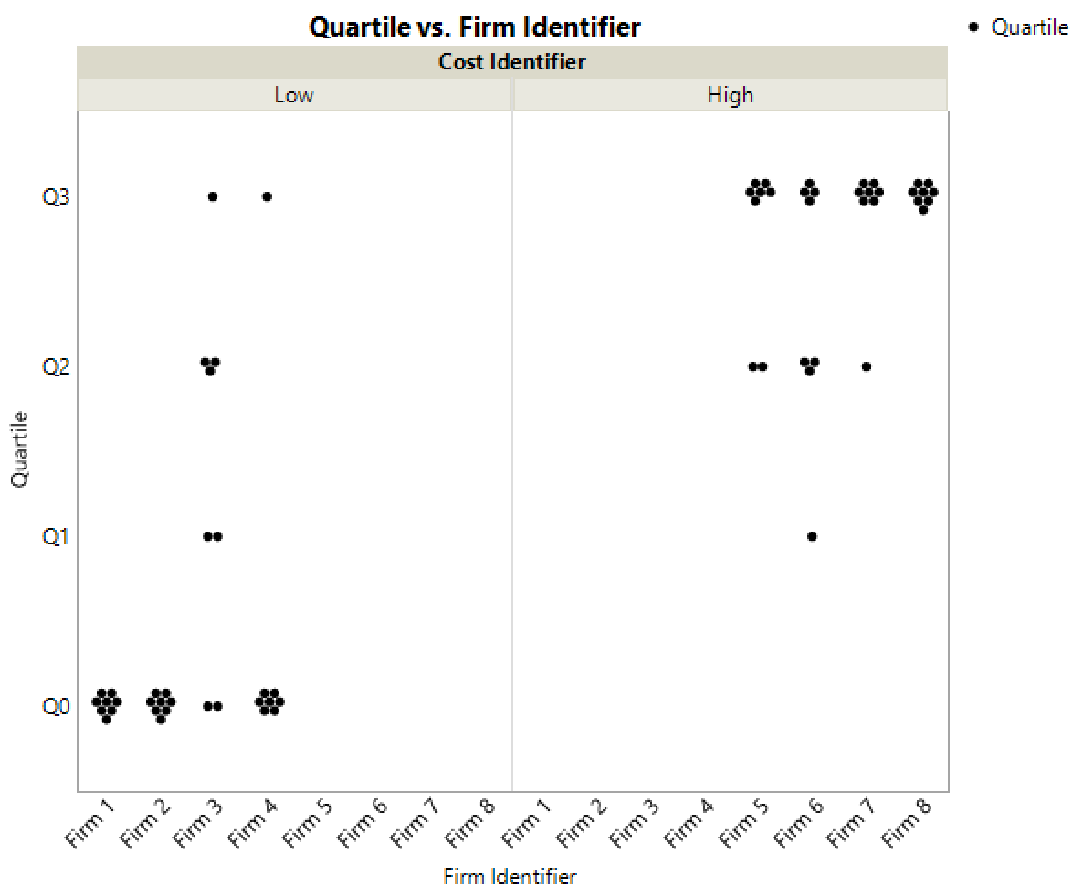

Figure 4: Price Persistence of Top and Bottom Funds

This allows me to capture if a fund has priced significantly above or below the expected price from the statistical model. It is necessary to understand, that if a fund exhibits intra-distribution movement from, say Q_3 to Q_2, it does not necessarily mean that the fund lowered its price. Instead, it means that holding together the underlying conditions of the fund (age, size, alpha, switching costs and holdings) with prices, there is a mismatch. Firstly, I acquire the residuals of the four funds with lowest residuals, and the four with the highest residuals. The results are depicted in figure 4.

The results are unambiguous. Fund company 1, 2 and 4, the low-cost firms, all present almost exclusive persistence, having residuals in the lower quantile over the entire period. Fund company 3 is more dispersed, having residuals in all of the 4 relevant quantiles. The high-cost funds: 5, 6, 7 and 8 are primarily in the opposite side of the spectrum. One fund however, company 6, has a more dispersed profile similar to fund company 3. Overall, the pricing is rather consistent, especially for the low-cost funds. These funds have historically priced significantly lower than their expected prices, but have still failed to attract enough investor attention to raise prices. Conversely, many of the high-cost funds have shown consistent over-pricing compared to their expectation from the statistical model. Relating this result to the price dispersion models presented in section [3.1] and [3.2], the results are definitely in favour of the spatial price dispersion model. Although the residuals are changing for some funds over time, it is not in favour of temporal pricing dispersion. The result suggest, that the uninformed consumers remain uninformed, as the cost of search (or education) is too large for investors.

In order to generalise the result to the entire sample, I compute each fund company's quantile position at each cross-section. Then, the probability of persistence is calculated for each quantile following a simple probability definition by Sørensen (2008, p. 9):

$$P = \frac{\#\text{Successful outcomes}}{\#\text{Outcomes}}$$

For example, if half of the funds from the upper quantile in 2010 are still in the upper quantile in 2018, I allocate 50 percent probability of persisting over eight years. This naturally suggests, that large probabilities leads to spatial price dispersion while probabilities converging towards zero suggest temporal dispersion. Due to the unbalanced data set and for simplicity, the analysis only includes funds that are open throughout the entire period. This is considered to be the best alternative to the inclusion of all funds. Furthermore, the focus is on the upper and lower quantile of residuals. The probabilities for the upper and lower quantiles are computed for the entire period. The results are summarised in the table below:

Table 6: Price Dispersion Persistence

	Upper Quantile	Lower Quantile
1 Year	0.744	0.909
2 Years	0.697	0.837
3 Years	0.6511	0.818
4 Years	0.621	0.741
5 Years	0.604	0.709
6 Years	0.589	0.681
7 Years	0.418	0.596
8 Years	0.391	0.567

Note: The table depicts the residual price persistence for upper and lower quantile funds. The values are calculated as fractions.

The probabilities generally confirm the above results, as probabilities are rather large throughout the period, being larger for low-cost funds. 90 percent of the low-cost funds are still low-cost funds a year after, while 74 percent remain for the high-cost funds. The probabilities are, as expected, a negative function of time, so probabilities are declining. However, even after eight years in the quantile, 56 percent of the low-cost funds and 39 percent of the high-cost funds remain in their respective quantile, strongly suggesting spatial price dispersion. Conclusively, the unexplained variation of prices, after controlling for heterogeneity, can be explained by significant search costs as measured by the three developed proxies in section [6.1]. Furthermore, the inefficiency that this search cost provides, manifests itself as spatial price dispersion meaning that high-cost firms i.e. firms that charge a price above the justified price, remain high-cost while the low-cost firms remain low-cost. Therefore, despite of previous authors rejection of spatial price dispersion as an empirical phenomenon, the study finds that such dispersion still remains in the Danish mutual fund industry.

7 Discussion & Limitations

This study has been revolved around analysing price dispersion in the Danish mutual fund industry. As the study has a rather case oriented approach, it is naturally a concern that the results can not be generalised. It is important to emphasise however, that a generalisation of spatial price dispersion or search cost existence is not the intention. Instead, the study aims to shed light on a rather unexplored phenomenon for mutual funds and therefore, the study should be seen as a step in exploring the phenomenon in depth (Flyvbjerg, 2010, p. 470).

The most vulnerable analysis of the paper is naturally the development of the statistical model. As the later analysis is revolved around modelling of residuals from this model, inadequate mode ling is a concern. I have implemented several measures to avoid the most common issues in financial modelling as suggested by Petersen (2008), and conducted robustness tests to ensure the statistical and economical significance of estimates. It is quickly seen that the study relies on the assumption of perfect heterogeneity controls in the statistical model as discussed in section [5]. If the explanatory variables are an imperfect control for heterogeneity, the residual and search cost analysis is directly affected. I have applied the most common explanatory variables, as was suggested by the established literature. However, it has been a top priority to develop a simple model in order to advocate to intuition and to avoid overfitting by allocating too much thought to better explanatory power (Walpole, Myers, Myers, Ye, 2007, p. 508). Therefore, it is not inconceivable that there may exist more explanatory variables that will be significant in explaining the price variations.

Related to the discussion on modelling concerns, the switching cost variable is of particular importance. As discussed in section [5.4], the idea of developing a perfect quantitative proxy for switching costs may be delusional as deferred load only identifies the transactional costs. The annual loyalty survey on Danish banks conducted by EPSI Denmark, found that approximately 60 percent of the Danish customers have been in the same bank for 10 years, primarily because they had been there for many years (EPSI Denmark, 2017). Seeing as most fund companies are banks and pension funds, it is plausible that Danes have switching costs in the form of psychological costs defined by Belleflamme and Peitz (2010, p. 168) : ”The quality of the goods or service is taken on trust, the customer may be reluctant to switch to another supplier she has not yet learned to trust”. In order to develop a more nuanced and reliable proxy it may be an idea to produce a variable with a qualitative nature. Similarly, the search cost proxies developed in section [6.1] may have been identified in a more precise manner with such an approach.

A particular note has to be made on the role of marketing. Search costs are naturally declining with the marketing initiatives of a fund company, as marketing initiatives intuitively provide exposure to the market. Most theoretical approaches on price dispersion disregard the role of marketing (Belleflamme and Peitz, 2010, p. 157). It is quickly seen however, that such a cost heterogeneity within fund companies, will allow price dispersions in the market that are not necessarily an inefficiency in the economic sense. If the search costs, and thus the price dispersion, is perfectly explained by marketing initiatives on fund company level, the price dispersions may be a natural development of access based positioning (Nagle and Müller, 2017, pp. 156-157). Unfortunately, the acquisition of marketing budgets for fund companies is rather scarce and therefore, I strongly suggest that future research on this area develop on the ideas of including marketing into the analysis.

Finally, the study generally assumes heterogeneity on firm level, confer the discussion above, but also on investor level. More specifically, the results have not been generalised to a utility framework. Therefore, the study is only able to serve as an initial study on the area of investor welfare implications. Drawing on the theory of industrial organisation however, the documented inefficiency obviously lowers investor welfare. Horatscu and Syverson (2004) develop a utility framework for for this, however further developments will be efficient for the study of price dispersions in the future.

8 Conclusion

Conclusively, this study finds that price dispersion are consistently present in the mutual fund industry, even after correcting for investment strategies. These price dispersions are to some extent justified by heterogeneity among mutual fund products, as the developed statistical model suggests that roughly 43 percent of the price variation is explained by such heterogeneity. Rather counter-intuitively, Alpha, the excess monetary return is estimated to have a negative parameter suggesting that the price falls approximately one basis point per percentage alpha. Furthermore, characteristics that do not directly affect monetary returns for investors, such as fund size and fund age, play a critical role for the pricing decision of fund companies and for the investors choice of investment. Especially the parameter estimate for fund size shows economic significance. As suggested by Belleflamme and Peitz (2010, p. 167), switching costs are often important in the pricing decision, which is also caught by the model by utilising the Morningstar proxy, deferred load.

As it has been suggested by previous literature, the remaining price dispersion may be due to search costs. By utilising three proxies for this quantitatively complex variable, I find strong evidence for the existence of significant search costs in the industry. Finally, I utilise this result to study the nature of the remaining price dispersion. The results, both on a small sample and on all funds, suggests that the industry exhibits characteristics related to spatial price dispersion. This result is puzzling, as it suggests that mutual funds that have unjustifiably high costs show persistence in their pricing. As these funds generally manage to stay in the market over the eight year period, this suggests that investors are not able to identify which funds are high-cost funds.

Although the study has had a rather exploratory nature, as mentioned in section [1-2], the findings have implications for managers, investors and regulators alike. Managers should primarily draw attention to the heterogeneity control in section [4]. Although the price variation is not fully explained, the economic significance of the parameter estimates can work as a guideline for future pricing policy. Investors allocate positive differential value to fund size and fund age. Although fund age and size is not necessarily adjustable on a managerial level, both variables might be seen as an expression for something else - trust. Therefore, if managers are able to establish such a relationship with customers pricing for profit may become easier. Furthermore, the study finds that investors look to returns rather than alpha, when making their investment decisions.

Private investors should primarily focus on the findings in the latter part of the paper - search costs make the uninformed investor worse off. More specifically, some fund companies consistently charge unjustifiably large prices, which hurt the uninformed investors that invest in those funds. Investors, especially those who know that they are uninformed, should look to limit the search costs most effectively. Theoretically, this can be done through the utilisation of a clearinghouse. Such clearinghouses are present in the mutual fund industry: Morningstar and Bloomberg among others. The degree of spatial dispersion however, suggests one of two things: (1) Investors do not utilise the clearinghouses, (2) The clearinghouses do not provide adequate information to the uninformed investors.

Finally, the study suggests that regulators should continue their pursuit of providing transparency in the mutual fund industry, in order to limit the search costs. Although the European Parliament and SEC has taken steps in this direction, the model suggests that spatial dispersion is still at large in 2018.

9 References

- Anadu, K., Kruttli, M., McCabe, P., E., Osambela, E., & Shin, C. (2018), The Shift from Active to Passive Investing: Potential Risks to Financial Stability? *FEDS Working Paper* No. 2018-060

- Baye, M., Morgan, J. Scholten, P. (2006). Information, Search, and Price Dispersion, *Working Papers 2006-11, Indiana University*, Kelley School of Business, Department of Business Economics and Public Policy.

- Belleflamme, P., Peitz, M., (2010). Industrial Organization: Markets and Strategies, Ch. 12

- Bowerman, B., O'Connell, R., Koehler, A. (2004), South-Western College Pub

- Cochrane, J. (2005). Asset Pricing, Princeton University Press

- Cooper, M., Halling, M., Yang, W. (2016). The Mutual Fund Fee Puzzle.

- Dohoo, Ducrot, Fourichon, Donald, and Hurnik. An Overview of Techniques for Dealing with Large Numbers of Independent Variables in Epidemiologic Studies. *Preventive Veterinary Medicine* 29.3 (1997): 221-39.

- Fama, E., Macbeth, J. (1973), Risk, Return, and Equilibrium: Empirical Tests, *The Journal of Political Economy*, Vol. 81, No. 3. (May - Jun., 1973), pp. 607-636

- Flyvbjerg, B, (2010) Fem misforståelser om casestudiet (Five Misunderstandings about Case-Study Research). *In Svend Brinkmann and Lene Tanggaard, eds., Kvalitative metoder*,

- Haslem, J. (2017), Mutual Fund Economies of Scale: Nature and Sources, *The Journal of Wealth Management Summer* 2017, 20 (1) 97-124

- Henriksen, K (2010), Hvad betaler jeg for og til?
  Located at: http://www.morningstar.dk/dk/news/91809/hvad-betaler-jeg-for-og-til.aspx

- Hinterhuber, A., (2016). The six pricing myths that kill profits, *Business Horizons*, vol. 59(1), pages 71-83

- Holm, A. (2016), Videnskab i virkeligheden - en grundbog i videnskabsteori, Samfundslitteratur

- Iannotta, G., Navone, M., (2012). The cross-section of mutual fund fee dispersion, *Journal of Banking Finance*, vol. 36(3), pages 846-856.

- Lach, S. (2002). Existence and Persistence of Price Dispersion: An Empirical Analysis. *The Review of Economics and Statistics*, 84(3), 433-444.

- Nagle, T., Müller G. (2017), The Strategy and Tactics of Pricing: A Guide to Growing More Profitably 6th Edition

- Newey, W. K., and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–708.

- Newbold, P., Carlson, W., Thorne, B. (2013), Pearson Education Limited

- Nocera, J. (2019), John Bogle, The Apostle of Index Funds, Never Gave Up
Located at: https://www.bloomberg.com/opinion/articles/2019-01-17/vanguard-founder-jack-bogle-was-the-apostle-of-index-funds

- Petersen, Mitchell A. (2006), Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, Kellogg Finance Dept., *Working Paper* No. 329.

- Pratt, J., Wise, D., Zeckhauser, R. (1979). Price Differences in Almost Competitive Markets. *The Quarterly Journal of Economics*, 93(2), 189-211.

- Salop, S., Stiglitz, J. (1977). Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion. *The Review of Economic Studies*, 44(3), 493-510.

- Schaefer, S., (2018), Passive beats active management
Located at: https://www.london.edu/lbsr/passive-beats-active-management-its-not-scepticism-its-arithmetic

- Stavnsager Rasmussen, E., Østergaard, P., Beckmann, S. C. (2006). Essentials of Social Science Research Methodology. Odense: Syddansk Universitetsforlag.

- Syverson, C., Hortaçsu, A., (2004), Product Differentiation, Search Costs, And Competition in the Mutual Fund Industry: A Case Study of SP 500 Index Funds. *The Quarterly Journal of Economics*. 119. 403-456

- Söderberg, H. (2017), EPSI Rating Bank 2017
  Located at: http://www.epsi-denmark.org/report/banksektoren-2017/

- Varian, Hal R, (1980). A Model of Sales. *American Economic Review, American Economic Association*, vol. 70(4), 651-659

- Walpole, R., Myers, R., Myers, S., Ye, K., 2007, Probability & Statistics for Engineers & Scientists, Pearson Education International)

- White, M.J., (2013), Hedge Funds – A New Era of Transparency and Openness
  Located at: https://www.sec.gov/news/speech/2013-spch101813mjw

- Wooldridge, J. (2010). Econometric Analysis of Cross Section and Panel Data, MIT Press

- Wooldridge, J., (2016), Introductory Econometrics : A Modern Approach, Cengage Learning, Inc