ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Behavioural Economics

Performance persistence in U.S. mutual funds' returns

Student name: Eri Kooistra

Student ID number: 413710

Supervisor: PhD Candidate A.C. Peker

Second assessor:

Date final version: 20-01-2020

Abstract

Performance persistence signifies the ability of a mutual fund to gain returns on investment above a certain reference point on a continuous basis. This paper studies the impact of implementing the streak selection bias into the investigation about persistence in mutual fund performance. If the probability of flipping heads is 0.5, the streak selection bias implies that the probability conditional on an ongoing streak of flipping heads is lower than the underlying probability of 0.5. This streak selection bias was discovered by Miller and Sanjurjo (2018). Imagine a sample of similar mutual funds for which the probability of performing better than half of its peers is 0.5. If this fund experiences a streak of one period or more of performing better than half of its counterparts, the streak selection bias states that the conditional probability of performing better than half of the funds again in the next period is lower than the underlying probability of 0.5. A data set is constructed covering 27,718 U.S. equity funds over the 2010-2018 period which is used to assess the effect of accounting for the streak selection bias when performance persistence is measured. This is done by comparing two different research methods. One method does not take the streak selection bias into account and uses the uncorrected probability of 0.5 for a fund to perform better than half of the funds regardless of any ongoing streaks. The other method corrects for the streak selection bias by using the corrected conditional probability. This conditional probability differs in streak length. The results show that under the uncorrected probability persistence in mutual fund performance is only present in the short run. Mutual fund performance under the corrected probability persists both in the short and long run. This indicates that accounting for the streak selection bias increases the duration for which performance persistence in mutual funds occurs.

Keywords: Performance Persistence, Streak Selection Bias, Contingency Tables, Binomial Tests

JEL codes: G00, G02, G11, G17, G19

Disclaimer

Acknowledgements

With handing in this final master’s thesis, the end of my academic career at the Erasmus University Rotterdam is coming to an end. First and foremost, I would like to thank my supervisor Cem Peker for supervising me throughout the writing process of my thesis. He was always eager to help and open for discussion combined with great knowledge about the topic. Furthermore, I would like to show my appreciation to my parents who have always supported me.

Table of Contents

1. Introduction 6
2. Theoretical framework 9
2.1 Performance persistence in the securities market 9
2.2 The 'hot hand' in sports 11
2.3 Streak selection bias 12
2.4 Hypotheses 13
3. Methodology 15
3.1 Estimating performance measure alpha 15
3.2 Contingency tables & Pearson's chi-squared test 17
3.3 Binomial tests 19
3.4 Corrected benchmark 20
3.5 Manual calculations 21
4. Data 23
4.1 Data selection 23
4.2 Data modifications 25
5. Results and discussion 28
5.1 Contingency tables using the canonical benchmark 28
5.2 Contingency tables using the corrected benchmarks 31
5.3 Binomial tests using canonical benchmark 34
5.3.1 Winning streaks 34
5.3.2 Losing streaks 36
5.4.1 Winning streaks 37
5.4.2 Losing streaks 39
5.5 Discussion 40
6. Conclusion 42
Appendix 48

List of tables

TABLE 1 - STREAK SELECTION BIAS VISUALIZED BY COIN FLIP EXAMPLE 13
TABLE 2 - CONTINGENCY TABLE EXAMPLE 1 17
TABLE 3 - CONTINGENCY TABLE EXAMPLE 2 19
TABLE 4 - CORRECTED BENCHMARK 21
TABLE 5 - MANUAL CALCULATION CONTINGENCY TABLE EXAMPLE 22
TABLE 6 - OVERVIEW OF ALL THE VARIABLES OF THE FINAL DATA SET 26
TABLE 7 - CONTINGENCY TABLE CANONICAL BENCHMARK 29
TABLE 8 - CONTINGENCY TABLES CANONICAL BENCHMARK 30
TABLE 9 - CONTINGENCY TABLE CORRECTED BENCHMARK 32
TABLE 10 - CONTINGENCY TABLES CORRECTED BENCHMARKS 33
TABLE 11 - BINOMIAL TESTS WINNING STREAKS CANONICAL BENCHMARK 35
TABLE 12 - BINOMIAL TESTS LOSING STREAKS CANONICAL BENCHMARK 36
TABLE 13 - BINOMIAL TESTS WINNING STREAKS CORRECTED BENCHMARKS 38
TABLE 14 - BINOMIAL TESTS LOSING STREAKS CORRECTED BENCHMARKS 39
TABLES 15-39 - APPENDIX A TO E...48

1. Introduction

During the last decades, bond and equity mutual fund investors, on average, have failed to beat the markets emphasising how challenging it is to outperform index funds for a longer period of time on a regular basis (Sommer, 2019). Despite the fair amount of experience or the possession of in-depth knowledge about the do's and don'ts concerning the investment climate, nearly every mutual fund manager is unable to consistently exceed the portfolio related market indices. Because only few succeed in beating the market, it appears obvious that investing in actively managed funds is not a decision without risks. Nonetheless, in July of this year the assets of passively managed U.S. equity funds exceeded the assets of actively managed U.S. funds for the first time in history showing how popular investing in actively managed funds still is among investors (Gittelsohn, 2019). According to the belief of many investors that actively managed funds are more profitable than passively managed funds when picked out carefully, one is faced with a choice between a vast amount of actively managed mutual funds. As these funds are desperately trying to find common and alternative ways to beat the market, another thoroughly investigated topic still keeps being ambiguous in its findings: the capability of mutual funds to perform better than their peers for a considerable length of time.

If mutual funds would consistently outperform their rivals or, one could argue that these mutual funds have the ability to build portfolios which are more profitable than those of their counterparts on a continuous basis. This phenomenon is called 'performance persistence' and dates from 50 years ago (Jensen, 1969). Jensen (1969) investigated the correlation between the average value of portfolio returns in excess of the benchmark over different periods and found persistence only among bad performing funds. The latter is more recently underlined by three other researchers (Carhart, 1995; Shukla & Trzcinka, 1994). This particular form of performance persistence is called negative performance persistence. On the other hand, Brown and Goetzmann (1994) found performance persistence in good and bad performing funds. One of the early researchers that investigated the possible existence of persistence in mutual fund performance did not find any evidence pointing in the direction of performance persistence whatsoever (Carlson, 1970). When performance persistence among mutual funds is detected, the duration of performance persistence is another point of discussion. Whereas Hendricks, Patel, and Zeckhauser (1993) found performance persisting over one year at most, Grinblatt and Titman (1994) find performance persistence up to five years. Goetzmann and Ibbotson (1994) even find performance persistence during a one-month, one-year and two-year period. This shows that the duration of performance persistence found by researchers is divergent.

Performance persistence is not an event appearing solely in the financial environment. In sports it is called the 'hot hand' symbolizing the false belief of supporters deeming the possibility of a basketball player scoring a shot more plausible when his shot is following a streak of consecutive hits (Gilovich, Vallone & Tversky, 1985). More than thirty years later, evidence given by Miller and Sanjurjo (2018) in favour of the existence of the ‘hot hand’ addressed statistical flaws in the conventional approach used to investigate the possible existence of the ‘hot hand’. Precisely, a streak selection bias is emerging that indicates the dependency of current outcomes on previous streaks. To illustrate, the probability of flipping heads is 0.5 for every coin flip, Pr(H) = 0.5. However, as stated by Miller and Sanjurjo (2018), the conditional probability of flipping heads if the coin flip is following a streak of consecutive flip of heads, Pr(H | streak), is less than the underlying probability of flipping heads. So, Pr(H | streak) < 0.5. The size of the streak selection bias varies with streak and sequence length. For instance, the conditional probability for flipping heads if it is the 26^{th} flip out of 58 flips total when you are experiencing a streak of 13 times flipping heads is different from the conditional probability for flipping heads if it is the 14^{th} flip out of 29 flips total when you are experiencing a streak of 7 times flipping heads. Translating the coin flip example according to performance persistence in this paper, the conditional probability, given a streak of K[1] wins, of a mutual fund to perform better than half of its peers is lower than 0.5. The same goes for the probability of becoming a loser if it follows a sequence of K losses. The assumption made in this paper to legitimise the latter two statements is that the probability for a fund of having higher returns than half of its peers is 0.5[2]. Miller and Sanjurjo (2018) corrected the outcomes of Gilovich et al. (1985) for the streak selection bias and came to a contrary conclusion. They dispelled the theory that dismisses the ‘hot hand’ as a fallacy and, as a matter of fact, their results validated the view of the basketball crowd. Up until today, to the best of my knowledge, this new improved way of testing for the existence of the ‘hot hand’ is merely applied to the world of basketball. Yet, it has to set foot in the broadly studied field of finance and in particular, the securities market. The aim of this paper is to investigate if the discovery of a streak selection bias influences the presence and length of performance persistence. Hence, the research question of this paper is as follows:

> ***What is the effect of correcting for the streak selection bias on short- and long-term performance persistence?***

This research question will be divided into four developed hypotheses capturing the short and long-run aspect of performance persistence. These will be presented further on in this paper. A comparison will be made between the outcomes which are based on the use of the earlier mentioned probability of 0.5 which is not corrected for the streak selection bias and the outcomes which are based on the use of conditional probabilities accounting for the streak selection bias. These probabilities will be named the canonical benchmark and the corrected benchmark, respectively. In this paper, short-term performance persistence symbolizes the ability of mutual funds to significantly outperform their competitors uninterruptedly for two years or less. Long-term performance persistence signifies the ability of mutual funds to significantly

[1] Where K > 0
[2] This is an assumption which will be elaborated on further in the remaining of this paper.

outperform their competitors for more than two years uninterruptedly.

To answer the research question, data is gathered from CRSP concerning 27,718 U.S. equity funds between 2010-2018. An in-sample period between 2010 and 2014 is used to fit the model with an ordinary least squared regression and to predict out-of-sample returns for the 2015-2018 period where the four your period between 2015 and 2018 is divided into 16 periods of three months, i.e. 16 quarters. Subsequently, those out of sample returns are used to compute a performance measure alpha following the Carhart four-factor model (Carhart, 1997). Based on this alpha, funds are ranked as a winner – performing better than half of their counterparts – or loser – performing worse than half of their counterparts – by computing the median performance value for every given period. The ranking of all periods is used to create 2x2 contingency tables for every streak length. One categorical variable consists of funds experiencing a winning or losing streak for a consecutive number of periods. The other categorical variable indicates if those funds are a winner or loser in the subsequent period. The test corresponding to the use of contingency tables is the Pearson's chi-squared test for independence. This test calculates if the two variables are independent by looking at the difference between the observed and expected frequency of every cell. If the cells displaying funds experiencing a winning streak being winners in the next period and funds experiencing a losing streaks being losers in the next period show a significantly higher observed than expected frequency, the two variables are not independent and performance persistence for that particular streak length with regard to the next period is apparent. The second method used are one-sided binomial tests which simultaneously serve as a robustness check.

Being the first of its kind, this paper aims to apply the streak selection bias into research about performance persistence in the financial world. This study finds results indicating that the corrected benchmark show performance persistence in the short and long run while the canonical benchmark only displays short-term performance persistence. Remarkably, the corrected benchmark lengthens the duration for which positive and negative performance persistence is present. Both contingency tables and binomial tests show similar results.

The remainder of this research paper is organised as follows. The next chapter provides background information covering several aspects. Chapter 3 describes the methodology, followed by the constructed data set in chapter 4. After that, the results are discussed in chapter 5. Lastly, the paper is concluded by summarizing the most imported findings, mentioning the limitations and putting forward recommendations for future research, all in chapter 6.

2. Theoretical framework

In this section, I will start with discussing the relevant literature and origin of performance persistence in the securities market. Thereafter a transition is made to the world of sports where a similar phenomenon arises, called the ‘hot hand’ in subsection 2.2. Finally, the critique of Miller and Sanjurjo (2018) will be dwelled on in paragraph 2.3. Lastly, in subchapter 2.4, a preliminary conclusion is defined.

2.1 Performance persistence in the securities market

Actively trading mutual funds are expected to build portfolios which yield the highest possible gain for every single period and are chosen by investors based on the belief of their ability to outperform their peers. As the securities market is a speculative market, it is interesting to investigate whether mutual funds are able to perform better than their rivals on a continuous basis. In regard to the earlier description of performance persistence, the following question captures the financial ‘hot hand' phenomenon for mutual funds: if a mutual fund yields portfolio returns that are higher than the median portfolio returns of all funds which are similar to that fund for a given period, is this mutual likely to choose the more profitable securities again for its portfolio in the subsequent period? Complementary, if a mutual fund yields portfolio returns that are lower than the median portfolio returns of all funds which are similar to that fund for a given period of time, is this mutual fund likely to choose the less profitable securities again for its portfolio in the subsequent period? In more financial wording, is there persistence in mutual fund performance? The answer to that specific question has several parts that need reviewing, some of which this study is going to elaborate on in the following section. Mostly, I will review papers that investigated short-run or long-run persistence in mutual funds’ performance since it is in line with the framework of this research paper. As the chosen timeframe for long-run performance persistence differs considerably among previous researchers, this theoretical framework part classifies all the discussed research investigating performance persistence over a time horizon that is longer than two year as long-run performance persistence. Short-run performance persistence will refer to performance persisting up to two years.

Grinblatt and Titman (1992) studied the relationship between past and future performance of mutual funds and found evidence of positive and negative long-run persistence. They attributed the consistent excess returns to managerial skill. The chosen benchmarks did not account for these abnormal earning differences because they were biased. Moreover, the continual divergences in service charges and transaction expenses between funds did not account for these abnormal earning differences too. Earlier research of the two already confirmed some growth and aggressive-growth funds persistently performed better than their peers due to a newly introduced bias-free benchmark (Grinblatt & Titman, 1989).

Based on data including quarterly returns of no-load, growth, open-end equity funds over the 1974–1988 period, significant results showed positive and negative short-run persistence in their relative performance (Hendricks et al., 1993). The evidence was strongest when a one-year estimation period was used. Investors who adapted strategies which took this arbitrage opportunity into account beat the mediocre mutual funds substantially leading to higher earnings. Mutual funds did not show significant positive persistence in surpassing their peers' benchmarks. On the contrary, evidence for steadily underperforming mutual funds was found which they named as managers having 'icy hands'. The findings of Hendricks et al. (1993) are four years later however refuted by Carhart (1997) explaining the 12-month momentum effect, described by Jegadeesh and Titman (1993), to be responsible for the discovered short-run persistence. If equity funds still show short-run persistence, this can be attributed to familiar components[3] generating stock earnings and investment and strategy costs but not to the art of picking the best stocks. Moreover, the significant underperformance by the worst performing mutual funds, found by Hendricks et al (1993), are not clarified and not in line with their view of a manager's ability to pick the most profitable stocks.

Not long after statements have been made about the substantial impact of fund style and expenses on persistence in mutual fund managers' performance, Kahn and Rudd (1995) investigated long-run persistence in mutual fund performance controlling for these factors specifically leading to findings in support of the assumption of long-run persistence among fixed-income mutual funds. Equity funds did not show meaningful signs supporting long-run persistence. Nonetheless, investing in fixed-income mutual funds was not recommended as managerial costs charged and transaction expenses of the respective funds passed on to the investor limiting the benefits of picking past winners. On average, fixed-income funds underperform the market and being able to choose past winners does not outweigh the negative monetary consequences transactions and managers bring along.

Huij and Derwal (2008) found that information about mutual funds' results of past periods could be used to build a strategy in order to generate earnings different from the norm due to existing short-run persistence in performance among investment grade corporate (high-quality) and high-yield (low quality) bond funds. These findings imply mutual fund managers' capability of picking the above average profitable bonds for their portfolio to be true. If these skilled managers produced exceptional returns on a permanent basis, they endured 'hot hands'. While Kahn and Rudd (1995) did not suggest to invest in actively managed mutual funds, Huij and Derwal (2008) favour the strategy of advanced investors which adapt to exploiting valuable insights about persistence since extraordinary profits can be contrived outmanoeuvring the associated benchmarks. On the other hand, index funds are still the most evident choice for investors who

[3] The two models that were used to measure mutual fund performance are the Capital Asset Pricing Model (CAPM) of Lintner (1965) and Sharpe (1964) as well as the four-factor model of Carhart (1995), which is an expanded version of the three-factor model of Fama and French (1993).

do not possess the adequate competences to take advantage of persistence in mutual fund performance (Huij & Derwall, 2008).

2.2 The 'hot hand' in sports

Finding its origin in finance, the existence of performance persistence might also be present in other areas, such as the area of basketball where it is named as having a 'hot hand' (Gilovich et al., 1985). Can an athlete indeed have a 'hot hand'? Since the initial time Gilovich et al. (1985) investigated the possible existence of the 'hot hand' in the world of basketball almost 35 years ago, a substantial number of them dedicated their time trying to solve this question. If someone is experiencing a 'hot hand', typically is meant that this person is 'on a roll' or 'in the zone' which are analogous terms used to point out that somebody is in the middle of a streak of consecutive successes and because of that, the likeliness of extending that streak becomes considerably larger.

Analysts started introducing the 'hot hand' phenomenon in the game of basketball where fans and players are inclined to think the probability of scoring a shot is higher if it follows a previous scored shot than if it follows a previous missed shot (Gilovich et al., 1985). Statistically defined, the different shots are not considered to be independent; they are considered autocorrelated. Gilovich et al. (1985) investigated the actual existence of the 'hot hand' by performing multiple experiments with both professional and recreational basketball players and found evidence that the phenomenon is nothing more than a widely believed fallacy. People tend to suppose there is such a thing as the 'hot hand' due to an inferior understanding of the laws of chance regarding big and small sequences and how these sequences represent the way they are generated[4]. Additionally, another general misconception called the memory bias is partly responsible for the belief in the existence of a 'hot hand' (Gilovich et al., 1985). This bias implies that streaks of consecutive hits and misses are more likely to be remembered than a sequence of shots where hits and misses alternate. As a consequence, the number of times streak shooting occurs, is overestimated. Referring back to the evidence, Gilovich et al. (1985) found that the number of sequences including consecutive hits among basketball players did not significantly differ from the expected number based on a binomial model assuming that the possibility of hitting a shot is equal to the possibility of missing a shot. If there is no significant disparity between the expected and real number of consecutive hits, what does explain the differences in an individual's scoring discrepancy between different periods? A real increase in performance level is not the reason why in some periods a player scores more in comparison to other periods, but the success rate of previous shots could affect the eagerness to shoot and increase the times this player receives the ball from fellow teammates after which he will score more eventually (Tversky &

[4] A process that follows a random walk generates sequences which are expected to display the characteristics of that process (Tversky, & Kahneman, 1974).

Gilovich, 1989). The increased scoring is however not directly attributable to players having a ‘hot hand’.

These results were first questioned by Larkey, Smith, and Kadane (1989), who modified the existing dataset of Gilovich et al. (1985). Larkey et al. (1989) argued that measuring longer sequences in a real game was too challenging wherefore they limited the sequence length of an individual player in which he could experience having a ‘hot hand’ to 20 shooting attempts. This alteration of research method resulted in a remarkable finding headlining Vinnie Johnson from the Detroit Pistons. He showed signs of streak shooting which caused Larkey et al. (1989) to view him as a living example of an athlete who was undeniably able to experience having a ‘hot hand' and moreover, the first one who could achieve this during a real game. At the end of their research, Larkey et al. (1989) reinforced the feeling that belief in the existence of a ‘hot hand’ among athletes was rightfully grounded, something not considered possible at all by Gilovich et al. (1985).

2.3 Streak selection bias

Recent research, carried out by Miller and Sanjuro (2018), is in line with the conclusion of Larkey et al. (1989) but discovered the belief in the existence of having a ‘hot hand’ to be correct due to a mistake in the method of selection by Gilovich et al. (1985). In other words, having a ‘hot hand’ is an event that actually does exist. Consider every throw by a basketball player in the Gilovich et al. (1985) experiment as an i.i.d.[5] Bernoulli trial in which there are two outcomes: ‘success’, or in this example ‘hit’, and the outcome of ‘failure’, or in this example ‘miss’ where p denotes the probability of success and q denotes the probability of failure (Uspensky, 1937). Following from the two events being complementary, the sum of these probabilities equals 1. Miller and Sanjurjo (2018) give evidence suggesting that sequences, where every outcome follows a Bernoulli trial, impose the fraction of successes regarding outcomes directly following consecutive successes is estimated lower than the fundamental probability of success. This finding can be more clearly illustrated with the aid of an example. Imagine a sequence of three flips with a coin where you are investigating the fraction of heads (H) following a previous flip of heads or multiple successive flips of heads. The likeliness to flip heads or tails (T) in this example is the same, namely 0.5. The average of the proportion of all possible sequences where heads follows previous heads is summarised in the following table.

[5] Independent and identically distributed

TABLE 1 - STREAK SELECTION BIAS VISUALIZED BY COIN FLIP EXAMPLE

sequence	Proportion of heads (H) following heads (H)
TTT	-
TTH	-
THT	0
HTT	0
THH	1
HTH	0
HHT	1/2
HHH	1
Average	5/12

Note, this table signifies the averaged expected fraction of heads (H) following a flip of heads directly where the underlined letter signifies the outcome of the flip. Retrieved from Miller and Sanjurjo (2018).

Here you can see that the average fraction of heads following one or more successive heads is lower than the underlying probability to flip heads which is 0.5. Summarized, the main difference between the biased conventional research method and the corrected research method by Miller and Sanjurjo (2018) is the altering in the assumed probabilities for success if it follows immediate previous success. Gilovich et al. (1985) presumed $P(hit \mid k\ hits) - P(hit \mid k\ misses) = 0$ explaining an unaltered probability of success regardless of history while Miller and Sanjurjo (2018) presume that $P^i(hit \mid k\ hits) - P^i(hit \mid k\ misses) < 0$ revealing a decreased probability of success when the outcome directly follows previous obtained success(es). The magnitude of this bias diminishes when the sequence length increases and increments as a consequence of an increase in the number of consecutive successes. This will be further highlighted in the methodology section. I will incorporate the latest view of Miller and Sanjurjo (2018) into my research method and come up with newly calculated conditional probabilities of becoming a winner (loser) when a streak of becoming a winner (loser) for consecutive periods is experienced. These conditional probabilities will be called the corrected benchmark. That there is no clear answer to the question if a 'hot hand' exists among managers in the active trading environment yet could be due to the streak selection bias. This thesis could contribute to existing literature if results with the corrected benchmark from Miller & Sanjurjo (2018) differ from the results of existing literature which adopt the often-used canonical benchmark referring to the probability of 0.5 that does not take the streak selection bias into account.

2.4 Hypotheses

In order to aid in answering to the comprehensive research question, hypotheses are developed in this section which will serve as a guidance throughout this paper. Continuing on discussed literature, a substantial number of researchers investigated performance persistence in U.S. equity funds. A few papers that analysed performance persistence in equity funds are highlighted in this section, all of which employ the canonical benchmark expecting that the probability of becoming a winner or loser in the subsequent period is fixed and quantified as 0.5. Malkiel (1995) studied the existence of performance persistence exploring U.S. equity funds over the 1971-1991 period. He found that performance persisted from one year to the next one but merely in the decade of the 1970s. Kahn and Rudd (1995) looked at performance persistence in U.S. equity funds and examined if performance persisted between a 4-year period and a 3-year period. So, if funds are winners (losers) in the first four-year period, are they likely to be winners (losers) in the next 3-year period? The outcomes of the research did not give evidence of performance persistence. Some researchers explored the existence of performance persistence in equity funds outside the U.S., but the inferences related to their results differ substantially. Persistence in performance is sometimes found from one year to the other (Jan & Hun, 2004; Filip, 2004). On the other hand, similar research failed to find evidence indicating that performance persists into the next year (Casarin, Piva & Pelizzon, 2001; Iqbal & Tauni, 2016). Jan and Hun (2004) discover performance persistence for funds being winners in the last three years in regard to the following three years. Contrary, no performance persistence was found in Spanish equity funds when a performance between March 1997 and October 1999 was compared to their performance between November 1999 and June 2002 (Vicente & Ferruz, 2005). As the existing literature concerning equity funds is not uniform in their view towards the existence and length of performance persistence, the following two hypotheses are developed:

Hypothesis 1 (H1): there is no short-term persistence (≤2 year) under the canonical benchmark.

Hypothesis 2 (H2): there is no long-term persistence (>2 year) under the canonical benchmark.

As described briefly in the introduction, short-term performance persistence is present if performance persists at least for one quarter, i.e. three months, and a maximum of eight quarters, i.e. two years. Choosing a maximum of two years for short-term persistence is in line with a previous paper by Jackowicz and Kowalewski (2011) investigating short and long-run performance persistence in the Central-European banking industry. Long-term performance persistence is present if performance persists for at least nine quarters, i.e. two years and three months. The latter implies that all the four streak lengths (1, 2, 3, 4, 5, 6, 7 and 8) need to indicate performance persistence with regard to the next period. Since there exists no literature assessing the impact of the corrected benchmarks on financial data yet, it is tricky to develop a train of thought about the latter. This is why a cautious approach towards the adaption of the Miller and Sanjurjo (2018) is appropriate. The following two hypotheses capture this:

Hypothesis 3 (H3): there is no short-term persistence (≤2 year) under corrected benchmarks.

Hypothesis 4 (H4): there is no long-term persistence (>2 year) under corrected benchmarks.

3. Methodology

This chapter describes the empirical methods employed in this paper. In subsection 3.1, the designed models are discussed to extract the measure of mutual fund performance on which every period's relative ranking of mutual funds is based. Subsequently, the use of contingency tables is discussed subsection 3.2. Next, in subsection 3.3, the use of binomial tests is clarified which simultaneously serve as a robustness check of the main findings. In subchapter 3.4 the corrected benchmarks are deliberated. Finally, in paragraph 3.5, the use of manual calculations is presented.

3.1 Estimating performance measure alpha

For the purpose of uncovering if performance persistence is present among mutual funds, an appropriate measure that embodies a mutual fund's ability to conduct relative profitable portfolios is required. An essential specification of this explicit indicator is that it should merely characterize the contribution at the hands of the mutual fund, not the comprehensive influence driven by market factors. This said, numerous performance gauges are capable of serving as an instrument to examine mutual fund performance persistence often yielding similar results. In 1961, the first proxy for measuring fund performance was developed aiding in the understanding of average excess fund returns relative to its risk commonly known as the Treynor Ratio (Treynor, 1961). The Treynor ratio measures the relative volatility between the market (beta[6]) and the portfolio. Shortly after this discovery, Bill Sharpe adjusted the Treynor Ratio by replacing beta with the standard deviation of the portfolio (Sharpe, 1966). This is what we now know as the Sharpe Ratio. During the years, various asset pricing models are developed, some with more explanatory variables than others. Also looking at risk-adjusted returns of mutual funds and additionally comparing it to a related benchmark, the multi-period model of Jensen enabled him to introduce alpha as a performance measure which is referred to as Jensen's alpha and revealed the forecasting ability of a manager or fund to invest in lucrative stocks (Jensen, 1968, 1969). This was an extension of the initial single-period capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). Jensen's market model is the foundation of further performance measurements. Fama and French (1993) added two risk factors to the traditional baseline CAPM which represent asset classes performing substantially better than the aggregated market naming it the three-factor model. One factor reflected the exposure of a portfolio to the returns of companies with a small market capitalization and the other the returns of company stocks with a high book-to-market ratio or value stocks. Due to the wrongful belief in performance persistence by Hendricks et al., Carhart (1997) added another risk factor to the model which explained the short-term persistence in mutual fund performance and called this the momentum effect extending it to a four-factor model. Conforming the scope of this paper that is keen to illustrate the mutual fund distinctive share of its

[6] Representing the stock's systematic risk compared to the market.

total returns, picking one of the four discussed CAPM's seems to be evident. Since half of this paper investigates short-term persistence in mutual fund performance, the momentum factor cannot be ignored in the model as it is vital in avoiding biased results. Thence, the Carhart four-factor model (1997) is the most appropriate model to us.

A vast body of research preceded in using the model for measuring mutual fund performance (Bollen & Busse, 2004; Tonks, 2005; Ter Horst, Nijman & Verbeek, 2001). As mentioned earlier, the model of Carhart was implemented heavily within the field of expertise as it captures an important factor in explaining short-term performance persistence supplementing the supposedly insufficient Fama-French three-factor model (1993). The four factors from the Carhart four-factor model explain a considerable fraction of the fluctuation in transversal stock returns. The model takes on the following shape:

$$R^s(t) = \alpha^s + RF(t) + \beta_{MKT} * MKTRF(t) + \beta_{SMB} * SMB(t) + \beta_{HML} * HML(t) + \beta_{UMD} * UMD(t) + \varepsilon(t) \qquad (1)$$

where $R^s(t)$ is the monthly return per share as *s* stand for share, $RF(t)$ represents the risk-free rate of return (monthly treasury bill rate) and α^s (alpha) is the inexplicable portion of return that is not explained by market factors, hence the fraction that is attributable to mutual fund's ability to generate better than average portfolios, $MKTRF(t)$ is the value-weighted excess market return on all index stocks, $SMB(t)$ – Small Minus Big – is the difference between small-sized companies' return and big-sized companies' return, $HML(t)$ – High Minus Low – stands for the difference between the return of value stocks with a high book-to-market ratios and growth stocks with lower book-to-market ratios, $UMD(t)$ – Up Minus Down – means the difference in excess returns of the top half performing mutual funds 'winners' and value losses of the bottom half performing funds 'losers' based on *alpha* (α) and $\varepsilon(t)$ is the error term, all at time t.

The next step involves the separation of an out-of-sample period and an in-sample period in line with existing literature on mutual fund performance (Kahn & Rudd, 1995; Morey, 2003; Huij, & Derwall, 2008). An in-sample period is used to fit the model whereas the out-of-sample period is used to perform tests on. For the in-sample period, a standard ordinary least squares regression (OLS) suffices making use of the abovementioned variables except the risk-free rate which does not comes into place promptly. Elucidated, the total monthly return per share will be regressed on the excess return on the market portfolio (MKTRF), the size effect (SMB), the value premium (HML) and the momentum factor (UMD) over the 2010–2014 period. The obtained beta coefficients are utilised to predict out of sample total returns. To end up with the excess return part which is explained by the hand of the mutual fund, the difference in the actual total return and the predicted total return minus the risk-free rate signifies alpha. The formula is given below:

$$y^s(t) - \hat{y}(t) - RF(t) = \alpha^s(t) \tag{2}$$

where y^s(t) is the actual total return, $\hat{y}$ is the predicted total return, *RF*(t) is the risk-free rate and a^s(t) is the constant expressing the fund's ability to outperform its rivals, all at time t. Winners for any given period will be the top half performing mutual funds and losers will be the bottom half performing mutual funds for any given period based on *alpha (α)*. Every period the median of alpha will be determined on which the ranking of funds with respect to alpha is based.

3.2 Contingency tables & Pearson's chi-squared test

To conduct the main findings of this research, this paper will use 2x2 contingency tables combined with the Pearson's chi-square tests of independence to determine if two categorical variables are independent or related to each other, i.e. if there is performance persistence or not. The two categorical variables are:

1: Winning streaks and losing streaks

2: Winner or loser in the period following a streak

The null hypothesis of the Pearson's chi-square test is: Variable 1 and 2 are independent which in this case means there is no performance persistence. To illustrate the basic concept concerning the use of contingency tables, an example is given of how performance persistence among mutual funds between two periods can be measured. Imagine there are only two periods, period 1 and period 2, and we want to find out if mutual fund performance persists in these two periods.

TABLE 2 - CONTINGENCY TABLE EXAMPLE 1

		Period 2	
		W	L
Period 1	W	8 (5)	2 (5)
	L	2 (5)	8 (5)

Expected frequencies in parentheses
$\chi^2 = 7.2$

Note, W are winners and L are losers for a given period following section 3.1. The numbers which are not parentheses in every cell indicate how often a combination of the two categorical variables actually appeared. This is called the observed frequency.

The deviation of the actual frequency from the expected frequency is used to compute the test statistic. A simplified version of the original function used to calculate the test statistic can be displayed like this (Pearson, 1900):

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}} \tag{3}$$

where X^2 is the value of the test statistic, $\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(O_{i,j} - E_{i,j}\right)^2}{E_{i,j}}$ symbolises the aggregate value of all the observed frequencies minus the expected frequencies squared and divided by the expected frequency for each particular cell.

The test statistic is:

$(8 – 5)^2 / 5 + (2 – 5)^2 / 5 + (2 – 5)^2 / 5 + (8 – 5)^2 / 5 = 7.2$

The test statistic of the Pearson's chi-square test is compared to the critical values in appendix A. The degrees of freedom equal one which can be calculated using the following equation multiplication:

$$(r - 1)(c - 1) \tag{4}$$

where r are the number of rows and c are the number of columns. Hence, a 2x2 contingency table works with one degree of freedom[7]. Say, we want to see if the obtained test statistic is significant at a 0.05 significance level. The threshold value of the chi-square distribution with one degree of freedom at a 0.05 significance level is 3.481. As 7.2 > 3.481, we can reject the null hypothesis stating that the two categorical variables are independent. Therefore, performance persistence between period 1 and 2 can be confirmed at a 0.05 significance level. Important to note, if the observed frequencies would be reversed so we have 8 WL, 2 WW, 8 LW and 2 LL, instead of 8 WW, 2 WL, 8 LL and 2 LW, the test statistic would be the same and the null hypothesis would be rejected. Now however, performance persistence cannot be confirmed as winners in period 1 are now more likely to be losers in period 2 and losers in period 1 are more likely to be winners in period 2. This indicates a relationship between the two categorical variables that is the opposite of performance persistence.

The chi-square test is a non-parametric test often used to give insight at a nominal scale (McHugh, 2013). The main benefit of this approach is the nonrequirement of the homoscedasticity assumption which means that the error terms across all the independent variables are not necessarily expected to be equal. The use of contingency tables to is adopted by researchers before (Huij & Derwall, 2008; Kahn & Rudd, 1995). However, instead of dividing the time span in only two periods, this paper separates the total sample period

[7] (2-1) * (2-1) = 1

into 16 periods of equal length. This way, performance persistence is measured 15 times using separate contingency tables for every possible streak length. To visualize this approach, another example is given investigating performance persistence in mutual funds experiencing a streak of 10 wins or losses with regard to the next period.

TABLE 3 - CONTINGENCY TABLE EXAMPLE 2

		Period 11	
		W	L
Period 10	WWWWWWWWWW	80 (70)	60 (70)
	LLLLLLLLLL	64 (70)	76 (70)

Expected frequencies in parentheses
$\chi^2 = 3.9$

Note, WWWWWWWWWW and LLLLLLLLLL are winning and losing streaks concerning every possible streak of 10 periods of consecutive winners or losers starting at any point in the given time span of the total of 16 periods.

This contingency table captures all the mutual funds which experience a streak of 10 wins or losses and, in addition, have an observation for the following period. To be clear, as this paper has a maximum of 16 periods, this streak could only start in period 1, 2, 3, 4 or 5 if it has an observation for the next period. Just like this example, all streak lengths – 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 – are investigated.

3.3 Binomial tests

In addition to the use of contingency tables and Pearson's chi-squared test, this paper makes use of one-sided binomial tests. The advantage of this method is that it enables you to investigate winning and losing streaks separately, strengthening the validity of the results. To illustrate the rationale behind the binomial test, a brief explanation accompanied with an example will be given. The binomial test is used to check if the proportion of a sample is expected to be a certain value x. Only two possible outcomes can be obtained. Applied to the research of this paper, imagine the probability of 35 funds experiencing a winning streak of six periods to become a winner in the subsequent period is 0.5. Becoming a loser is the other possibility and has a probability of 0.5 too. The binomial test examines, by looking at the distribution of this sample, if it is likely that the distribution of winners and losers in the seventh period for this specific streak is 0.5. The formula that is used to compute the binomial test statistic is displayed below:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{5}$$

where k = the number of successes, n = the sample size and p = the probability of success or, in this specific case, the probability of becoming a winner in the seventh period. $\binom{n}{k}$ is computed as:

$$\frac{n!}{k!(n-k)!} \tag{6}$$

If 25 funds become a winner in the seventh period, the test statistic will be $P(X = 25) = \binom{35}{25}0.5^{25}(1-0.5)^{35-25} = 0.0083$. The probability of 25 or more out of 35 becoming a winner when the proportion of winners is 0.5 in the seventh period is 0,83%. The null hypothesis states that two possible outcomes are equally likely to happen, i.e. 0.5 and is rejected because 0.0083 < 0.05 which means that the actual proportion of funds becoming a winner in the seventh period after experiencing a streak of six periods of being a winner is higher than 0.5.

3.4 Corrected benchmark

Before tests are performed using corrected benchmarks, an important assumption needs to be made. As previously stated in the data section, funds get ranked by way of a relative ranking method. In every period, the median alpha is calculated whereafter funds get ranked as a winner when their alpha is above the median and as a loser when their alpha is below the median. The number of funds performing better than the median are by definition equal to the number of funds performing worse than the median (Kahn & Rudd, 1995; Huij & Derwall, 2007). External and internal drivers other than those discussed in the Carhart four-factor model influencing an investor's probability of being either a winner or a loser are ignored as that is beyond the scope of this study. Therefore, the probability of becoming a winner or a loser is 0.5. This is the canonical benchmark which does not take the streak selection bias into account. Under the null hypothesis, stating that the probability of becoming a winner or a loser is not different from 0.5, there is no performance persistence. On the other hand, the corrected benchmark does take the streak selection bias into account. This corrected benchmark is computed for every possible streak of 16-period sequences as the size of the streak selection bias varies in streak length. When assuming the corrected benchmark, the null hypothesis states that this corrected benchmark is the conditional probability when there is no performance persistence. The corrected benchmarks are conducted the same way as the Miller and Sanjuro (2018) example in the theoretical framework part for flipping a coin that could be either heads or tails with probability 0.5. The corrected benchmarks for a sequence of 16 periods is displayed below one the next page:

TABLE 4 - CORRECTED BENCHMARK	
Proportion of W after	expectation
1 W	7/15
2 W’s	213/533
3 W’s	254/717
4 W’s	64/189
5 W’s	76/227
6 W’s	113/337
7 W’s	79/234
8 W’s	164/481
9 W’s	29/84
10 W’s	47/134
11 W’s	327/913
12 W’s	59/160
13 W’s	37/96
14 W’s	5/12
15 W’s	1/2

Note that Proportion of W after = the corrected probability of being a winner in the next period after multiple consecutive wins are experienced.

3.5 Manual calculations

As discussed in the previous sector, the probability of a fund to perform better than its peers, is by definition 0.5 due to the median being the benchmark. Contingency tables computed by Stata do therefore not suffice. Stata evaluates the frequency of all possible outcomes and comes up with expected frequencies based on this distribution, whereas this paper assumes a probability of 0.5 of becoming a winner or loser for both winning and losing streaks beforehand. In order to still use contingency tables as a way of obtaining test results, they have to be modified manually. The Pearson chi-squared test also requires a manual approach. This can be clarified by a visual example. Take a look at the following contingency table:

TABLE 5 - MANUAL CALCULATION CONTINGENCY TABLE EXAMPLE

		Period 8 W	Period 8 L
Period 7	WWWWWWW	1,500 (1,585)	1,670 (1,585)
	LLLLLLL	1,843 (1,635)	1,427 (1,635)

Expected frequencies in parentheses

$\chi^2 =$

Note, WWWWWWW and LLLLLLL are winning and losing streaks concerning every possible streak of seven periods of consecutive wins or losses starting at any point in the given time span of 16 periods.

The expected values are calculated by adding up the actual winners and losers in period 8 for a specific winning or losing streak and dividing it by two. The test statistic is afterwards calculated as described in section 3.2. In order to make conclusions about performance persistence for both winning and losing streaks, both winners have to outlast losers for winning streaks and losers have to outlast winners for losing streaks. The test statistic can be computed as described in section 3.2.

The manual calculation regarding contingency tables where the corrected benchmarks are used, follows the same design as table 5 but now the expected frequencies for WWWWWWWW and LLLLLLLL are multiplied with the specific corrected benchmark of streak length seven (79/234). WWWWWWWL and LLLLLLLW are multiplied by (1 – 79/234).

4. Data

In this chapter, the data collection procedure and important adjustments made to the data set which allowed me to work with a legitimate dataset are explained in subsection 4.1 and 4.2 respectively.

4.1 Data selection

To give one the most recent view about the matter, a sample concerning US non-index equity fund returns over the 2015-2018 out-of-sample period will be investigated with an in-sample period dating from 2010 to 2014. As described in the introduction, the US mutual fund industry is with 18,408.33 billion euros the largest in terms of total absolute asset value which makes it the most enhanced mutual fund market in the world ("Value of assets of mutual funds in selected countries worldwide in 2018," 2018). May this paper yield surprising discoveries, then it concerns the biggest group of investors reachable. Moreover, it is, to the best of my knowledge, the most complete data set. This research is scoped by excluding all funds but equity funds. As stocks are more volatile than bonds, every choice made by a mutual fund affects the profit more substantially. The impact on future investments of any pioneering scientific findings of this paper will therefore have bigger consequences. A variety of databases exists whence data about the security market can be withdrawn from. Such as, Morningstar, CRSP or Datastream. For this paper, the CRSP (Center for Research in Security Prices) database, which covers a great deal of general and financial information about US mutual funds from 1962 onwards, is seized. The main advantage of using CRSP rather than other databases is that it offers a survivorship-bias-free sample (Elton, Gruber & Blake, 2001)[8]. In the first 25 years of research, papers investigating performance persistence often suffer from survivorship bias (Brown, Goetzmann & Ibbotson, 1992). Survivorship bias arises when firms go bankrupt and are no longer included in the performance analysis as a result of extinction. Databases which suffer from survivorship bias only include funds which have observations throughout the complete duration of the sample period. Consequently, the results of the study which are prone to this bias could be overestimated. Databases Since CRSP offers a data set which includes observations of funds that are not required to have observations throughout the complete duration of the sample period causing the ranking of funds to be as accurate as possible. Funds with only one observation are also included in the data set even though they don't qualify to be included in both contingency tables and binomial tests since they lack the opportunity to experience a streak. These observations are however incorporated in the ranking of funds as it guarantees survivorship-bias-free results.

[8] Elton et al. (2001) argue that still two problems arise when using the CRSP database, namely the omission bias which states that only small firms (<15 million total net asset) share monthly return data and the occasional overstatement of monthly returns due to numerous distributions on the same day. The omission bias is not an issue in this paper since big firms (see appendix B) share monthly return data nowadays too just like the upward-biased returns is not a problem as it would concern every mutual fund changing nothing to the rank order.

To make the data set compatible with this paper, the appropriate funds have to be selected and particular adjustments to certain variables have to be made. Two different type of data sets were acquired. One carried mutual fund specifications on a quarterly basis where the second presented monthly mutual fund and economic financial data. The original merged dataset has, after 3,290 duplicates were deleted, 3,347,511 unique observations, although not short of any misspecifications. The first distinct problem that was noticed concerned missing values. The variable expressing the monthly return per share included observations with a missing value or displayed a non-numerical value: 'R'. As a consequence, those data points were dropped resulting in a decreasing sample of 55,693 observations. Before the next piece of data-scrapping will be discussed, one is enlightened with the essential requirements every mutual fund has to satisfy. To make the data set useful within the scope of this paper, a few criteria (see appendix C) had to be evaluated aiding in answering the research question eventually. Starting off with an evident criterion that seeks explanation, is the fund objective criterion. In other words, what is the mutual fund's style? Does the mutual fund invest in equity or fixed-income securities? It is plausible that no fund invests without an objective. Nevertheless, the reality of that assumption is contradictory stumbling upon several mutual funds lacking an objective throughout the duration. Due to the absence of the possibility of deriving a fund's objective, all the non-relevant observations (732) were removed[9]. Since this paper looks at performance persistence in actively trading equity funds, fixed-income mutual funds (I, ICQH, ICQY, IF, IG, IGD, IGDI, IGDS, IGT, IM, IMM, IU, IUI and IUS) are erased together with mixed fixed-income and equity funds (M), mortgage-backed (OM) funds, currency funds (OC) or other funds (O). Consequently, 1,303,921 observations were excluded from the data set.

Continuing with a criterion that indicates if a mutual fund is indeed actively managed. A mutual fund is actively managed if its portfolio is not a complete copy of an expansive market index, such as the S&P 500, reducing the buying and selling of stocks. On the condition that a fund actively, manually picks stocks for its portfolio(s), a fund is entitled to be viewed as fairly better than its counterparts. By simple reasoning, if a fund flourishes by making a positive return on its portfolio(s) as he perpetually follows a passive strategy by tracking the S&P 500 index, no active changes in the portfolio are made in the coming periods taking away a funds eligibility to be regarded as better than its peers by making significant beneficial portfolio choices. For this reason, all the mutual funds that are pure – not actively managed – index funds (D) are not taken into consideration when the data is being analysed clearing out 179,999 observations[10] (see appendix A). Those mutual funds that are characterized as B or E remain in the dataset

[9] An important assumption is made in this paper regarding mutual funds that report an objective code on an irregular basis. For missing periods, the fund will take on the same style as the first more recent point in time where the fund's objective is available within observations of that particular fund.

[10] The aforementioned assumption in footnote 12 feels no urge to hold here as missing values imply a mutual fund is not depending on the index in defining the fund's investment strategy.

as they use the index as a tool to compose their portfolios but do not passively rely entirely on market indexes. Ostensibly, the selection procedure looks finished, yet some values are worth mentioning as their remarkable value in terms of magnitude cannot be neglected demanding a necessary act of revise. Based on Huij and Derwall (2008), five extreme outliers from the same fund with a monthly total return of more than 180 percent are detected and in so doing simultaneously cut out. Eventually, the final data set which is used to create alpha contains 1,807,161 singular observations spread over 27,718 mutual funds.

4.2 Data modifications

Monthly fund return data of 2010-2014 is used to fit the model. When the in-sample regression was stored in memory, a second new variable is developed which follows automatically from the out-of-sample prediction of the fitted model. This variable displays the predicted total returns of the out-of-sample periods. From here, the variable of interest is deducted. The variable of interest is alpha which serves as a performance measurement which is used to rank the funds afterwards. Note that these out-of-sample alphas are estimated monthly whereas this paper investigates persistence among mutual funds in multiples of three-month periods. Therefore, the alpha estimates are averaged into quarterly data. This compression decreased the data set to 609,288 observations[11]. The last variable that is essential in preparing the data set such that it is ready to be used in order to perform tests, is the one that indicates whether a mutual fund performed better or worse than the median calculated over all the funds in a given period[12]. Finally, all the observations from 2010–2014 are dropped as they are not to any use anymore which resulted in the ultimate data set used to examine my hypotheses consisting of 287,657 quarterly observations over the 2015–2018 period. As funds had to be detected which were experiencing a streak, a new variable was conducted for every period. Funds that were experiencing a streak received a different name. An overview of all variables from the final data set is given in table 4 below:

[11] This is not exactly one third of the monthly observations because some funds lacked the observations of all three months in certain quarters.

[12] In some periods, a few (<10) mutual funds could not be labeled as a winner or a loser as their value was identical to the median. This was however a marginal appearance and did not affect my research substantially.

TABLE 6 - OVERVIEW OF ALL THE VARIABLES OF THE FINAL DATA SET		
N = 287,657		
Variable	**Meaning**	**Description**
Quarter	Quarterly date	Date in Stata format counting from 1960
FundID	Fund identifier	Unique number for each mutual fund
Performance	Mutual fund performance measure	Value of alpha measuring mutual fund performance in percentages
WL	Winning or losing streak one period	Indicates whether mutual fund is winner or loser in every period
WL_2	Winning or losing streak two periods	Indicates whether mutual fund is winner or loser for two consecutive periods consecutively where WW is winner for two periods and LL is loser for two periods
WL_3	Winning or losing streak three periods	Indicates whether mutual fund is winner or loser for three consecutive periods where WWW is winner for three periods and LLL is loser for three periods
WL_4	Winning or losing streak four periods	Indicates whether mutual fund is winner or loser for four consecutive periods where WWWW is winner for four periods and LLLL is loser for four periods
WL_5	Winning or losing streak 5 periods	Indicates whether mutual fund is winner or loser for five consecutive periods where WWWWW is winner for five periods and LLLLL is loser for five periods
WL_6	Winner or loser period 6	Indicates whether mutual fund is winner or loser for six consecutive periods where WWWWWW is winner for six periods and LLLLLL is loser for six periods
WL_7	Winner or loser period 7	Indicates whether mutual fund is winner or loser for seven consecutive periods where WWWWWWW is winner for seven periods and LLLLLLL is loser for seven periods
WL_8	Winner or loser period 8	Indicates whether mutual fund is winner or loser for eight consecutive periods where WWWWWWWW is winner for eight periods and LLLLLLLL is loser for eight periods
WL_9	Winner or loser period 9	Indicates whether mutual fund is winner or loser for nine consecutive periods where WWWWWWWWW is winner for nine periods and LLLLLLLLL is loser for nine periods
WL_10	Winner or loser period 10	Indicates whether mutual fund is winner or loser for 10 consecutive periods where WWWWWWWWWW is winner for 10 periods and LLLLLLLLLL is loser for 10 periods
WL_11	Winner or loser period 11	Indicates whether mutual fund is winner or loser for 11 consecutive periods where WWWWWWWWWWW is winner for 11 periods and LLLLLLLLLLL is loser for 11 periods
WL_12	Winner or loser period 12	Indicates whether mutual fund is winner or loser for 12 consecutive periods where WWWWWWWWWWWW is winner for 12 periods and LLLLLLLLLLLL is loser for 12 periods
WL_13	Winner or loser period 13	Indicates whether mutual fund is winner or loser for 13 consecutive periods where WWWWWWWWWWWWW is winner for 13 periods and LLLLLLLLLLLLL is loser for 13 periods

WL_14	Winner or loser period 14	Indicates whether mutual fund is winner or loser for 14 consecutive periods where WWWWWWWWWWWWWW is winner for 14 periods and LLLLLLLLLLLLLL is loser for 14 periods
_1_WL	Winner or loser in every period brought forward in time one period	Shows for every period if fund was winner (W) or loser (L) but brought forward every observation one period.
_1_WL_B	Winner or loser in every period brought forward in time one period numeric	Shows for every period if fund was winner (1) or loser (0) but brought forward every observation one period.

Note, column 1 represents the variable name as it is used in the data set, column 2 explains the variable label and column 3 describes its use and how it is measured.

5. Results and discussion

The following chapter discusses the results of this research investigating the possible existence of performance persistence in US, non-index equity funds' return over the 2015–2018 period in the short and long run. As described in the methodology section, I will look at the outcomes of contingency tables period for period using Pearson's chi-squared test for independence with the canonical benchmarks first. After that, the contingency tables will be revised using corrected benchmarks following Miller's and Sanjurjo's (2018) research method. Afterwards, robustness tests are performed to validate the main findings using binomial tests for winners and losers separately in subsection 5.4. Lastly, in subsection 5.5, the results are contextualized by comparing it with the main findings of previous literature.

5.1 Contingency tables using the canonical benchmark

As described in the first part of this chapter, the results of various contingency tables, as portrayed in the methodology section, using canonical benchmarks will be discussed first. The null hypothesis of this method assumes that the two categorical variables are independent meaning there is no persistence and that becoming a being a winner or loser under this null hypothesis is an i.i.d. Bernoulli trial with a fixed probability not different from 0.5 regardless of any ongoing winning or losing streaks. As the total number of winners and losers in the subsequent periods for winning and losing streaks vary, the expected frequencies for those respective cells differ. The discrepancy in the number of total streaks of wins and losses held by certain funds that qualify for continuing their streak in the following period can be explained by the absence of a subsequent observation[13].

Although this paper examines possible winning or losing streaks up to 16 periods, contingency tables could only be reproduced when the total possible streak did not exceed a streak length of 13 periods. Remaining funds that were experiencing a 13-period streak of being a loser did not have any observations for the 14th period. Only six streaks of being a winner for 13 periods straight were followed up by a subsequent observation of being a winner or a loser. Because a table consisting of two cells is not eligible to perform a test, it is not regarded as a contingency table and as a result, no test statistic could be extracted. One contingency table will be highlighted to guide as an example of how the Pearson's chi-square test is conducted and what its implications are with regard to performance persistence. All the funds experiencing a streak of 4 periods of being a winner or loser with an observation in the subsequent period will be briefly discussed.

[13] The absence of subsequent observations for any particular streak can have various reasons, such as bankruptcy, change in investment vehicle(s) or a lack of data.

Streak of 4 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The null hypothesis of this test states that the two categorical variables are independent, i.e. there is no performance persistence. And, under this null hypothesis the fixed probability of becoming a winner in the subsequent period for funds experiencing a winning streak of four periods or the probability of becoming a loser in the subsequent period for funds experiencing a losing streak of four periods is not different from 0.5. The contingency table looks like this:

TABLE 7 - CONTINGENCY TABLE CANONICAL BENCHMARK

		Period 5	
		W	L
Period 4	WWWW	10,536 (9,867)	9,197 (9,867)
	LLLL	7,537 (9,945)	12,352 (9,945)

Expected values in parentheses
$\chi^2 = 1.3*10^3$

Contingency table for sequences of five periods, W = Winner, L = Loser in period 5, WWWW = Winner in period 1, 2, 3 and 4, LLLL = Loser in period 1, 2, 3 and 4 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4 and 5 starting from the first quarter of 2015 but represent every possible sequence of five following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark of 0.5.

The test statistic is:

$(10{,}536 - 9{,}867)^2 / 9{,}867 + (9{,}197 - 9{,}867)^2 / 9{,}867 + (7{,}537 - 9{,}945)^2 / 9{,}945 + (12{,}352 - 9{,}945)^2 / 9{,}945 = 1{,}256.5$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $1.3*10^3 > 6.635$ causes the H_0 to be rejected. As the number of winners is higher than the number of losers in period 5 for winning streaks and the number of losers is higher than winners in period 5 for losing streaks, it points in the direction of performance persistence. Since the first three contingency tables also confirmed performance persistence at a 0.01 significance level (see table 8), this confirms performance persistence up to four periods with the canonical benchmark at a 99% confidence interval.

How all the different contingency tables were calculated separately can be found in appendix D. Keep in mind, all the contingency tables from table 8 test the exact same null hypothesis as this specific example stating that the categorical variables are independent, i.e. there is no performance persistence and the fixed probability for becoming a winner (loser) when experiencing a winning (losing) streak under this null hypothesis is 0.5. An overview of all these contingency tables for different streak lengths and their implications are given in table 5 below:

TABLE 8 - CONTINGENCY TABLES CANONICAL BENCHMARK

Streak length	N	X^2	H_0 rejected	OW>EW	OL>EL	Persistence
1	265,155	$1.9*10^3$***	Yes	Yes	Yes	Yes
2	134,676	$1.6*10^3$***	Yes	Yes	Yes	Yes
3	69,964	$2.9*10^3$***	Yes	Yes	Yes	Yes
4	39,622	$1.3*10^3$***	Yes	Yes	Yes	Yes
5	21,408	$1.1*10^3$***	Yes	Yes	Yes	Yes
6	12,140	780.7***	Yes	Yes	Yes	Yes
7	7,213	179.9***	Yes	No	Yes	No
8	3,641	124.3***	Yes	Yes	Yes	Yes
9	2,044	117.3***	Yes	No	No	No
10	717	85.0***	Yes	No	Yes	No
11	248	1.7	No	No	Yes	No
12	97	61.8***	Yes	Yes	No	No

***p<0.01, **p<0.05. *p<0.1

Note that these are results of contingency tables for different streak lengths using the canonical benchmark of 0.5. Streak length = the number of periods for which funds are winners or losers successively, N = the total number of winning and losing streaks with an observation for the subsequent period, X^2 = test statistic of Pearson's chi-squared test for independence, H_0 is an indication if the null hypothesis stating that the two categorical variables are independent is rejected, OW>EW indicates if the number of wins for winning streaks is higher than expected, OL>EL indicates if the number of losers is higher than expected and Persistence = indication of persistence.

In order to confirm persistence, the H_0 has to be rejected, the number of wins has to be higher than expected (W>E=Yes) and the amount of losses has to be higher than expected (L>E=Yes), as is the case in the first six different streak lengths. For all streak lengths between 1 and 6 therefore, performance persistence is confirmed at a 0.01 significance level with a $1.9*10^3$, $1.6*10^3$, $2.9*10^3$, $1.3*10^3$, $1.1*10^3$, 780.7 test statistic respectively. When streak length increases with 1 period, the test result of 179.9 is still significant at a 0.01 significance level. However, persistence cannot be confirmed. More funds that are experiencing a losing streak tend to be a loser instead of winner indicating persistence. But, as funds experiencing a winning streak are more likely to become a loser in the subsequent period too, this does not indicate persistence. After the first six streak lengths, only streaks of eight periods show signs of performance persistence again at a 0.01 significance level. Because there is no indication of persistence for streaks of seven periods, we cannot say that performance persists for eight periods.

Summarised, performance persists for six periods. On average, funds that are performing good (bad) are more likely to perform good (bad) in the future when they are experiencing a winning (losing) streak of 6 periods or less. The first hypothesis (H1) stating that under the canonical benchmark of 0.5 there is no short-run performance persistence, is rejected. For streak lengths 1, 2, 3, 4, 5 and 6 performance persist significantly at a 0.01 significance level. Thus, performance under the canonical benchmark persists in the short run for six periods. The second hypothesis (H2) stating that under the canonical benchmark there is no long-run performance persistence, is not rejected as performance persists for not more than six periods consecutively. To see what the effect of the corrected benchmark is, the contingency tables are manually recalculated in the following chapter.

5.2 Contingency tables using the corrected benchmarks

Without the use of corrected benchmarks, we saw that performance persists for six consecutive periods. Miller and Sanjurjo (2018) pointed out that standard benchmarks could be misleading, because the conditional probability of being a winner in the next period is almost always – except for streaks of 15 periods of being a winner or being a loser – lower than the standard benchmark of 0.5 when the outcome is part of a streak of wins. The same goes for losing streaks. The null hypothesis of this method assumes that the two categorical variables are independent meaning there is no persistence and becoming a winner (loser) when experiencing a winning (losing) streak under this null hypothesis the conditional probability is not different from the length specific corrected benchmark. To implement this finding, the canonical benchmarks are displaced by the corrected benchmarks of section 3.4 and are used to create newly adjusted contingency tables. One contingency table will be highlighted to guide as an example of how the Pearson's chi-square test is conducted and what its implications are with regard to performance persistence. All the funds experiencing a streak of seven periods of being a winner or loser with an observation in the subsequent period will be discussed.

<u>Streak of 7 wins or losses.</u>

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of seven with regard to the subsequent period is 79/234 which can be found in section 3.4. The null hypothesis of this test states that the two categorical variables are independent, i.e. there is no performance persistence, and under this null hypothesis the probability of becoming a winner (loser) in the subsequent period for funds experiencing a winning (losing) streak of seven periods is not different from 79/234. The adjusted table looks like this:

TABLE 9 - CONTINGENCY TABLE CORRECTED BENCHMARK

		Period 8 W	Period 8 L
Period 7	WWWWWWW	1,309 (1,418)^{O} (957)^{N}	1,527 (1,418)^{O} (1,879)^{N}
	LLLLLLL	1,766 (2,189)^{O} (2,899)^{N}	2,611 (2,189)^{O} (1,478)^{N}

New(N) and old(O) expected values in parentheses
$\chi^2 = 1.5*10^3$

Contingency table for sequences of eight periods, W = Winner, L = Loser in period 8, WWWWWWW = Winner in period 1, 2, 3, 4, 5, 6 and 7, LLLLLLL = Loser in period 1, 2, 3, 4, 5, 6 and 7, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7 and 8 starting from the first quarter of 2015 but represent every possible sequence of eight following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark of 0.5. The new expected values for WWWWWWWW, WWWWWWWL, LLLLLLLW and LLLLLLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$(1{,}309 - 957)^2 / 957 + (1{,}527 - 1{,}879)^2 / 1{,}879 + (1{,}766 - 2{,}899)^2 / 2{,}899 + (2{,}611 - 1{,}478)^2 / 1{,}478 = 1{,}506.7$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $1.5*10^3 > 6.635$ causes the H_0 to be rejected. As the number of winners is higher than the number of losers in period 5 for winning streaks and the number of losers is higher than winners in period 5 for losing streaks, it points in the direction of performance persistence. Since the first six contingency tables also confirmed performance persistence at a 0.01 significance level (see table 10), this confirms performance persistence up to seven periods with the canonical benchmark at a 99% confidence interval.

How all the different contingency tables were calculated separately can be found in appendix E. Keep in mind, all the contingency tables from table 10 test the same null hypotheses stating that the categorical variables are independent, i.e. there is no performance persistence. However, every conditional probability for becoming a winner (loser) when experiencing a winning (losing) streak under this null hypothesis differs with every streak length. An overview of the corrected contingency tables is given below in table 6:

TABLE 10 - CONTINGENCY TABLES CORRECTED BENCHMARKS						
Streak length	N	X^2	H_0 rejected	OW>EW	OL>EL	Persistence
1	265,155	$6.1*10^3$***	Yes	Yes	Yes	Yes
2	134,676	$13.4*10^3$***	Yes	Yes	Yes	Yes
3	69,964	$18.4*10^3$***	Yes	Yes	Yes	Yes
4	39,622	$10.4*10^3$***	Yes	Yes	Yes	Yes
5	21,408	$6.9*10^3$***	Yes	Yes	Yes	Yes
6	12,140	$4.6*10^3$***	Yes	Yes	Yes	Yes
7	7,213	$1.5*10^3$***	Yes	Yes	Yes	Yes
8	3,641	$1.0*10^3$***	Yes	Yes	Yes	Yes
9	2,044	26.4***	Yes	Yes	Yes	Yes
10	717	70.5***	Yes	No	Yes	No
11	248	33.0***	Yes	Yes	Yes	Yes
12	97	37.5***	Yes	Yes	No	No

***p<0.01, **p<0.05. *p<0.1

Note, these are results of contingency tables for different streak lengths using corrected benchmarks. Streak length = the number of periods for which funds are winners or losers successively, N = the total number of winning and losing streaks with an observation for the subsequent period, X^2 = test statistic of Pearson's chi-squared test for independence, H_0 rejected indicates if the null hypothesis stating that the two categorical variables are independent is rejected, OW>EW indicates if the number of wins for winning streaks is higher than expected, OL>EL indicates if the number of losers is higher than expected and Persistence = indication of persistence.

Comparing the outcomes of table 10 for streak lengths one till six using the corrected benchmarks with the test results where the canonical benchmark was used, we see no change in outcomes regarding persistence. However, when we look at the number of periods for which performance persists uninterruptedly, we see a prolongation of three periods. Whereas under the canonical benchmark performance persisted up to six periods, the use of the corrected benchmarks makes performance persist up to 9 periods. More detailed, the first change takes place at a streak length of seven periods. The corrected benchmark alters the observed frequencies in relation to the expected frequencies for winning streaks when compared to the canonical benchmark. First, the observed frequency of winning streaks was less than the expected frequency while using the canonical benchmark. With the introduction of the corrected benchmark, the observed frequency exceeds the expected frequency. This, in addition to an increased indication of persistence in losing streaks, confirms performance persistence for seven periods at a 0.01 significance level. Performance persistence was already endorsed for streaks of eight concerning the next period when the canonical benchmark was used, and the corrected benchmark underlines this finding. Thus, until this point, performance persists for eight periods straight at a 0.01 significance level. Already lengthening the time span of significant performance persistence amongst mutual fund managers with two periods, the use of the corrected benchmark induces a prolongation of performance persistence even more. A test statistic of 26.4 lengthens performance persistence further for funds experiencing a winning or losing streak of nine periods at a 0.01 significance level. Performance persistence can be confirmed for nine successive periods before the series gets interrupted. The null hypothesis testing if 10-period winning or losing streak funds have an increased probability of continuing their streak, is rejected at a 0.01 significance level. As a result, persistence for 10 periods is not confirmed. For streaks of 11 periods signs indicating performance persistence can be noticed again. As the indication of performance persistence is lacking for streaks of 10 periods, we cannot infer that performance persists for 11 periods.

The third hypothesis (H3) stating that under the corrected benchmarks there is no short-run performance persistence, is rejected. In the short run, streak lengths of 1, 2, 3, 4, 5, 6, 7 and 8 periods show significant signs of performance persistence at a 0.01 significance level. Hence, performance under the corrected benchmark persists in the complete short run. The fourth hypothesis (H4) stating that under the corrected benchmarks there is no long-run performance persistence, is also rejected as performance persists up to 9 periods successively. Although the canonical benchmark and the corrected benchmark indicate performance persistence in the short run, persistence in the long run is only confirmed when the corrected benchmarks are used. Thus, adjusting for the streak selection bias extends the time span for which performance in mutual funds persist when using contingency tables.

5.3 Binomial tests using canonical benchmark

The second method that evaluates performance persistence among U.S. non-index mutual funds is the binomial test as formulated in the methodology section. In contrast to contingency tables, binomial tests enable us to look at positive and negative performance persistence separately. Induced suspicions about possible performance persistence in certain streak lengths that could not be confirmed, are scrutinised in this and the following sub-chapter.

5.3.1 Winning streaks

We start by looking at one-sided binomial tests of winning streaks where the canonical benchmark of 0.5 is used. Interesting to note, aside from the results, is that the longest streak length qualifying for a potential extended streak consists of 14 periods. This is two periods more than was possible using contingency tables now that we consider winning and losing streaks separately. Nonetheless, the sample size of the last three streak lengths are too small to say something meaningful about them. An overview of all the winning streaks with sufficient data is given below in table 7:

TABLE 11 - BINOMIAL TESTS WINNING STREAKS CANONICAL BENCHMARK

			p=0.5			
Streak	N	K	k	P	H_0 rejected	persistence
W	132,913	72,008	66,456.5	0.5417	Yes***	Yes
WW	67,697	37,883	33,848.5	0.5596	Yes***	Yes
WWW	35,770	20,622	17,885	0.5765	Yes***	Yes
WWWW	19,733	10,536	9,866.5	0.5339	Yes***	Yes
WWWWW	9,917	5,335	4,958.5	0.5379	Yes***	Yes
WWWWWW	5,051	3,012	2,525.5	0.5963	Yes***	Yes
WWWWWWW	2,836	1,309	1,418	0.4615	No	No
WWWWWWWW	1,159	665	579.5	0.5737	Yes***	Yes
WWWWWWWWW	627	275	313.5	0.4386	No	No
WWWWWWWWWW	258	55	129	0.2131	No	No
WWWWWWWWWWW	48	23	24	0.4791	No	No
WWWWWWWWWWWW	17	10	8.5	0.5882	No	No
WWWWWWWWWWWWW	6	3	3	0.5000	No	No
WWWWWWWWWWWWWW	1	1	0.5	1.0000	No	No

Note, these are binomial tests with the canonical benchmark of 0.5 for winning streaks, p = the probability for a fund to be a winner in the next period, streak = the streak length of winning periods, N is the total number of winners and losers together for the following period, K = the observed number of winners in the subsequent period, k = the expected number of winners in the subsequent period, P = the observed probability for a fund to be a winner in the next period ex post, H_0 indicates whether the null hypothesis stating that the proportion of winners in the following period is not different from 0.5, is rejected and persistence = the confirmation of persistence.

In line with the outcomes of the contingency tables of subsection 5.1, consecutive winners are likely to keep on winning up to six consecutive periods at a 0.01 significance level. Then, no significant result pointing out performance persistence can be found with concern to the next streak length. When funds are winners during seven consecutive periods, there is no reason to believe that the actual proportion of being a winner again in the next period is significantly higher than 0.5. Hence, the null hypothesis does not get rejected. Winning streaks of eight periods show signs of performance persistence again. However, performance persistence cannot be confirmed for eight periods as there is no evidence indicating performance persistence for seven periods. After this, the canonical benchmark fails to give evidence suggesting performance persistence for the remaining streak lengths. The deductions about performance persistence concerning winning streaks correspond with the use of the canonical benchmarks when the canonical benchmark is employed. That is to say, using a fixed probability of becoming a winner equal to 0.5 regardless of any ongoing streak, winning streaks persist significantly up to six periods. Referring to the developed hypotheses is done in the next part as positive and negative performance persistence are required to be discussed together. The hypotheses describe performance persistence in general which is about positive and negative performance persistence.

5.3.2 Losing streaks

Founding similar results for winning streaks while using contingency tables and one-sided binomial tests, we continue our research by looking at one-sided binomial tests involving losing streaks with the canonical benchmark. An overview of all the streaks of interest is displayed in the following table:

TABLE 12 - BINOMIAL TESTS LOSING STREAKS CANONICAL BENCHMARK

			p=0.5			
Streak	N	K	k	P	H_0 rejected	persistence
L	132,242	60,475	66,121	0.4573	Yes***	Yes
LL	66,979	30,239	33,489.5	0.4514	Yes***	Yes
LLL	34,194	12,939	17,097	0.3784	Yes***	Yes
LLLL	19,889	7,537	9944.5	0.3789	Yes***	Yes
LLLLL	11,491	4,047	5,745.5	0.3521	Yes***	Yes
LLLLLL	7,089	2,519	3,544.5	0.3553	Yes***	Yes
LLLLLLL	4,377	1,766	2,188.5	0.4034	Yes***	Yes
LLLLLLLL	2,482	993	1,241	0.4000	Yes***	Yes
LLLLLLLLL	1,417	904	708.5	0.6379	No	No
LLLLLLLLLL	459	226	229.5	0.4923	No	No
LLLLLLLLLLL	200	91	100	0.4550	No	No
LLLLLLLLLLLL	80	75	40	0.9375	No	No

Note that these are binomial tests with the canonical benchmark of 0.5 for losing streaks, p = the probability for a fund to be a winner in the next period, streak = streak length of losing periods, N is the total number of winners and losers together for the following period, K = the observed number of winners in the subsequent period, k = the expected number of winners in the subsequent period, P = the observed probability for a fund to be a winner in the next period ex post, H_0 indicates whether the null hypothesis stating that the proportion of winners in the following period is not different from 0.5, is rejected and persistence = the confirmation of persistence.

The results from the binomial tests of table 12 share similarities with the results from the canonical contingency tables. During the first six periods of a losing streak, funds are more plausible to continue their active losing streak an additional period at a 0.01 significance level. Contingency tables enforcing the same canonical probability of 0.5 already implied this. However, the dissimilarity between these two research methods, occurs at a streak length of seven periods. At first, referring to the usage of contingency tables, there was no evidence of persistence as the observed count of winning streaks suggested the opposite making it impossible to draw a general conclusion about performance persistence. But, as one can see now, the earlier suspicion about possible persistence in losing streaks of seven periods is confirmed. Streaks of being a loser in seven successive periods are significantly more likely to be extended an additional period. This outcome recurs as streaks of losing in eight periods consecutively are looked into. The proportion of losing streaks ending after eight consecutive periods is 0.4 which results in a rejection of the null hypothesis at a 0.01 significance level. Thus, based on binomial testing, performance persistence in losing streaks persist up to eight consecutive periods. Contrary to the contingency tables, this finding includes funds which showed performance persistence in their previous seven periods. After the streak length of eight losing periods, no signs indicating persistence can be found. The first hypothesis (H1) stating that under the canonical benchmark of 0.5 there is no short-run performance persistence, is rejected. For streak lengths 1, 2, 3, 4, 5, 6, 7 and 8 negative performance persist significantly at a 0.01 significance level with regard to the following period. Winning streaks show short-run persistence for six periods. Thus, performance under the canonical benchmark persists in the short run. The second hypothesis (H2) stating that under the canonical benchmark there is no long-run performance persistence, is not rejected. Positive performance persists up to six periods consecutively and negative performance persists up to eight periods uninterruptedly which is too short to be qualified as long-term persistence. To see what the effect of the corrected benchmark is, the binomial tests are recalculated in the following chapter.

5.4 Binomial tests using corrected benchmark

In this final part of the results, positive and negative performance persistence will be examined separately by adopting one-sided binomial tests whilst accounting for a present streak selection bias concerning the dependency of outcomes on previous obtained streaks. The corrected benchmarks are identical to the previous ones employed to modify the canonical contingency tables. The expected number of observations is the total number of actual observations multiplied by the corrected benchmark.

5.4.1 Winning streaks

To complement the evaluation of this research in which the impact of a supposedly existing streak selection bias on performance persistence is investigated, we apply the corrected benchmarks to one-sided binomial tests for winning streaks. After already finding thought-provoking results after the implementation of corrected benchmarks to contingency tables, it is important to validate these results by altering the used research method. An overview of the newly corrected binomial test including winning streaks only is presented below in table 9:

TABLE 13 - BINOMIAL TESTS WINNING STREAKS CORRECTED BENCHMARKS							
Streak	p	N	K	k	P	H_0 rejected	persistence
W	7/15	132,913	72,008	62,026.1	0.5417	Yes***	Yes
WW	213/533	67,697	37,883	27,053.4	0.5596	Yes***	Yes
WWW	254/717	35,770	20,622	12,671.7	0.5765	Yes***	Yes
WWWW	64/189	19,733	10,536	6,682.1	0.5339	Yes***	Yes
WWWWW	76/227	9,917	5,335	3,320.2	0.5379	Yes***	Yes
WWWWWW	113/337	5,051	3,012	1,693.7	0.5963	Yes***	Yes
WWWWWWW	79/234	2,836	1,309	957.5	0.4615	Yes***	Yes
WWWWWWWW	164/481	1,159	665	395.2	0.5737	Yes***	Yes
WWWWWWWWW	29/84	627	275	216.5	0.4386	Yes***	Yes
WWWWWWWWWW	47/134	258	55	90.5	0.2131	No	No
WWWWWWWWWWW	327/913	48	23	17.2	0.4791	Yes*	Yes
WWWWWWWWWWWW	59/160	17	10	6.3	0.5882	Yes*	Yes
WWWWWWWWWWWWW	37/96	6	3	2.3	0.5000	No	No
WWWWWWWWWWWWWW	5/12	1	1	0.4	1.0000	No	No

Note, these are binomial tests with corrected benchmarks for winning streaks, p = the probability for a fund to be a winner in the next period, streak = the streak of winning periods, N is the total number of winners and losers together for the following period, K = the observed number of winners in the subsequent period, k = the expected number of winners in the subsequent period, P = the observed probability for a fund to be a winner in the next period ex post, H_0 indicates whether the null hypothesis stating that the proportion of winners in the next period is not different from the corrected benchmark, is rejected and persistence = the confirmation of persistence.

The first nine streak lengths all indicate performance persistence between the respective period and its sequential period at a 0.01 significance level. Despite the actual proportion of winners in the subsequent period at streaks lengths seven and nine being lower than 0.5, the corrected benchmarks still reveal performance persistence. The lower corrected expected probability of becoming a winner in the next period if funds follow a certain streak of winning, means that the actual proportion of winners is not required to be higher than 0.5 to obtain significant results. Now that performance persistence is confirmed until the ninth period without insignificant interruptions for particular streak lengths, the outcomes of performance persistence for winning streaks of 10 periods is reviewed. The distinct portion of 0.2131 winners is too small to be of any significance. As a result, the null hypothesis cannot be rejected, and persistence cannot be confirmed. Unlike this being the first streak length without evidence pointing in the direction of performance persistence, the next streak length again give reason to believe that performance persistence is present. Above median performance perseveres when funds experience a 11-period winning streak at a 0.1 significance level. Performance persistence is however not confirmed up to 11 periods as the streak length of 10 does not indicate performance persistence. The last two streak lengths show a higher actual winner count than expected. Due to a small sample size performance persistence cannot be confirmed, nonetheless. Based on table 9, we can conclude that mutual funds perform significantly better than their peers for nine periods straight if all of their previous obtained results are better than the period specific median of alpha at a 0.01 significance level. How these results relate to the developed hypotheses is discussed in the next part as winning and losing streaks are required to be considered together to say something meaningful about the hypotheses. The hypotheses capture positive and negative performance persistence as a whole.

5.4.2 Losing streaks

In the final part of the results, individual losing streaks are tested for performance persistence while correcting for the streak selection bias. As already came forward in the part examining various contingency tables, the lack of data points among longer losing streaks enables this research to investigate performance persistence to a maximum of 12 consecutive periods. An overview of all the singular losing streaks and their binomial tests results is given in the following table:

TABLE 14 - BINOMIAL TESTS LOSING STREAKS CORRECTED BENCHMARKS

Streak	p	N	K	k	P	H_0 rejected	persistence
L	8/15	132,242	60,475	70,529.1	0.4573	Yes***	Yes
LL	320/533	66,979	30,239	40,212.5	0.4514	Yes***	Yes
LLL	462/717	34,194	12,939	22.080.6	0.3784	Yes***	Yes
LLLL	125/189	19,889	7,537	13,154.1	0.3789	Yes***	Yes
LLLLL	160/227	11,491	4,047	8099.4	0.3521	Yes***	Yes
LLLLLL	224/337	7,089	2,519	4712.0	0.3553	Yes***	Yes
LLLLLLL	155/234	4,377	1,766	2899.3	0.4034	Yes***	Yes
LLLLLLLL	317/481	2,482	993	1,635.7	0.4000	Yes***	Yes
LLLLLLLLL	55/84	1,417	904	927.8	0.6379	Yes*	Yes
LLLLLLLLLL	87/134	459	226	298.0	0.4923	Yes***	Yes
LLLLLLLLLLL	586/913	200	91	128.4	0.4550	Yes***	Yes
LLLLLLLLLLLL	101/160	80	75	50.5	0.9375	No	No

Note that these are binomial tests with corrected benchmarks for winning streaks, p = the probability for a fund to be a winner in the next period ex ante, streak = the streak of losing periods, N is the total number of winners and losers together for the following period, K = the observed number of winners in the subsequent period, k = the expected number of winners in the subsequent period, P = the observed probability for a fund to be a winner in the next period ex post, H_0 indicates whether the null hypothesis stating that the proportion of winners in the following period is not different from the corrected benchmark, is rejected and persistence = the confirmation of persistence.

Similar to the results of losing streaks within the binomial tests section adopting the canonical benchmark, table 10 shows performance persisting for the first eight streak lengths at a 0.01 significance level. A switch in deduction about the existence of performance persistence, compared to analogous binomial tests not correcting for the streak selection bias, takes place at streaks of nine periods. Out of 1,417 observations, funds experiencing a losing streak of nine periods do not become winners in the following period more often than 904 times. Consequently, the null hypothesis stating that the fraction of losing funds turning into a winner is not significantly different from 55/84 is rejected. Hence, performance persistence is confirmed at a 0.1 significance level. Then 459 streaks of solely losing qualify to possibly extend their streak. 223[14] of these continue being a loser which is enough to reject the null hypothesis confirming negative performance persistence at a 0.01 significance level. The majority of funds keep on losing in the 12th period after they were losers during 11 successive periods. Also, this result is significant at a 0.01 significance level. All in all, losing funds are more likely to keep on becoming losers in the following period if they are experiencing a losing streak up to eleven periods. Performance persistence in losing streaks is observable throughout 11 periods nonstop at a 0.1 significance level. The findings of the corrected binomial tests underline the findings by the corrected contingency tables. The third hypothesis (H3) stating that under the corrected benchmarks there is no short-run performance persistence, is rejected. For streak lengths 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 negative performance persists significantly at a 0.1 significance level. Regarding winning streaks, performance persists significantly for nine periods. Hence, performance under the corrected benchmark persists in the complete short run. The fourth hypothesis (H4) stating that under the corrected benchmarks there is no long-run performance persistence, is rejected as performance persists up to nine periods for winning streaks and up to 11 periods for losing streaks successively indicating long-term performance persistence.

All in all, the canonical benchmark of 0.5 showed short-term performance persistence but no long-term performance persistence. On the other hand, the corrected benchmarks displayed short and long-term performance persistence. This implies that the number of periods for which performance persists consecutively is higher when the corrected benchmarks are used. The corrected benchmarks, which are built on the conditional probability taking the streak selection bias into account, are lower than the canonical benchmark of 0.5. Briefly mentioned before, this corrected benchmark causes the actual number of funds continuing their streak for some streak lengths to be higher than expected, where earlier this was not the case, leading to a significant indication of performance persistence. With the canonical benchmark, the actual number of funds was not higher than expected. This explains the impact of the streak selection bias.

5.5 Discussion

This study investigates the impact of the streak selection bias on the existence and length of short and long-term performance persistence in U.S. equity funds. The outcomes of the corrected benchmarks are compared to the outcomes where the canonical benchmarks were used. The use of the canonical benchmark resulted in an indication of short-run persistence to some extent. Contingency tables show that performance persists for six periods under the canonical benchmark. On the contrary, long-term performance persistence cannot be confirmed. This however changes when the corrected benchmarks are used. Performance persists for nine periods under the corrected benchmarks indicating performance persistence in the short run completely and in the long run to some degree. The outcomes of the contingency

[14] 459-226

tables are underlined by the binomial tests that served as a robustness test. Although negative performance persists for eight periods when examined separately with the canonical benchmark, it does not alter the view towards performance persistence in the short and long run. The results of this paper give incentive to believe that performance persistence exists in the short term regardless of the chosen benchmark. There is however reason to believe that the corrected benchmarks do change the way the existence of long-term performance persistence is perceived. It is nevertheless important to put the latter into perspective as the framing of short and long-term persistence affects the interpretation of the results. What does become evident after this paper is that the implementation of the streak selection bias has an impact on the duration of performance persistence. If the streak selection bias is not taken into account by future researchers examining the existence of performance persistence, not incorporating the streak selection bias will most certainly lead to downward biased results. In some cases, the existence could be denied when in fact performance persistence does exist. The implementation of the streak selection bias showed that the differences in performance are not temporary but seem to be rather structural. An important implication, if this is indeed true, is that the number of competitors on the market will shrink as the worst performing funds will not survive in the future. As a result, the competitiveness will decrease, and mutual funds increase the fees charged to investors. The regulatory authority of financial markets could take measures to guarantee competitiveness in the market. For instance, create different levels of taxation where winners of a given period pay more in comparison to losers if they gain capital. The efficient market hypothesis states that stock prices are a combination of the information available and rational expectations of investors. Price fluctuations that are not induced by publicly available information can therefore not be explained and as a consequence cannot be predicted. As this paper has found that funds are able to gain more portfolio profit than 50% of their peers over consecutive periods, this could mean that some funds have an advantage over most of the other funds in the gathering of information, like inside information. This is however a firm suspicion which has to be taken rather lightly. This paper also found evidence of the underperformance of funds for successive periods what could point out that some mutual funds suffer from insufficient information gathering. If access to sufficient information about possible price developments causes the difference in performance, the government has to improve transparency in the market. If all mutual funds have the possibility to access the same pool of information, differences between funds' performance will become less severe.

6. Conclusion

Right now, there are probably thousands, maybe millions of investors weighing the pros and cons of investing in actively managed mutual funds versus investing in passively managed mutual funds. Depending on their attitude towards risk or their degree of risk aversion, they finally decide whether to invest in one of these two funds. Actively managed funds have a higher possible gain and passively managed funds are safer investment vehicles. Still, a lot remains unclear about the return on investment between the two types of mutual funds up until this day. As beating the market on a regular basis seems to be an impossible task due to inexplicable external factors, the potential ability of an actively managed mutual fund to be able to perform better than its peers continually appears to be a more vital investigation. A recent, thought provoking discovery by Miller and Sanjurjo (2018) about a present streak selection bias questioned the credibility of Gilovich et al.'s (1985) research about the 'hot hand' being a fallacy. This paper tries to give important insights about the impact of this streak selection bias on the existence and duration of persistence in mutual fund performance.

To improve the validity of this paper, a survivor-bias-free database is used, and two methodologies are employed, one of which simultaneously serves as a robustness check. Namely, 2x2 contingency tables and one-sided binomial tests respectively. Both research methods are first utilized the conventional way by taking 0.5 as the probability of performing better (worse) than half of their counterparts in the next period regardless of any ongoing winning (losing) streaks. This already gave evidence pointing out in the direction of performance persistence being existing. The contingency tables revealed performance persistence up to six periods consecutively. This confirms performance persistence being present in the short run but not in the long run when the canonical benchmark is used. Binomial tests showed performance persistence up to six periods for winning streaks and when it comes to losing streaks, performance persists up to eight successive periods. Subsequently, the same approach as before was adopted except the canonical benchmark was corrected for the streak selection bias. To correct for the streak selection bias, new benchmarks were computed taking the interdependency of previous consecutive winning (losing) periods on future probability of being a winner (loser) into account. Looking at the results of these kind of contingency tables intertwined with the corrected benchmarks, the time span for which performance persisted, extended for three periods confirming the presence of performance persistence in the short and long run. Considering binomial tests for winning streaks making use of corrected benchmarks, performance persists now up to nine periods straight which is a lengthening of three periods[15]. For losing streaks, performance persistence increases in duration with three additional periods from eight to 11 periods.

[15] Note that the use of the canonical benchmark already showed performance persistence in funds that were experiencing an eight-period streak of being a winner for the following period so technically only two additional streak lengths were added.

All in all, both the 2x2 contingency tables and one-sided binomial tests employed with the canonical benchmark already gave evidence of positive and negative performance persistence in the short run. In the long run, performance persistence could not be confirmed. This changes with the implementation of the streak selection bias. The corrected benchmark gave evidence indicating short and long-term performance persistence. Both binomial tests concerning winning and losing streaks and contingency tables showed an increase of three periods for which performance persistence lasts. As already discussed in the discussion part, this study reveals the importance of bringing forth the streak selection bias in the ongoing investigation about performance persistence in mutual funds. If researchers do not implement the streak selection bias, they can be at risk of coping with downward biased results or, in a worst-case scenario, they cannot find indications of performance persistence when indeed there are reasons to believe that it exists. As the results from the corrected benchmark illustrate that persistence in mutual fund performance is more a structural phenomenon than a temporary event, the worst performing funds could be in danger of bankruptcy. A reasonable assumption clarifying the continual difference in performance could be that the best performing funds have more or better stock-related information access. Information about stocks could help predict the future price. To diminish the gap between best and worst performing funds, the government could make the securities market more transparent. Companies listed on the stock exchange could be forced to share a half-yearly (financial) update about their product or service which contains the same information for everyone.

Although this paper found some interesting results, there are some limitations that have to be mentioned. The chosen time span is the first limitation of this research. Because one period has a length of three months, the number of funds that accomplish a streak of 16 periods of being a winner or loser is zero. If periods would have a length of six months, the number of funds experiencing a streak of being a winner or loser for four years could be higher. An approach like this would enable future research to say more about long-term performance persistence. A second shortcoming of this paper is visible in the methodology. The estimation of alpha could be not completely accurate. The in-sample period is used to fit the model of an OLS with the total return as the dependent variable. This model forecasts total returns for the out-of-sample period after which the period-specific risk-free rate of return is subtracted. The most preferable way would be to fit the model of an OLS with the excess return as the dependent variable and compare it with the real excess return, but this was not possible with this estimation method. However, since the results are based on rankings, this flaw does not affect the validity of this paper as it concerns every observation in the same way.

The capability of the streak selection bias, in case it is correctly implemented, to increase the belief of persistence in mutual fund performance being an existing phenomenon, suggests that it cannot be neglected. It is hard to make strong inferences about the magnitude of the streak selection bias as it depends considerably on the chosen period length and framing of short and long-term persistence. That the implementation of the streak selection bias has an effect at all, can nevertheless be confirmed. Since this research is scoped by examining equity funds only, it would be interesting to broaden the foundation that is now lain by exploring the bond fund market for example. Or, examine different kind of equity funds and see whether the effect is more or less substantial for certain types of those funds. Wandering if maybe in the distance future the importance of the streak selection bias is acknowledged in the financial environment as it could contribute to the conviction of the existence of relative performance persistence, one other question continues to exist. Will investors ever be able to continuously outperform the market?

References

Bollen, N. P., & Busse, J. A. (2004). Short-term persistence in mutual fund performance. *The Review of Financial Studies*, *18*(2), 569–597.

Bryan, A., & Li, J. (2016). Performance persistence among US mutual funds. *Analyst*, *1*(312), 384–4979.

Brown, S. J., & Goetzmann, W. N. (1994, July). Attrition and mutual fund performance. In *Journal of Finance* (Vol. 49, No. 3, pp. 1055-1056).

Brown, S. J., & Goetzmann, W. N. (1995). Performance persistence. *The Journal of finance*, *50*(2), 679–698.

Camerer, C. F. (1989). Does the Basketball Market Believe in the 'Hot Hand,'?. *American Economic Review*, *79*(5), 1257–1261.)

Carhart, M. (1995). Survivor bias and mutual fund performance. *University of Southern California working paper*.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57–82.

Carlson, K. A., & Shu, S. B. (2007). The rule of three: How the third event signals the emergence of a streak. *Organizational Behavior and Human Decision Processes*, *104*(1), 113–121.

Carlson, R. S. (1970). Aggregate performance of mutual funds, 1948–1967. *Journal of Financial and Quantitative Analysis*, *5*(1), 1-32.

Casarin, R., Piva, A., & Pelizzon, L. (2001). Italian equity funds: Efficiency and performance persistence.

Elton, E. J., Gruber, M. J., & Blake, C. R. (2001). A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *The Journal of Finance*, *56*(6), 2415–2430.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Filip, D. (2011). Performance persistence of equity funds in Hungary. *Contemporary Economics*, *5*(1), 18-34.

Goetzmann, W. N., & Ibbotson, R. G. (1994). Do winners repeat? Patterns in mutual fund return behaviour. *Journal of Portfolio Management*, *20*(2), 9-18.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, *17*(3), 295–314.

Grinblatt, M., & Titman, S. (1992). The persistence of mutual fund performance. *The Journal of*

Finance, *47*(5), 1977–1984.

Grinblatt, M., & Titman, S. (1994). A study of monthly mutual fund returns and performance evaluation techniques. *Journal of financial and quantitative analysis*, *29*(3), 419-444.

Gittelsohn, J. (2019). End of Era: Passive Equity Funds Surpass Active in Epic Shift. Retrieved from https://www.bloomberg.com/news/articles/2019-09-11/passive-u-s-equity-funds-eclipse-active-in-epic-industry-shift

Hendricks, D., Patel, J., & Zeckhauser, R. (1993). Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of finance*, *48*(1), 93–130.

Huij, J., & Derwall, J. (2008). "Hot Hands" in bond funds. *Journal of Banking & Finance, 32(4)*, 559–572.

Iqbal, A., & Tauni, M. Z. (2016). Performance persistence in institutional investment management: The case of Chinese equity funds. *Borsa Istanbul Review*, *16*(3), 146-156.

Jackowicz, K., Kowalewski, O., & Kozłowski, Ł. (2011). The Short and LongTerm Performance Persistence in the Central European Banking Industry. *Contemporary Economics*, *5*(4), 18-31.

Jan, Y. C., & Hung, M. W. (2004). Short-run and long-run persistence in mutual funds. *The Journal of Investing*, *13*(1), 67-71.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of finance*, *23*(2), 389–416.

Jensen, M. C. (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of business*, *42*(2), 167–247.

Kahn, R. N., & Rudd, A. (1995). Does historical performance predict future performance? *Financial Analysts Journal*, *51*(6), 43–52.

Koehler, J. J., & Conley, C. A. (2003). The "hot hand" myth in professional basketball. *Journal of Sport and Exercise Psychology*, *25*(2), 253–259.

Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the "hot hand". *Chance*, *2*(4), 22–30.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The journal of finance*, *20*(4), 587–615.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of finance*, *50*(2), 549-572.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica: Biochemia medica*, *23*(2), 143–149.

Miller, J. B., & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica*, *86*(6), 2019–2047.

Morey, M. R. (2003). Should you carry the load? A comprehensive analysis of load and no-load mutual fund out-of-sample performance. *Journal of banking & finance*, *27*(7), 1245–1271.

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157–175.

Rao, J. M. (2009). *Inference of Subjective Sequences: When the gambler's fallacy becomes the hot hand fallacy. Yahoo*. Research Labs, Mimeo.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425–442.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of business*, *39*(1), 119–138.

Shukla, R., & Trzcinka, C. (1994). Persistent performance in the mutual fund market: tests with funds and investment advisers. *Review of Quantitative Finance and accounting*, *4*(2), 115-135.

Sommer, J. (2019, July 9). Investors Are Usually Wrong. I'm One of Them. *The New York Times*. Retrieved from https://www.nytimes.com/2019/07/26/business/stock-bond-investing.html

Ter Horst, J. R., Nijman, T. E., & Verbeek, M. (2001). Eliminating look-ahead bias in evaluating persistence in mutual fund performance. *Journal of Empirical Finance*, *8*(4), 345–373.

Titman, S., & Grinblatt, M. (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Bussines*, *62*(3).

Tonks, I. (2005). Performance persistence of pension-fund managers. *The Journal of Business*, *78*(5), 1917–1942.

Treynor, J. L. (1961). Market value, time, and risk. *Time, and Risk (August 8, 1961)*.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

Tversky, A., & Gilovich, T. (1989). The cold facts about the "hot hand" in basketball. *Chance*, *2*(1), 16–21.

Uspensky, J. V. (1937). Introduction to mathematical probability.

Value of assets of mutual funds in selected countries worldwide in 2018. (2018). Retrieved from https://www.statista.com/statistics/270289/amount-of-fund-assets-in-selected-countries-of-the-world/

Vicente, L., & Ferruz, L. (2005). Performance persistence in Spanish equity funds. *Applied Financial Economics*, *15*(18), 1305-1313.

Appendix

APPENDIX A. CRITICAL VALUES OF THE CHI-SQUARE DISTRIBUTION

Upper-tail critical values of chi-square distribution with *v* degrees of freedom

v	0.9	0.95	0.975	0.99	0.999
1	2.706	3.481	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588

Table 15

APPENDIX B. TOTAL NET ASSETS SUMMARY STATISTICS

Variable	N	μ	σ	Min.	Max.
Total Net Assets	27,718	408.34	2,053.50	0.1	101,687

Table 16: Summary statistics of the total net assets aggregated at the mutual fund level, N = number of funds, μ = the mean in millions of U.S. dollars, σ = the standard deviation, min = the lowest recorded total net assets value, whereas 0.1 is an equivalent of all the reported net assets values < \$100,000 and max = the highest recorded total net asset value. Note that the net assets values are first averaged for every single mutual fund before the mean is calculated.

APPENDIX C. RELEVANT CRITERIA ON WHICH SAMPLE IS EVALUATED

Variable name	Variable description
Index Fund Flag	Identifies if a fund is an index fund: - B = Index-based fund — utilizes indices as its primary filter for the purchase and sale of securities. This is accomplished by investing in the components of one or more indices, or by investing in a small percentage of securities within the index in an attempt to capture the best performers. In each case, the option to invest a portion of assets outside the securities held by the index is left open. - D = Pure Index fund — objective is to match the total investment performance of a publicly recognized securities market index. The fund will hold virtually all securities in the noted index with weightings equal to those in the index. - E = Index fund enhanced — objective is to exceed the total investment performance of a publicly recognized securities market. This is accomplished by investing primarily in derivatives based on the index itself and/or the securities within the index, or by utilizing different weightings for the securities held by the index
CRSP Objective Code	CRSP mapping of Strategic Insights, Wiesenberger, and Lipper objective codes into a continuous series, providing continuity. For a complete description of the objective codes, see http://www.crsp.com/products/documentation/crsp-style-code

Table 17

APPENDIX D. MANUAL CALCULATIONS CONTINGENCY TABLES CANONICAL BENCHMARK

<u>Streak of 1 win or loss.</u>

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 2	
		W	L
Period 1	W	72,008 (66,457)	60,905 (66,457)
	L	60,475 (66,121)	71,767 (66,121)

Expected values in parentheses
$\chi^2 = 1.9*10^3$

Table 18: Contingency table for sequences of two periods, W = Winner, L = Loser in period 2 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1 and 2 starting from the first quarter of 2015 but represent every possible sequence of two following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(72{,}008 - 66{,}457)^2 / 66{,}457 + (60{,}905 - 66{,}457)^2 / 66{,}457 + (60{,}475 - 66{,}121)^2 / 66{,}121 + (71{,}767 - 66{,}121)^2 / 66{,}121 = 1891.7$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $1.8*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to two periods with the canonical benchmark at a 99% confidence interval.

Streak of 2 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 3	
		W	L
Period 2	WW	37,883 (33,849)	29,814 (33,849)
	LL	30,239 (33,490)	36,740 (33,490)

Expected values in parentheses
$\chi^2 = 1.6*10^3$

Table 19: Contingency table for sequences of three periods, W = Winner, L = Loser in period 3, WW = Winner in period 1 and 2, LL = Loser in period 1 and 2 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2 and 3 starting from the first quarter of 2015 but represent every possible sequence of three following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(37{,}883 - 33{,}849)^2 / 33{,}849 + (29{,}814 - 33{,}849)^2 / 33{,}849 + (30{,}239 - 33{,}490)^2 / 33{,}490 + (36{,}740 - 33{,}490)^2 / 33{,}490 = 1{,}592.7$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $1.6*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to three periods with the canonical benchmark at a 99% confidence interval.

Streak of 3 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 4	
		W	L
Period 3	WWW	20,622 (17,885)	15,148 (17,885)
	LLL	12,939 (17,097)	21,255 (17,097)

Expected values in parentheses
$\chi^2 = 2.9*10^3$

Table 20: Contingency table for sequences of four periods, W = Winner, L = Loser in period 4, WWW = Winner in period 1, 2 and 3, LLL = Loser in period 1, 2 and 3 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3 and 4 starting from the first quarter of 2015 but represent every possible sequence of four following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(20{,}622 - 17{,}885)^2 / 17{,}885 + (15{,}148 - 17{,}885)^2 / 17{,}885 + (12{,}939 - 17{,}097)^2 / 17{,}097 + (21{,}255 - 17{,}097)^2 / 17{,}097 = 2{,}860.2$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $2.9*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to four periods with the canonical benchmark at a 99% confidence interval.

Streak of 5 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 6	
		W	L
Period 5	WWWWW	5,335 (4,959)	4,582 (4,959)
	LLLLL	4,047 (5,746)	7,444 (5,746)

Expected values in parentheses
$\chi^2 = 1.1*10^3$

Table 21: Contingency table for sequences of six periods, W = Winner, L = Loser in period 6, WWWWW = Winner in period 1, 2, 3, 4 and 5, LLLLL = Loser in period 1, 2, 3, 4 and 5 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5 and 6 starting from the first quarter of 2015 but represent every possible sequence of six following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(5{,}335 - 4{,}959)^2 / 4{,}959 + (4{,}582 - 4{,}959)^2 / 4{,}959 + (4{,}047 - 5{,}746)^2 / 5{,}746 + (7{,}444 - 5{,}746)^2 / 5{,}746 = 1{,}061.3$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $1.1*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to six periods with the canonical benchmark at a 99% confidence interval.

Streak of 6 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 7	
		W	L
Period 6	WWWWWW	3,012 (2,526)	2,039 (2,526)
	LLLLLL	2,519 (3,545)	4,570 (3,545)

Expected values in parentheses
$\chi^2 = 780.7$

Table 22: Contingency table for sequences of seven periods, W = Winner, L = Loser in period 7, WWWWWW = Winner in period 1, 2, 3, 4, 5 and 6, LLLLLL = Loser in period 1, 2, 3, 4, 5 and 6 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6 and 7 starting from the first quarter of 2015 but represent every possible sequence of seven following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(3{,}012 - 2{,}526)^2 / 2{,}526 + (2{,}039 - 2{,}526)^2 / 2{,}526 + (2{,}519 - 3{,}545)^2 / 3{,}545 + (4{,}570 - 3{,}545)^2 / 3{,}545 = 780.7$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 780.7 > 6.635 causes the H_0 to be rejected confirming performance persistence up to seven periods with the canonical benchmark at a 99% confidence interval.

Streak of 7 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 8	
		W	L
Period 7	WWWWWWW	1,309 (1,418)	1,527 (1,418)
	LLLLLLL	1,766 (2,189)	2,611 (2,189)

Expected values in parentheses
$\chi^2 = 179.9$

Table 23: Contingency table for sequences of eight periods, W = Winner, L = Loser in period 8, WWWWWWW = Winner in period 1, 2, 3, 4, 5, 6 and 7, LLLLLLL = Loser in period 1, 2, 3, 4, 5, 6 and 7 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7 and 8 starting from the first quarter of 2015 but represent every possible sequence of eight following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(1{,}309 - 1{,}418)^2 / 1{,}418 + (1{,}527 - 1{,}418)^2 / 1{,}418 + (1{,}766 - 2{,}189)^2 / 2{,}189 + (2{,}611 - 2{,}189)^2 / 2{,}189 = 179.9$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 179.9 > 6.635 causes the H_0 to be rejected at a 99% confidence interval. However, the outcomes for positive and negative persistence have different implications. For winning streaks, the results suggest performance non-persistence, because the observed count for losers is higher than expected and lower than expected for winners. Concerning, negative streaks, the results suggest performance persistence, because the observed count for losers is higher than expected and lower than expected for winners. Since, contingency tables give generic implications about performance persistence, no conclusions can be drawn regarding performance persistence overall for streaks of seven periods with regard to the following period using the canonical benchmark.

Streak of 8 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 9	
		W	L
Period 8	WWWWWWWW	665 (580)	494 (580)
	LLLLLLLL	993 (1,241)	1,489 (1,241)

Expected values in parentheses
$\chi^2 = 124.3$

Table 24: Contingency table for sequences of nine periods, W = Winner, L = Loser in period 9, WWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7 and 8, LLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7 and 8 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8 and 9 starting from the first quarter of 2015 but represent every possible sequence of nine following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(665 - 580)^2 / 580 + (494 - 580)^2 / 580 + (993 - 1{,}241)^2 / 1{,}241 + (1{,}489 – 1{,}241)^2 / 1{,}241 = 124.3$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $124.3*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence, concerning funds still having a streak after eight periods, for the ninth period with the canonical benchmark at a 99% confidence interval.

Streak of 9 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 10	
		W	L
Period 9	WWWWWWWWW	275 (314)	352 (314)
	LLLLLLLLL	904 (709)	513 (709)

Expected values in parentheses
$\chi^2 = 117.3$

Table 25: Contingency table for sequences of 10 periods, W = Winner, L = Loser in period 10, WWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8 and 9, LLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8 and 9 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 starting from the first quarter of 2015 but represent every possible sequence of 10 following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(275 - 314)^2 / 314 + (352 - 314)^2 / 314 + (904 - 709)^2 / 709 + (513 - 709)^2 / 709 = 117.3$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 117.3 > 6.635 causes the H_0 to be rejected confirming performance non-persistence, concerning funds still having a streak after nine periods, for the 10th period with the corrected benchmark at a 99% confidence interval.

Streak of 10 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 11	
		W	L
Period 10	WWWWWWWWWW	55 (129)	203 (129)
	LLLLLLLLLL	226 (230)	233 (230)
		Expected values in parentheses $\chi^2 = 85.0$	

Table 26: Contingency table for sequences of 11 periods, W = Winner, L = Loser in period 11, WWWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10, LLLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 and χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 starting from the first quarter of 2015 but represent every possible sequence of 11 following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(55 - 129)^2 / 129 + (203 - 129)^2 / 129 + (226 - 230)^2 / 230 + (233 - 230)^2 / 230 = 85.0$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 85.0 > 6.635 causes the H_0 to be rejected at a 99% confidence interval. However, the outcomes for positive and negative persistence have different implications. For winning streaks, the results suggest performance non-persistence, because the observed count for losers is higher than expected and lower than expected for winners. Concerning, negative streaks, the results suggest performance persistence, because the observed count for losers is higher than expected and lower than expected for winners. Since, contingency tables give generic implications about performance persistence, no conclusions can be drawn regarding performance persistence overall for streaks of 10 periods with regard to the following period using the canonical benchmark.

Streak of 11 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 12	
		W	L
Period 11	WWWWWWWWWWW	23 (24)	25 (24)
	LLLLLLLLLLL	91 (100)	109 (100)

Expected values in parentheses
$\chi^2 = 1.7$

Table 27: Contingency table for sequences of 12 periods, W = Winner, L = Loser in period 12, WWWWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, LLLLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 and χ^2 *= the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 starting from the first quarter of 2015 but represent every possible sequence of 12 following periods starting at any point in time.*

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(23 - 24)^2 / 24 + (25 - 24)^2 / 24 + (91 - 100)^2 / 100 + (109 - 100)^2 / 100 = 1.7$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 1.7 < 2.706 < 3.481 < 6.635 causes the H_0 to be not rejected.

Streak of 12 wins or losses.

To calculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the appropriate canonical benchmark of 0.5 to compute the expected frequencies. The adjusted table looks like this:

		Period 13	
		W	L
Period 12	WWWWWWWWWWWW	10 $(9)^{O}$ $(6)^{N}$	7 $(9)^{O}$ $(11)^{N}$
	LLLLLLLLLLLL	75 $(40)^{O}$ $(50)^{N}$	5 $(40)^{O}$ $(30)^{N}$

Expected values in parentheses
$\chi^2 = 61.8$

Table 28: Contingency table for sequences of 13 periods, W = Winner, L = Loser in period 13, WWWWWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, LLLLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 starting from the first quarter of 2015 but represent every possible sequence of 13 following periods starting at any point in time.

The expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5.

The test statistic is:

$$(10 - 9)^2 / 9 + (7 - 9)^2 / 9 + (75 - 40)^2 / 40 + (5 - 40)^2 / 40 = 61.8$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 61.8 > 6.635 causes the H_0 to be rejected at a 99% confidence interval. However, the outcomes for positive and negative persistence have different implications. For winning streaks, the results suggest performance persistence, because the observed count for winners is higher than expected and lower than expected for losers. Concerning, negative streaks, the results suggest performance non-persistence, because the observed count for losers is lower than expected and higher than expected for winners. Since, contingency tables give general implications about performance persistence, no conclusions can be drawn, concerning funds still having a streak after 12 periods, regarding general performance persistence in streaks of 13 periods.

APPENDIX E. MANUAL CALCULATIONS CONTINGENCY TABLES CORRECTED BENCHMARKS

<u>Streak of 1 wins or losses.</u>

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of one is 7/15 which can be found in section 4.3. The adjusted table looks like this:

		Period 2	
		W	L
Period 1	W	72,008 $(66,457)^O$ $(62,026)^N$	60,905 $(66,457)^O$ $(70,887)^N$
	L	60,475 $(66,121)^O$ $(70,529)^N$	71,767 $(66,121)^O$ $(61,713)^N$

New(N) and old(O) expected values in parentheses
$\chi^2 = 6.1*10^3$

Table 29: Contingency table for sequences of two periods, W = Winner, L = Loser in period 2, χ^2 = the Pearson's chi-squared test for independence test statistic and p = corresponding p-value. Important, these four combinations are not only quarter 1 and 2 starting from the first quarter of 2015 but represent every possible sequence of two following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WW, WL, LW and LL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(72,008 - 62,026)^2 / 62,026 + (60,905 - 70,887)^2 / 70,887 + (60,475 - 70,529)^2 / 70,529 + (71,767 - 61,713)^2 / 61,713 = 6,083.2$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $6.1*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to two periods with the corrected benchmark at a 99% confidence interval.

Streak of 2 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of two with regard to the subsequent period is 213/533 which can be found in section 4.3. The adjusted table looks like this:

		Period 3	
		W	L
Period 2	WW	37,883 (33,849)^{O} (27,053)^{N}	29,814 (33,849)^{O} (40,644)^{N}
	LL	30,239 (33,490)^{O} (40,213)^{N}	36,740 (33,490)^{O} (26,766)^{N}

New(N) and old(O) expected values in parentheses
$\chi^2 = 13.4*10^3$

Table 30: Contingency table for sequences of three periods, W = Winner, L = Loser in period 3, WW = Winner in period 1 and 2, LL = Loser in period 1 and 2, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2 and 3 starting from the first quarter of 2015 but represent every possible sequence of three following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWW, WWL, LLW and LLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(37{,}883 - 27{,}053)^2 / 27{,}053 + (29{,}814 - 40{,}644)^2 / 40{,}644 + (30{,}239 - 40{,}213)^2 / 40{,}213 + (36{,}740 - 26{,}766)^2 / 26{,}766 = 13{,}411.8$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $13.4*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to three periods with the corrected benchmark at a 99% confidence interval.

Streak of 3 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of three with regard to the subsequent period is 254/717 which can be found in section 4.3. The adjusted table looks like this:

		Period 4	
		W	L
Period 3	WWW	20,622 $(17,885)^O$ $(12,672)^N$	15,148 $(17,885)^O$ $(23,098)^N$
	LLL	12,939 $(17,097)^O$ $(22,081)^N$	21,255 $(17,097)^O$ $(12,113)^N$

New(N) and old(O) expected values in parentheses
$\chi^2 = 18.4*10^3$

Table 31: Contingency table for sequences of four periods, W = Winner, L = Loser in period 4, WWW = Winner in period 1, 2 and 3, LLL = Loser in period 1, 2 and 3, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3 and 4 starting from the first quarter of 2015 but represent every possible sequence of four following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWW, WWWL, LLLW and LLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(20{,}622 - 12{,}672)^2 / 12{,}672 + (15{,}148 - 23{,}098)^2 / 23{,}098 + (12{,}939 - 22{,}081)^2 / 22{,}081 + (21{,}255 - 12{,}113)^2 / 12{,}113 = 18{,}408.5$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $18.4*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to four periods with the corrected benchmark at a 99% confidence interval.

Streak of 4 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of four with regard to the subsequent period is 64/189 which can be found in section 4.3. The adjusted table looks like this:

		Period 5	
		W	L
Period 4	WWWW	10,536 $(9{,}867)^{O}$ $(6{,}682)^{N}$	9,197 $(9{,}867)^{O}$ $(13{,}051)^{N}$
	LLLL	7,537 $(9{,}945)^{O}$ $(13{,}154)^{N}$	12,352 $(9{,}945)^{O}$ $(6{,}735)^{N}$

New(N) and old(O) expected values in parentheses
$\chi^2 = 10.4*10^3$

Table 32: Contingency table for sequences of five periods, W = Winner, L = Loser in period 5, WWWW = Winner in period 1, 2, 3 and 4, LLLL = Loser in period 1, 2, 3 and 4, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4 and 5 starting from the first quarter of 2015 but represent every possible sequence of five following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWW, WWWWL, LLLLW and LLLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(10{,}536 - 6{,}682)^2 / 6{,}682 + (9{,}197 - 13{,}051)^2 / 13{,}051 + (7{,}537 - 13{,}154)^2 / 13{,}154 + (12{,}352 - 6{,}735)^2 / 6{,}735 = 10{,}444.1$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $10.4*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to five periods with the corrected benchmark at a 99% confidence interval.

Streak of 5 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of five with regard to the subsequent period is 76/227 which can be found in section 4.3. The adjusted table looks like this:

		Period 6	
		W	L
Period 5	WWWWW	5,335 $(4{,}959)^{O}$ $(3{,}320)^{N}$	4,582 $(4{,}959)^{O}$ $(6{,}597)^{N}$
	LLLLL	4,047 $(5{,}746)^{O}$ $(7{,}644)^{N}$	7,444 $(5{,}746)^{O}$ $(3{,}847)^{N}$

New(N) and old(O) expected values in parentheses
$\chi^2 = 6.9*10^3$

Table 33: Contingency table for sequences of six periods, W = Winner, L = Loser in period 6, WWWWW = Winner in period 1, 2, 3, 4 and 5, LLLLL = Loser in period 1, 2, 3, 4 and 5, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5 and 6 starting from the first quarter of 2015 but represent every possible sequence of six following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWWW, WWWWWL, LLLLLW and LLLLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(5{,}335 - 3{,}320)^2 / 3{,}320 + (4{,}582 - 6{,}597)^2 / 6{,}597 + (4{,}047 - 7{,}644)^2 / 7{,}644 + (7{,}444 - 3{,}847)^2 / 3{,}847 = 6{,}894.3$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $6.9*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to six periods with the corrected benchmark at a 99% confidence interval.

Streak of 6 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of six with regard to the subsequent period is 113/337 which can be found in section 4.3. The adjusted table looks like this:

		Period 7	
		W	L
Period 6	WWWWWW	3,012 $(2{,}526)^{O}$ $(1{,}694)^{N}$	2,039 $(2{,}526)^{O}$ $(3{,}357)^{N}$
	LLLLLL	2,519 $(3{,}545)^{O}$ $(4{,}712)^{N}$	4,570 $(3{,}545)^{O}$ $(2{,}377)^{N}$

New(N) and old(O) expected values in parentheses
$\chi^2 = 4.6*10^3$

Table 34: Contingency table for sequences of seven periods, W = Winner, L = Loser in period 7, WWWWWW = Winner in period 1, 2, 3, 4, 5 and 6, LLLLLL = Loser in period 1, 2, 3, 4, 5 and 6, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6 and 7 starting from the first quarter of 2015 but represent every possible sequence of seven following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWWWW, WWWWWWL, LLLLLLW and LLLLLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(3{,}012 - 1{,}694)^2 / 1{,}694 + (2{,}039 - 3{,}357)^2 / 3{,}357 + (2{,}519 - 4{,}712)^2 / 4{,}712 + (4{,}570 - 2{,}377)^2 / 2{,}377 = 4{,}586.8$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $4.6*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to seven periods with the corrected benchmark at a 99% confidence interval.

Streak of 8 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of eight with regard to the subsequent period is 164/481 which can be found in section 4.3. The adjusted table looks like this:

		Period 9	
		W	L
Period 8	WWWWWWWW	665 $(580)^O$ $(395)^N$	494 $(580)^O$ $(764)^N$
	LLLLLLLL	993 $(1,241)^O$ $(1,636)^N$	1,489 $(1,241)^O$ $(846)^N$

New(N) and old(O) expected values in parentheses
$\chi^2 = 1.0*10^3$

Table 35: Contingency table for sequences of nine periods, W = Winner, L = Loser in period 9, WWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7 and 8, LLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7 and 8, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8 and 9 starting from the first quarter of 2015 but represent every possible sequence of nine following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWWWWWW, WWWWWWWWL, LLLLLLLLW and LLLLLLLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(665 - 395)^2 / 395 + (494 - 764)^2 / 764 + (993 - 1,636)^2 / 1,636 + (1,489 - 846)^2 / 846 = 1,021.4$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. $1.0*10^3 > 6.635$ causes the H_0 to be rejected confirming performance persistence up to nine periods with the corrected benchmark at a 99% confidence interval.

Streak of 9 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of nine with regard to the subsequent period is 29/84 which can be found in section 4.3. The adjusted table looks like this:

		Period 10	
		W	L
Period 9	WWWWWWWWW	275 $(314)^O$ $(216)^N$	352 $(314)^O$ $(411)^N$
	LLLLLLLLL	904 $(709)^O$ $(928)^N$	513 $(709)^O$ $(489)^N$

New(N) and old(O) expected values in parentheses
$\chi^2 = 26.4$

Table 36: Contingency table for sequences of 10 periods, W = Winner, L = Loser in period 10, WWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8 and 9, LLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8 and 9, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 starting from the first quarter of 2015 but represent every possible sequence of 10 following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWWWWWWW, WWWWWWWWWL, LLLLLLLLLW and LLLLLLLLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(275 - 216)^2 / 216 + (352 - 411)^2 / 411 + (904 - 928)^2 / 928 + (513 – 489)^2 / 489 = 26.4$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 26.4 > 6.635 causes the H_0 to be rejected confirming performance persistence up to nine periods with the corrected benchmark at a 99% confidence interval.

Streak of 10 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of 10 with regard to the subsequent period is 47/134 which can be found in section 4.3. The adjusted table looks like this:

		Period 11	
		W	L
Period 10	WWWWWWWWWW	55 $(129)^{O}$ $(90)^{N}$	203 $(129)^{O}$ $(168)^{N}$
	LLLLLLLLLL	226 $(230)^{O}$ $(298)^{N}$	233 $(230)^{O}$ $(161)^{N}$

New(N) and old(O) expected values in parentheses
$\chi^2 = 70.5$

Table 37: Contingency table for sequences of 11 periods, W = Winner, L = Loser in period 11, WWWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10, LLLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 starting from the first quarter of 2015 but represent every possible sequence of 11 following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWWWWWWWW, LLLLLLLLLLL, WWWWWWWWWWL and LLLLLLLLLLW are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(55 - 90)^2 / 90 + (203 - 168)^2 / 168 + (226 - 298)^2 / 298 + (233 - 161)^2 / 161 = 70.5$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 70.5 > 6.635 causes the H_0 to be rejected at a 99% confidence interval. However, the outcomes for positive and negative persistence have different implications. For winning streaks, the results suggest performance non-persistence, because the observed count for losers is higher than expected and lower than expected for winners. Concerning, negative streaks, the results suggest performance persistence, because the observed count for losers is higher than expected and lower than expected for winners. Since, contingency tables give generic implications about performance persistence, no conclusions can be drawn regarding performance persistence overall for streaks of 10 periods with regard to the following period using the corrected benchmark.

Streak of 11 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of 11 with regard to the subsequent period is 327/913 which can be found in section 4.3. The adjusted table looks like this:

		Period 12	
		W	L
Period 11	WWWWWWWWWWW	23 $(24)^O$ $(17)^N$	25 $(24)^O$ $(31)^N$
	LLLLLLLLLLL	91 $(100)^O$ $(128)^N$	109 $(100)^O$ $(72)^N$
		New(N) and old(O) expected values in parentheses $\chi^2 = 33.0$	

Table 38: Contingency table for sequences of 12 periods, W = Winner, L = Loser in period 12, WWWWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, LLLLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 starting from the first quarter of 2015 but represent every possible sequence of 12 following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWWWWWWWWW, WWWWWWWWWWWL, LLLLLLLLLLLW and LLLLLLLLLLLL are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(23 - 17)^2 / 17 + (25 - 31)^2 / 31 + (91 - 128)^2 / 128 + (109 – 72)^2 / 72 = 33.0$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 33.0 > 6.635 causes the H_0 to be rejected confirming performance persistence up to nine periods with the corrected benchmark at a 99% confidence interval.

Streak of 12 wins or losses.

To recalculate the contingency table and the Pearson chi-squared test for independence test statistic, the table below is modified with the corrected benchmark. The corrected benchmark used for streaks of 12 with regard to the subsequent period is 59/160 which can be found in section 4.3. The adjusted table looks like this:

		Period 13 W	Period 13 L
Period 12	WWWWWWWWWWWW	10 $(9)^O$ $(6)^N$	7 $(9)^O$ $(11)^N$
	LLLLLLLLLLLL	75 $(40)^O$ $(50)^N$	5 $(40)^O$ $(30)^N$

New(N) and old(O) expected values in parentheses
$\chi^2 = 37.5$

Table 39: Contingency table for sequences of 13 periods, W = Winner, L = Loser in period 13, WWWWWWWWWWWW = Winner in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, LLLLLLLLLLL = Loser in period 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11, χ^2 = the Pearson's chi-squared test for independence test statistic. Important, these four combinations are not only quarter 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 starting from the first quarter of 2015 but represent every possible sequence of 13 following periods starting at any point in time.

The old expected values are calculated by multiplying the total amount of observed frequencies of every row with the canonical benchmark which is 0.5. The new expected values for WWWWWWWWWWWWW, LLLLLLLLLLLL, WWWWWWWWWWWWL and LLLLLLLLLLLW are calculated by multiplying the total observed amount per row with the corrected benchmark and (1 - corrected benchmark).

The test statistic is:

$$(10 - 6)^2 / 6 + (7 - 11)^2 / 11 + (75 - 50)^2 / 50 + (5 - 30)^2 / 30 = 37.5$$

Now that the test statistic is obtained, the answer has to be compared with the threshold value (appendix G) to see if H_0 holds at a 0.1, 0.05 or 0.01 significance level. 37.5 > 6.635 causes the H_0 to be rejected at a 99% confidence interval. However, the outcomes for positive and negative persistence have different implications. For winning streaks, the results suggest performance persistence, because the observed count for losers is lower than expected and higher than expected for winners. Concerning, negative streaks, the results suggest performance non-persistence, because the observed count for losers is lower than expected and higher than expected for winners. Since, contingency tables give generic implications about performance persistence, no conclusions can be drawn regarding performance persistence overall for streaks of 12 periods with regard to the following period using the corrected benchmark.