Essays on Risk Management & Investor Irrationality – Analyses of Stocks, Bonds, and Real Estate

DISSERTATION

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

(Dr. rer. pol.)

eingereicht an der

Fakultät für Wirtschaftswissenschaften

der Universität Regensburg

von

Carsten Fritz

Tag der Disputation

27.09.2021

Berichterstatter

Prof. Dr. Steffen Sebastian

Prof. Dr. Klaus Röder

Table of Contents

List of Tables

List of Figures

1 Preface

1.1 Risk and Irrationality

Financial markets are as complex as ever due to an accelerating development in the last decades. This development is attended by the presence of risk since risk-taking is the natural economic necessity to generate excess returns. Although the exact meaning of risk depends on the context, the term generally refers to the possibility of a loss or other undesirable outcome, not to be confused with uncertainty (Knight, 1921). Accompanying risk, the irrational behavior of the average investor is another factor that influences the return. Systematic violations of the efficient market hypothesis in the 1980's led to the emergence of behavioral finance attributing anomalies in asset returns to investor irrationality (e.g. Kahneman and Tversky, 1979; Mehra and Prescott, 1985; Black, 1986; Kahneman and Riepe, 1998).

Traditional finance defines a rational investor as one who makes choices that attempt to maximize his utility, or return on investment for a given level of risk. Such an investor may not be biased by emotions like fear, greed and anxiety, behavioral patterns, or external factors which are not directly affecting his stated utility (Statman, 2011). Traditional finance theory and research, as well as common financial models, still treat investors and their investment decisions as if they are not affected by these biases (e.g. Markowitz, 1952; Fama, 1970; Ross, 1976). Unfortunately, these approaches may be a detriment to investors, as evidenced by poor returns historically achieved by the average investor compared to returns that were possible to achieve with a simple passive investment bearing the same risk (Dalbar, 2010). Although researchers like Ross (2005) strongly argue that concepts from behavioral finance do not provide any benefit compared to traditional finance theory, neoclassical finance and behavioral finance do not have to be mutually exclusive as shown by Barberis et al. (2001).

Typically, investors support upside risk rather than downside risk which makes taking care of appropriate risk assessment, measurement and management even more important. These tasks are further complicated by the dynamic nature of risk. According to Gervais, Kaniel, and Mingelgrin (2001), investors like to relate to the stock market as a human which suffers or benefits from different moods like being bad-tempered or high-spirited. There are days where it can overreact or days where it can just behave normal. Consensus in the literature on the average investor and his investment experience suggest that humans are not wired for disciplined investing since many investors follow their emotions whilst markets have rewarded discipline and long-term investing (e.g. Kahneman and Tversky, 1979; Grossman and Stiglitz, 1980; Malkiel, 1995; Hirshleifer and Shumway, 2003). Unfortunately, peoples' (cognitive) abilities are limited which forces the majority of them to adopt rules of thumb to guide them through their decision-making. Baker and Wurgler (2007) as well as Seiler and Lane (2015) provide clear evidence that the (real estate) investment decision is not based solely on economic fundamentals, but also on emotions and personal beliefs. It is not only stock investors that underperform or show poor timing performance. Also, those who invest in a diversified portfolio of, for example, stocks, bonds, and real estate do not keep up with respective benchmarks namely indexes of similar securities. As argued by Cochrane (2005), the discussion about financial markets being efficient including the irrational behavior of individuals, is a question of whether asset pricing theory describes the ways the world does work or rather the ways it should actually work. However, it is evident that all those biases mentioned directly affect the trading behavior and performance of investors.

First of all, identifying those risk factors that the average investor has to face along his investment experience is important to become aware of potential pitfalls. Investors face situations in which the future financial performance of their assets is uncertain. Risk factors influencing this future performance can be classified as either systematic or non-systematic. As a result, Bhansali (2011) concludes, that investors should diversify their exposure across risk factors and not across asset classes (see also Asl and Etula, 2012; Bender et al. 2014). Thus, information on the type of risk factor that may influence the cash flow and quantifying the actual impact including potential correlations is materially useful to investors. Knowing if and how risk can be mitigated, for example via diversification, helps the investor to take appropriate action concerning the investment. Another option tackling risk is to hedge against the occurring downside. One of such risk factors, especially considering (direct) real estate, is inflation or respectively deflation. Thus, the topic of deflation and hedging capabilities of different asset classes against deflation as well as its influence on asset values are part of this thesis.

In order to help the investor in his decision making and to preserve him from irrationality driven actions, risk as a whole or risk factors in particular have to be modeled properly (French and Gabrielli, 2004). There are deterministic (Mollart, 1988) as well as stochastic models. The latter allow the explicit modelling of risk using the corresponding distribution functions of all risk factors deemed necessary (e.g. Hoesli et al. 2006). Since risk modelling involves empirically sophisticated models as well as an extensive history of risk factors and return, it is mostly conducted by institutional investors or regulatory authorities. Stochastic modelling enables the deduction of risk measures allowing the investor to assess his own decision in the context of other investments. Additionally, the investor is able to deduce a certain amount of risk capital needed as a cushion against adverse market movements and to monitor and forecast the risk and return ratio he is accepting in order to derive an efficient portfolio allocation and to be less prone to his own emotions and other potential biases.

Other than showing ways to circumvent several aspects of irrationality affecting investment strategies, it is also important to reveal to the average investor the magnitude of performance loss actually realized compared with a simple buy & hold or passive investment strategy. Of even greater importance to illustrate is the time when the performance is lost as well as the drivers of potential out- or underperformance. For example, Frazzini and Lamont (2008) provide evidence suggesting that the average investor tends to invest at the wrong time, buying too late into a rally and selling too late when prices decline. Not only retail investors suffer from this behavioral pattern due to potential unsophistication, but also institutional investors show poor timing performance. Thus, being aware of the timing of cash flows into and out of a security as well as the resulting consequences may help investors to be aware of their own irrational behavior when investing.

Besides classic asset classes like stocks and bonds, real estate has evolved into an integral part of the asset allocation of institutional and private investors. Advantages of (direct) real estate as an investment include its assumed inflation hedging potential, relatively stable rental income and low correlation with other asset classes (e.g. Maurer and Sebastian, 2002; Adrangi et al. 2004). There are also disadvantages that have to be faced regarding direct real estate investment. High transaction costs and limited liquidity combined with the disability to quickly sell the investment without having to face price discounts are amongst the main trade-offs (Hoesli and Lekander, 2008). Changes in consumer prices, and therefore inflation or deflation, are economic factors that investors must take into consideration when planning and managing their portfolios in order to mitigate the risk they are accepting. Thus, the first article of the thesis, "*Real Estate, Stocks, and Bonds as a Deflation Hedge*", analyzes the deflation-hedging capabilities of (direct) real estate investments alongside stocks and bonds for markets that suffered from long periods of deflation, namely Japan and Hong Kong. Extending the classic Fama and Schwert (1977) approach by implementing an ARIMA model framework enables to see if the nominal returns (e.g. real estate prices or rents, stock returns or bond returns) possess the ability to vary in one-to-one correspondence with changes of consumer prices and therefore provide a perfect hedge against inflation and accordingly deflation. If they do so, expected real return on the asset is uncorrelated with the changes in consumer prices.

In order to overcome the issues inherited by direct real estate investment, investors can also pursue an indirect form of investment like real estate investment trusts (REITs). Nevertheless, with these indirect investments, the investor has to face another set of risks and disadvantages, as well. Hence, the second article, "*AR-GARCH-EVT-Copula for Securitized Real Estate: An approach to improving risk forecasts*" is the first study to apply the so-called AR-GARCH-EVT-Copula model to bivariate portfolios, which contain securitized real estate in addition to the asset classes of stocks and bonds. This article analyses and compares the procedures of forecasting price risk metrics. The primary motivation for the paper is the stylized facts about financial market data. Due to the typical problems of financial time-series like heteroskedasticity, autocorrelation, leptokurtosis, skewness, fat tails and volatility clustering (McNeil and Frey, 2000; Liow, 2008), the correct forecast of price risk is subject to a lot of potential (econometric) pitfalls.

Asymmetric AR-GARCH models are needed to circumvent the issues of autocorrelation, heteroskedasticity and skewness. The remaining stylized facts can be modeled via extreme value theorem considering the peaks-over-threshold approach (Rossignolo et al. 2012). In addition, dynamic and asymmetric dependency appears to be necessary considering the modelling since real estate is an asset class, which co-moves to stocks and bonds in timely variant, skewed, and over-proportional fashion. Once, the correct model has been adopted, the resulting estimates can be compared to the most common approaches like variance-covariance and historical simulation via back-testing. These tests as well as the optical inspection show that such a complex model yields improved results considering the forecast of price risk and thus should be adopted by all (institutional) investors and regulatory authorities.

Investors are not only prone to risk and therefore to diminishing returns when it comes to forecasting. Especially evaluations of mutual fund performance have been a subject of interest since the introduction of financial services. A large body of literature has analyzed the empirical findings of investors chasing past returns and tending to time the market (e.g. Frazzini and Lamont, 2008, Greenwood and Schleifer, 2014; Chien, 2014). Nevertheless, market timing strategies are not doomed to automatically fail. Cochrane (2011) shows that if returns are somewhat predictable, then investors may achieve higher Sharpe ratios with a good timing-strategy. Whether investors can enhance their returns by selecting well-performing funds or timing their cash flows to the fund is still an ongoing debate with results covering all sides. Authors like Gruber (1996) and Zheng (1999) find evidence for a "smart money" effect whereas Frazzini and Lamont (2008) show that there is a "dumb money" effect, as well. Yet, the influence of investor timing decisions on their actual returns is less clear. Thus, comparing simple buy & hold investment strategies with money-weighted returns the average investor receives, is a way to show if the investor is able to realize a timing-outperformance or if he should rather think about his investment behavior and flee into passive strategies. As always, this is not a unilateral and simple discussion. Thus, the last article of this thesis, "The *Performance Gap: When is average Investor Performance poor?*", illustrates when the performance of the average investor was good or poor and derives the drivers of the respective over- or underperformance.

Based on the entire aforementioned derived research, the following questions are central for the empirical studies of the thesis:

Paper 1: Real Estate, Stocks, and Bonds as a Deflation Hedge

I. Are consumer prices and real estate prices as well as real estate rents correlated?

II. Do real estate, stocks, and bonds provide hedging capabilities against expected and unexpected consumer price changes?

III. Does real estate provide a hedge against deflation in particular?

IV. How is the hedging capability behaving for different sub-classes of real estate?

Paper 2: AR-GARCH-EVT-Copula for Securitized Real Estate: An Approach to improving Risk Forecasts?

I. Does real estate as an asset co-move to stocks and bonds in timely variant, skewed, and over-proportional fashion?

II. Does the AR-GARCH-EVT-Copula approach provide more accurate price risk metric forecasts compared to the variance-covariance or historical simulation method?

III. Are there any visible patterns concerning the chosen copula for the various asset classes?

Paper 3: The Performance Gap: When is average Investor Performance poor?

I. When does the average investor show good or poor timing performance?

II. What are the drivers of the performance gap?

III. Which measure of money-weighted returns proposes a more realistic assumption: Internal rate of return or modified internal rate of return?

The thesis is structured as follows: The next chapters reproduce the empirical studies that are related to the abovementioned research questions. Every article is introduced by a page that states the full list of authors, the status of the article, and a short abstract. The last chapter contains the conclusion stating a summary of the articles, the definite answer of the derived hypotheses, the joint conclusions of the thesis, the research limitations, and potential future research possibilities. Some tables and figures are relegated to appendices at the end of the corresponding chapter. The notation is consistent within each chapter but can differ slightly across chapters.

1.2 Bibliography

Adrangi, B., Chatrath, A., & Kambiz, R. (2004). REIT Investments and Hedging Against Inflation. Journal of Real Estate Portfolio Management, 10(2), pp. 97-112.

Asl, F. M., & Etula, E. (2012). Advancing Strategic Asset Allocation in a Multi-Factor World. The Journal of Portfolio Management, 39(1), pp. 59-66.

Barberis, N., Ming, H., & Santos, T. (2001). Prospect Theory and Asset Prices. The Quarterly Journal of Economics, 116, pp. 1-53.

Bender, J., Hammond, B. P., & Mok, W. (2014). Can Alpha Be Captured by Risk Premia? The Journal of Portfolio Management, 40(2), pp. 18-29.

Bhansali, V. (2011). Beyond Risk Parity. The Journal of Investing, 20(1), pp. 137-147.

Black, F. (1986). Noise. The Journal of Finance, 41(3), pp. 528-543.

Chien, Y. L. (2014). The Cost of chasing Returns. Economic Synopses Federal Reserve Bank of St. Louis.

Chochrane, J. H. (2005). Asset Pricing. Princeton: Princeton University Press.

Cochrane, J. H. (2011). Presidential Address: Discount Rates. The Journal of Finance, 66(4), pp. 1047-1108.

Dalbar Inc. (2010). Quantitative Analysis of Investor Behavior. Marlborough.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), pp. 383-417.

Fama, E. F., & Schwert, W. G. (1977). Asset Returns and Inflation. Journal of Financial Economics, 5(2), pp. 115-146.

French, N., & Gabrielli, L. (2004). The Uncertainty of Valuation. Journal of Property Investment & Finance, 22(6), pp. 484-500.

Gervais, S., Kaniel, R., & Mingelgrin, D. H. (2001). The High-Volume Return Premium. The Journal of Finance, 56(3), pp. 877-919.

Greenwood, R., & Shleifer, A. (2014). Expectations of Returns and Expected Returns. Review of Financial Studies, 27(3), pp. 714-746.

Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. The American Economic Review, 70(3), pp. 393-408.

Gruber, M. J. (1996). Another Puzzle: The Growth in Actively Managed Mutual Funds. The Journal of Finance, 51(3), pp. 783-810.

Hirshleifer, D., & Shumway, T. (2003). Good Day Sunshine: Stock Returns and the Weather. The Journal of Finance, 58(3), pp. 1009-1032.

Hoesli, M., & Lekander, J. (2008). Real Estate Portfolio Strategy and Product Innovation in Europe. Journal of Property Investment and Finance, 2, pp. 162-176.

Kahneman, D., & Riepe, M. W. (n.d.). Aspects of Investor Psychology. The Journal of Portfolio Management, 24(4), pp. 52-65.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), pp. 263-291.

Kinght, F. H. (1921). Risk, Uncertainty and Profit. Boston: Mifflin.

Liow, K. H. (2008). Extreme Returns and Value at Risk in International Securitized Real Estate Markets. Journal of Property Investment and Finance, 26(5), pp. 418-446.

Malkiel, B. G. (1995). Returns from Investing in Equity Mutual Funds 1971 to 1991. The Journal of Finance, 50(2), pp. 549-572.

Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), pp. 77-91.

Maurer, R., & Sebastian, S. (2002). Inflation Risk Analysis of European Real Estate Securities. Journal of Real Estate Research, 24(1), pp. 47-77.

McNeil, A. J., & Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. Journal of Empirical Finance, 7(3-4), pp. 271-300.

Mehra, R., & Prescott, E. (1985). The Equity Premium: A Puzzle. Journal of Monetary Economics, 15(2), pp. 145-161.

Mollart, R. G. (1988). Monte Carlo simulations using Lotus 123. Journal of Property Valuation, 6(4), pp. 419-433.

Ross, S. (1976). The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory, 13(3), pp. 341-360.

Ross, S. A. (2005). Neoclassical Finance. Princeton: Princeton University Press.

Rossignolo, A. F., Fethi, M. D., & Shaban, M. (2012). Value-at-Risk Models and Basel Capital Charges. Journal of Financial Stability, 8(4), pp. 303-319.

Seiler, M. J., Seiler, V. L., & Lane, M. A. (2012). Mental Accounting and False Reference Points in Real Estate Investment Decision Making. Journal of Behavioral Finance, 13(1), pp. 17-26.

Statman, M. (2011). What Investors Really Want. New York: McGraw-Hill Education.

Wurgler, J., & Baker, M. (2007). Investor Sentiment in the Stock Market. Journal of Economic Perspectives, 21(2), pp. 129-152.

Zheng, L. (1999). Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability. The Journal of Finance, 54(3), pp. 901-933.

2 Real Estate, Stocks, and Bonds as a Deflation Hedge

Authors: Carsten Fritz, Benedikt Fleischmann, Steffen Sebastian

Submission to: The Journal of International Real Estate Review (IRER)

Status: published in 2019

Fritz, C., Fleischmann B., Sebastian, S. (2019). Real Estate, Stocks, and Bonds as a Deflation Hedge. International Real Estate Review, 22(1), pp. 1-26.

Abstract

With inflation rates remaining close to zero in all major developed economies for long periods, especially from 1998 - 2015, investors have become increasingly concerned about the potential effects of deflation on asset value. Negative inflation rates were observed between 1998 and 2009 in Hong Kong and Japan, and those economies faced several years of deflation. There is a rich body of literature on the effects of inflation hedging on the returns of stocks, bonds, and real estate. We examine asset returns for these products between 1986 and 2009, and use an ARIMA model to explore whether they offer a deflation hedge. We show that rents and real estate prices are closely linked to consumer prices, which confirms prior findings about inflation hedging. Because the relationship was generally positive and over proportional, we found

that real estate was not an effective hedge against deflation. In contrast, we found no relationships between stocks or bonds and inflation. Only for Japanese bonds were we able to find a significantly negative relationship with unexpected deflation.

2.1 Introduction

Continual low interest rates and the slow recovery of all major economies from the 2008 - 2009 recession have resurrected investors' fears of deflation. One particularly chilling possibility is of falling into a state where inflation turns negative, and we endure a Japan-style outcome. Direct real estate investment is often regarded as a hedge against inflation, while bond investments are typically associated with exposure to inflation risk. The cash flows and repayment of bond capital are fixed ex ante. Thus, any inflationary losses are fully captured in the purchasing power. Income sources from real estate investments, such as rent, however, are subject to renegotiation and renewal at regular intervals. Therefore, they are linked to a decrease in monetary value. Following this reasoning, real estate should decrease its nominal value under deflation (negative inflation rates), and bonds should protect against an increase in monetary value, since a negative interest rate is uneconomical. On the other hand, a deflationary environment is often accompanied by an economic crisis. Therefore, normal market relationships may be weak.

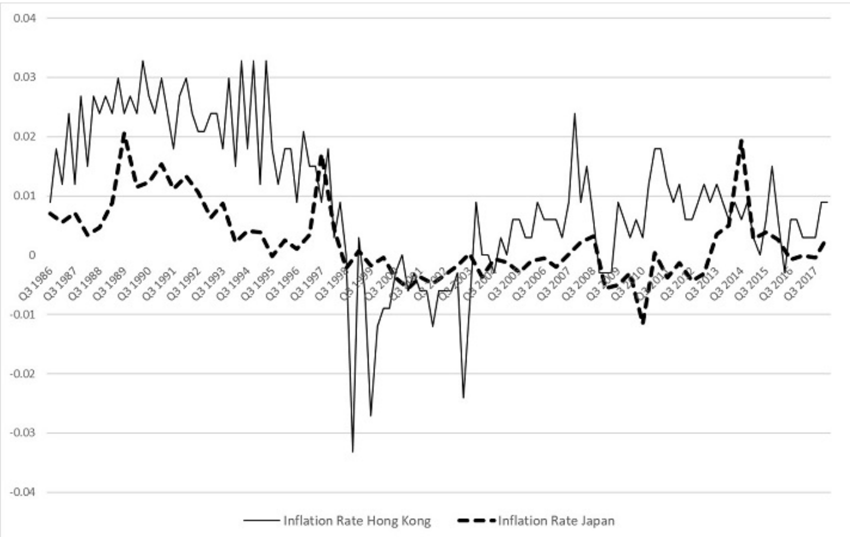
To further elaborate on what we mean by inflation and deflation protection, consider the following. The purchasing power of assets or protection against inflation is usually defined so that the nominal value of an investment increases in proportion to inflation. Thus, real value remains unchanged. In the context of an econometric investigation, we can define a perfect hedge against inflation or deflation as: The coefficients of the regression of the expected or unexpected inflation or deflation on the nominal return on an investment, considering the Fama and Schwert (1977) model, are not statistically or discernibly different from one. In other words, nominal returns increase or decrease, just as changes in inflation and deflation can be compensated for.

In an inflationary environment with rising prices, it is relevant to retain the purchasing power of the investment. This is not the case in a deflationary environment. Protection against deflation means that the nominal value remains constant, so the real value increases. From the viewpoint of an owner-occupier without leverage, deflation protection is only a protection against an illusory value loss. From the viewpoint of any other investor, e.g., an institutional investor, however, deflation protection is much more relevant. This is especially true when obligations/liabilities are fixed in nominal terms, and must be serviced by the returns on assets. Moreover, under most regulations, losses in nominal terms will affect the results in the annual report of the respective company.

Note that real estate assets, which are a good inflation hedge, may be especially prone to deflation risk. The best hedge against deflation would then be an investment in nominally denominated assets such as bonds. The aim of our paper is twofold: First, we analyze whether the deflation risk of real estate is symmetric to the inflation hedge characteristics.

Second, we examine whether stocks or bonds provide a hedge against deflation. We choose two markets that have experienced longer periods of deflationary regimes, Japan and Hong Kong. Both countries exhibited a measurable decrease in consumer prices over the 1998 - 2010 period. Hong Kong's inflation rates have generally been more positive since 2005; in Japan, monetary value remained very stable, around 0%, until the end of 2013. Figure 1 illustrates the inflation rates for both countries for our research period, and shows that both experienced overall deflationary phases beginning in 1998.

Figure 1: Inflation rates for Japan and Hong Kong



Notes: The inflation rate for Japan derives from half-yearly values compared to quarterly observations for Hong Kong. The issue of data availability will be discussed later.

The remainder of this chapter is organized as follows. Section 2.2 reviews the related literature, while section 2.3 provides a discussion of the Fama and Schwert (1977) framework, the extension of their model, and our data. Section 2.4 analyzes the data for the two given markets with respect to their inflationary and deflationary periods. Finally, the main findings are summarized and interpreted in section 2.5.

2.2 Literature Review

Early studies by Bodie (1976), Jaffe and Mandelker (1976) as well as Fama and Schwert (1977) found that nominal stock returns were negatively related to actual inflation. This relationship held even when the expected and unexpected components of inflation were examined separately. Although the Fama-Schwert (1977) model has been criticized for a lack of distinction between long-term equilibrium adjustments and short-term dynamic movements, it has been applied in numerous papers. In more recent studies, however, this classic model is usually supplemented by models, such as cointegration tests, that capture long-run relations.

Gultekin (1983) shows that results based solely on the relationship between stock returns and inflation can also be justified for many other countries. The finding that stocks provide a negative hedge against inflation seems counterintuitive at first, given that shares represent claims on cash flows from real assets. Among the various investment categories examined by Fama and Schwert (1977) for the U.S. (short- and long-term government bonds, residential real estate, stocks, human capital), residential real estate offered the only complete hedge against inflation. In contrast, short- and long-term government bonds provided protection only against expected, but not unexpected, inflation. Studies on the characteristics of commercial real estate have found it has at least a partial hedging capability. While American commercial real estate seems to provide a hedge against expected inflation, there is no clear evidence with regard to unexpected inflation (see, for example, Brueggeman et al. (1984), Hartzell et al. (1987), Gyourko and Linneman (1988), Rubens et al. (1989)). Additionally, Gyourko and Linneman (1988) attest the effective hedging capabilities of U.S. REITs, at least against expected inflation, for the 1972 - 1986 period. According to Park et al. (1990), REIT investments provide negative inflation hedging against both expected and unexpected inflation for the same period. In contrast to Gyourko and Linneman (1988), the rate of inflation is not predicted by means of an ARMA model, but rather by using three-month interest rates.

In comparison to direct real estate investments, the correlation of real estate stocks (REITs) to expected and unexpected inflation was determined to be zero or negative. The studies cited above use monthly to annual returns in their investigations. Boudoukh and Richardson (1993) examined long-term dependencies for the U.S. and U.K. markets for one- and five-year returns. They found that annual returns on stocks are negatively correlated with inflation, while five-year returns are positively correlated.

Yobaccio et al. (1995) extended Fama and Schwert's (1977) approach with a market parameter in order to examine the returns of different types of U.S. REITs for the 1972-1992 period. They find that REITs provide some hedging capabilities against anticipated, but not unanticipated, inflation. Liu et al. (1997) investigate the inflation hedging characteristics of property trusts in seven countries for 1980 - 1991. They find that U.S. property trusts, similarly to common stocks, are rather perverse inflation hedges. However, they found no evidence that real estate securities in other countries provided any better hedging capability against inflation than common stocks. In fact, in some countries, property trusts can be even more perverse as hedging instruments than common shares. In addition to short-term capabilities, Adrangi et al. (2004) also examine the long-term hedging characteristics of REITs by means of cointegration tests. They find no evidence that REITs significantly protect against inflation in the long run. Moreover, Maurer and Sebastian (2002) focus on the hedging properties of real estate securities in France, Germany, Switzerland, and the U.K. from 1980-2000. They find that only German open-end real estate funds, and not real estate stocks in Germany or other countries, provide hedging capabilities against anticipated inflation. Their approach follows an ARIMA time series model, with ex post inflation rates used as inflation predictors. They also determine shortfall risk measures for real returns of real estate stocks and German real estate funds. More recently, Obereiner and Kurzrock (2012) analyze German open-end real estate funds, special funds, and real estate stocks in order to test their hedging effectiveness against inflation. In addition to using the Fama and Schwert (1977) approach, they conducted cointegration and Granger causality tests, and found that real estate yields in the short run are almost independent of inflation. On the other hand, the cointegration tests demonstrated that real estate investments are long-term inflation hedges. The causality tests also indicated that real estate returns are influenced by inflation over the long run.

With a focus on the Hong Kong market, Ganesan and Chiang (1998) and Glascock et al. (2008) found real estate was less effective as an inflation hedge for the 1984-1994 and 1998-2006 periods. Ganesan and Chiang (1998) implemented cointegration methods, as well as the basic Fama and Schwert (1977) approach with quarterly data. They conclude that financial assets would have provided a better hedge against inflation in Hong Kong. Real assets, such as real estate, were of no use as a hedge during inflationary phases. Glascock et al. (2008) use short- and long-term methods and Granger causality tests, and conclude that real estate is not an effective hedge against inflation. They also construct subsamples for different types of properties, which show various inflation hedging characteristics.

During times of deflation, bonds are viewed as a typical hedge. Both bonds and equities have high real returns in a deflationary environment.

Several studies have shown that bonds can outpace stocks in terms of real returns during severe deflation (see for example, Dimson et al. (2012)). Hence, the real rate of return depends on the inflation rate, and seems to share a negative correlation coefficient. The correlation depends on the country and its specific macroeconomic factors, as well as on its central bank policy. TIPS (Treasury inflation-protected securities) are also used to hedge against deflation. According to Fleckenstein et al. (2014), the principal of a TIPS issue is protected against deflation because the amount received by a TIPS holder at maturity cannot be less than par. Hence, they feature a deflation floor.

However, bonds are not the only safe hedge against deflation. In an environment of falling prices, investors flee to perceived safe havens, such as gold. According to Day (2015), gold can appreciate during inflationary or deflationary phases. We further note that real estate may also be considered a safe haven. However, borrowing to hedge with real estate is a riskier investment during deflationary phases, because it makes repayment of debts more difficult.

Due to the lack of long-term deflationary phases in most OECD countries in recent years, few articles have explored such time intervals in a focused manner. To the best of our knowledge, a solid study covering the development of real estate prices and rents within deflationary market phases is lacking. Therefore, our study contributes to the literature by examining the suitability of real estate as a deflation hedge.

2.3 Methodology

Fama and Schwert's (1977) approach for quantifying the characteristics of inflation hedges has been widely applied in the literature to a host of investment categories in different countries. The model has been used to gauge the degree to which an investment's nominal returns depend on expected and unexpected changes in consumer prices. Because this approach only examines short-term dependencies, not long-term correlations, it has been modified a number of times. Moreover, even the more sophisticated versions have not yielded substantially different results to date. Hence, the international empirical literature still basically relies on the Fama and Schwert (1977) model. Therefore, we will also use it for our analysis here. We will extend the basic approach by means of an ARIMA model. A brief summary of the method follows.

According to Fisher (1930), the nominal interest rate on an investment can be divided into the real interest rate and price level changes. In reality, the problem of uncertain price level changes arises. Therefore, the nominal rate of interest I_{nom} can be written as the equilibrium of expected real interest rate $E(I^{real})$ plus the expected inflation rate $E(\pi)$ under uncertainty and imperfect foresight. Fama and Schwert extended Fisher's hypothesis to (nominal) risk-bearing investments:

$$E\left(R_{i,t}^{n}|\phi_{t-1}\right) = E\left(R_{i,t}^{r}|\phi_{t-1}\right) + E(\pi_t|\phi_{t-1}) \qquad (1)$$

where:

$E\left(R_{i,t}^{n}|\phi_{t-1}\right)$ = the anticipated nominal investment return for period t-1 to t, given information in t-1.

$E\left(R_{i,t}^{r}|\phi_{t-1}\right)$ = the anticipated real return on the investment for period t-1 to t, given information in t-1.

$E(\pi_t|\phi_{t-1})$ = the expected change in consumer prices for period t-1 to t, given information in t-1.

For Fisher's hypothesis to be valid, the anticipated real interest rate and the expected rate of inflation π_t^e, therefore the real and the monetary sectors, must be independent quantities. Under these conditions, any investment's hedge effectiveness can be examined using the following empirical regression model:

$$R_{i,t}^{n} = \alpha_i + \beta_i \pi_t^e + \gamma_i[\pi_t - \pi_t^e] + \epsilon_{i,t} \qquad (2)$$

where:

$R_{i,t}^{n}$ = the nominal return on the i^{th} asset for period t-1 to t.

π_t^e = the rate of inflation expected for period t-1 to t.

π_t = the realized rate of inflation for period t-1 to t.

$\pi_t - \pi_t^e$ = unexpected inflation for period t-1 to t.

$\epsilon_{i,t}$ = error term, $\epsilon_t \sim WN(0, \sigma^2)$.

The parameters α_i, β_i and γ_i must be estimated individually for each asset. The regression parameter β_i describes the hedge effectiveness of the ith investment with respect to the expected change in consumer prices. According to Fama and Schwert (1977), an asset is considered a perfect or complete hedge against expected inflation when β_i = 1, while an investment is regarded as a negative hedge against expected inflation if β_i < 0. A short position would then be an inflation hedge. The regression model's second predictive variable provides additional information about the nominal asset return's sensitivity to unexpected changes in consumer prices. If β_i as well as γ_i are not significantly different from 1, thus there is a complete hedge, then ex-post real returns and consumer price changes are uncorrelated. Table 1 summarizes the abovementioned dependence of hedge characteristics on the regression coefficient's value and on the direction of changes in consumer prices.

Table 1: Classification of inflation and deflation hedges

Value of β, γ coefficient	]−∞; 0[	0	]0; 1[	1	]1; 2[	2	]2;∞[
Hedge - Classification	negative hedge	no hedge	positive hedge				
			weak hedge	perfect hedge	overhedge		
Risk participation under inflation ($\Delta\pi = 1$)	over $\Delta R < -1$	complete $\Delta R = -1$	partial $-1 < \Delta R < 0$	none $\Delta R = 0$	over (risk turns to reward) $\Delta R > 0$	$\Delta R = 1$	$\Delta R > 1$
Chance Reward participation under deflation ($\Delta\pi = -1$)	over $\Delta R > 1$	complete $\Delta R = 1$	partial $0 < \Delta R < 1$	none $\Delta R = 0$	over (reward turns to risk) $-1 < \Delta R < 0$	$\Delta R = -1$	$\Delta R < -1$

Notes: The classification of the investment as either an inflation or a deflation hedge, and participation in risk from changes in real return ΔR or in reward from changes in the inflation rate Δπ, are shown to be dependent on the regression values. We account for the coefficients β, γ, where the nominal return is an endogenous variable, and inflation rates are exogenous.

Next, we test the estimated parameters against two hypotheses:

1.) $H_0\colon \beta_i = 0$ vs. $H_1\colon \beta_i \neq 0$

If the null hypothesis can be rejected statistically, it will indicate that the i^{th} investment, is either a positive or negative hedge against expected inflation, depending on the estimated parameter's sign (see Table 1). The second test evaluates the influence exerted by unexpected changes in consumer prices:

2.) $H_0: \gamma_i = 0$ vs. $H_1: \gamma_i \neq 0$

If the null hypothesis can be rejected statistically, it will indicate that the i^th investment is either a positive or a negative hedge against unexpected inflation, depending on the estimated parameter's sign (see Table 1).

The significant autocorrelation of the returns themselves, as well as of the residuals produced by the above regression equation, call for an extension of the Fama and Schwert (1977) approach with an ARMA model. We posit that the integration of past returns into this model is justified on economic grounds. Real estate properties normally cannot be traded as quickly as stocks, since they incur high transaction costs. Their market is also not transparent, so new information is absorbed into prices slowly. Therefore, we extend the previous Fama and Schwert (1977) model as follows:

$$R_{i,t}^n = \alpha_i + \beta_i \pi_t^e + \gamma_i [\pi_t - \pi_t^e] + \sum_{j=1}^{p} a_j R_{i,t-j}^n + \sum_{k=1}^{q} b_k \epsilon_{i,t-k} + \epsilon_{i,t} \quad (3)$$

Here the most recent asset returns p and residuals q have been factored into the regression. To keep parameterizing to a minimum, the correct number of p and q terms is calculated using the Schwarz Information Criterion (Bayesian Information Criterion, BIC), and the corrected coefficient of determination.

The capital markets model presented above is based largely on the expected inflation rate. Because it cannot be observed, this variable must be determined by other means. Models grounded in macroeconomics or univariate time-series models are the most commonly used in the literature (Fama and Gibbons (1982), Gultekin (1983), Hartzell et al. (1987), Limmack and Ward (1988) and Harvey (1989)). Fama and Schwert (1977) estimate expected inflation from a three-month Treasury bill's interest rate with a one-period lag. Assuming that the country's creditworthiness/likelihood of default remains unchanged, the real interest on a Treasury bill should remain constant over time. Because the nominal single-period interest rate is known ex ante, the expected future rate of inflation π^e can be obtained directly. It corresponds to the changes in the nominal interest rate, given the constant real interest rate.

$$\pi_t^e = R_{TBill,t}^n - \left(R_{TBill,t-1}^n - \pi_{t-1}\right) = R_{TBill,t}^n - R_{TBill,t-1}^r \qquad (4)$$

Fama (1975) was able to confirm this hypothetical constant real interest rate by studying American government bonds for the 1953-1971 period.

For univariate time-series models, the empirically observable, realized inflation rate is used, with the underlying stochastic process approximated for with an ARMA model.

$$\pi_t^e = c + \sum_{j=1}^{p} a_j \pi_{t-j}^n + \sum_{k=1}^{q} b_k \epsilon_{t-k} \qquad (5)$$

The inflation expectation synthesized at the end of period t-1 for period t is calculated here as the weighted mean of the past realized inflation rates p and the past disturbance terms q. This model implies that economic actors only use past changes in price levels to formulate their expectations of future price levels. Lizieri et al. (2008) compared various models for the U.S. and U.K. markets. Using an error correction mechanism, they concluded that there is little evidence of short-term adjustments to changes in either anticipated or unexpected inflation. In the long run, public market asset returns are linked to anticipated inflation. Therefore, adjustments to changes in inflation occur through an error-correcting adjustment process to the long-run relationship.

2.4 Data and Descriptive Statistics

Our data for Japan derives from usage-based indices for commercial and industrial properties and for residential real estate that are published annually in March and October by the Japan Real Estate Institute. The NIKKEI 225 Stock Average Price Index, the Government Bond Index, and the Consumer Price Index (CPI) data come from Thomson Reuters Datastream. The real estate indices limit the duration and frequency of the data, so we examine Japan on a semi-annual basis, from the first half of 1986 to the first half of 2010. For Hong Kong real estate data, we rely on the Hong Kong Rating and Valuation Department's transaction-based indices. The department publishes separate rent and price indices for the residential, office, commercial, and industrial real estate sectors. The indices have been published at least quarterly since 4Q1985. The time series for the Hang Seng Stock Price Index and for the CPI are also sourced from Thomson Reuters Datastream.

A Hong Kong bond index with sufficient historical data was not available, so our Hong Kong research is based on quarterly data from 1Q1986 through 4Q2009. All the time series are denominated in the local currency in order to avoid exchange rate problems. The nominal asset returns are the first differences of the logarithmized total return data.

Although there are other ways to measure inflation, such as the GDP deflator, we use the Consumer Price Index (CPI) as the proxy for inflation. CPI is a measure of the change in prices over time for a typical basket of consumer goods and services. Just as for asset returns, the realized changes in consumer prices are computed as the first differences π of the logarithmized price levels: $\pi_t = ln\left(\frac{CPI_t}{CPI_{t-1}}\right)$.

As the term suggests, expected inflation is based on market actors' expectations and on information available up until t-1 (Hamelink (1997)). Unexpected inflation describes the random error terms that refer to differences between expected and actual inflation. These errors stem from market inefficiencies that arise because complete information was not priced in ex ante. For our purposes, using Treasury bill returns for estimating expected inflation would lead to distorted results, because, as we noted earlier, negative interest rates for Treasury bills do not make economic sense. This is why we estimate an ARIMA (p, I, q) model in Equation (5) here to forecast inflation. Unexpected inflation then becomes simply the ex-post difference between inflation and the ARIMA projection.

Table 2 illustrates the categorization of the different phases. From 1986 - 1997 and from 2010 - 2017 the two markets experienced overall inflationary phases. The focus of our analyses is the deflationary phase from 1998 - 2010. Since we want to draw implications about the hedging properties during times of deflation, and also contrast these abilities with those in times of inflation, it is not relevant which inflationary phase is chosen for the analyses. Due to a more observations, we opted for the first inflationary period.

Table 2: Inflation rate for each subsample

	Inflation Rate Japan		**Inflation Rate Hong Kong**	
	Annual Mean	Standard Deviation	Annual Mean	Standard Deviation
1986 Q3 - 2018 Q1	0.48%	0.90%	3.77%	2.50%
1986 Q3 - 1997 Q4	1.66%	0.76%	8.60%	1.36%
1998 Q1 - 2010 Q1	-0.37%	0.35%	-0.41%	2.00%
2010 Q2 - 2018 Q1	0.19%	0.90%	3.11%	0.94%

Table 3 provides an overview of both countries' historical statistics for our sample period (mean, standard deviation, and autocorrelation), as well as the means and standard deviations for the two sub-periods, 1986 - 1997 and 1998 - 2010. For Hong Kong, the first sub-period marks a time of rising prices and substantial economic growth, while the second signals a deflationary phase and sinking rental returns. In Japan, the situation is similar. Hence, distinguishing between the two sub-periods enables us to shed additional light on the results across two different market phases.

Table 3: Annual means and standard deviations for return and price changes

		1986-2010		1986-1997		1998-2010		Autocorr. 1st order
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	
Hong Kong								
Residential properties	Prices	7.61%	13.28%	16.64%	10.84%	-1.22%	14.05%	0.47
	Rents	4.13%	7.54%	9.69%	5.68%	-1.31%	8.19%	0.51
Office properties	Prices	8.25%	18.93%	15.16%	17.96%	1.48%	19.42%	0.43
	Rents	4.35%	10.71%	9.99%	10.63%	-1.18%	10.15%	0.79
Commercial properties	Prices	9.90%	13.88%	9.90%	11.35%	3.88%	12.89%	0.43
	Rents	4.76%	6.22%	7.97%	9.12%	-2.05%	6.87%	0.45
Industrial properties	Prices	6.86%	12.18%	17.36%	12.50%	2.60%	14.32%	0.62
	Rents	2.91%	8.40%	10.17%	5.88%	-0.54%	5.38%	0.37
Stocks		10.87%	29.79%	15.75%	32.73%	6.09%	26.74%	-0.1
Inflation	realized	4.05%	2.84%	8.60%	1.36%	-0.41%	2.00%	0.74
Japan								
Residential properties	Prices	-0.99%	3.68%	2.00%	4.28%	-3.75%	1.20%	0.87
Commercial properties	Prices	-3.39%	5.79%	0.87%	6.87%	-7.30%	2.37%	0.94
Industrial properties	Prices	-1.37%	3.95%	2.34%	4.06%	-4.79%	1.67%	0.91
Stocks		-1.39%	22.94%	1.37%	21.45%	-3.94%	24.53%	0.06
Govt. bonds		4.15%	4.79%	6.47%	6.01%	2.01%	2.61%	0.06
Inflation	realized	0.60%	0.93%	1.66%	0.76%	-0.37%	0.35%	0.77

The overall results demonstrate that Hong Kong's real estate prices performed substantially better than those of Japan. In Hong Kong, prices increased between 6.86% for industrial properties and 9.9% for commercial properties. In Japan, we observed a negative trend ranging from -0.99% for residential properties to -3.39% for commercial property. Despite Hong Kong's substantially higher inflation rate of 4.05%, average real returns were positive throughout. Rents developed moderately, generally at about half the increase in prices. The highest returning asset in Hong Kong is stocks, with a 10.87% return; but it also exhibited the highest standard deviation, at 29.79%. In contrast, in Japan, stocks performed below average, with a 1.39% return at a higher standard deviation. Here, government bonds, with a 4.15% return, were the best-performing assets.

Because of how they are constructed, the transaction-based Hong Kong indices display significantly higher volatility (price changes ranging from 12.18% for industrial properties to 19.93% for office buildings); however, rents, ranging from 6.22% to 10.71%, were less volatile. On the other hand, Japan's appraisal-based price indices behaved very smoothly, their volatility ranging from 5.20% to 8.18%. The appraisers' methodology was also reflected in autocorrelations ranging from 0.87 to 0.94. In Hong Kong, the price index autocorrelations are significantly lower but also positive throughout. Thus, rents are generally more stable than prices. We note that Hong Kong's inflation volatility was nearly three times higher than that of Japan. Analysis of the two sub-periods reveals clear structural thresholds in both countries. From 1Q1986 through 4Q1997, Hong Kong experienced a period of extreme growth, marked by high price returns, strong growth in rents, and higher inflation. In stark contrast, the 1Q1998 through 4Q2009 period saw an average 0.41% rate of deflation, with rent levels simultaneously sinking in all four sectors.

Only industrial and commercial properties and stocks generated price returns of, respectively, 2.60%, 3.88% and 6.09%. From 1986 to 1997, Japan also experienced a growth phase, although it was marked overall by moderate growth rates and inflation. Government bonds returned an above-average 6.47%. From 1998-2010, our second sub-period, inflation averaged -0.37%, real estate and stocks simultaneously lost value disproportionately, and only bonds (like those in the first period) showed positive returns. Table 4 shows the cross-correlations between the realized rate of inflation and the nominal investment returns/changes in rents. Economic theory, in the context of information-efficient financial markets, posits that new information will be priced in immediately by market actors. In the real world, information efficiency is not a given, and therefore we control here simultaneously for intertemporal relationships. We study the correlations with the inflation rate lagged for up to twelve months. Especially for appraisal-based real estate indices, we assume actual market changes are gradually incorporated. But even transaction-based indices may feature market and information inefficiencies. To avoid capturing only short-term effects, we calculate the correlations in each case for quarterly (Hong Kong only), semi-annual, and annual frequencies.

Table 4: Maximum (cross-) correlations between the inflation rate lagged by k and the rent/price changes

		Cross-correlations with inflation 1986-2010						Cross-correlations with inflation 1986-1997						Cross-correlations with inflation 1998-2010					
		Quarterly		Semi-annual		Annual		Quarterly		Semi-annual		Annual		Quarterly		Semi-annual		Annual	
		k	ρ(Rn, πk)	k	ρ(Rn, πk)	k	ρ(Rn, πk)	k	ρ(Rn, πk)	k	ρ(Rn, πk)	k	ρ(Rn, πk)	k	ρ(Rn, πk)	k	ρ(Rn, πk)	k	ρ(Rn, πk)
Hong Kong																			
Residential properties	Prices	1	0.34**	0	0.5**	0	0.64**	1	0.13	0	0.15	0	0.17	0	0.2	0	0.42*	0	0,61
	Rents	0	0.5***	0	0.61***	0	0.68**	0	0.34**	0	0.43	0	0.44	0	0.38**	0	0.56**	0	0,65
Office properties	Prices	0	0.22	0	0.37	0	0.42	1	0.09	0	0.23	0	0.37	0	0.2	0	0.46*	0	0,47
	Rents	0	0.4*	0	0.45*	0	0.49	0	0.12	0	0.21	0	0.22	0	0.5**	0	0.6**	0	0,71
Commercial properties	Prices	0	0.32**	0	0.47	0	0.5	1	0.1	0	0.3	0	0.31	0	0.24	0	0.41	0	0,34
	Rents	0	0.47***	0	0.62***	0	0.76*	1	0.25*	0	0.46	0	0.65	0	0.28	0	0.42*	0	0,56
Industrial properties	Prices	0	0.33*	0	0.37*	0	0.36	0	0.16	0	0.34	0	0.33	0	0.53***	0	0.58**	0	0,63*
	Rents	0	0.36***	0	0.52**	0	0.6	1	0.19	0	0.25	0	0.33	0	0.39**	0	0.58**	0	0,78**
Stocks		1	0.01	0	0.1	0	0.12	1	0.08	0	0.08	1	0.37	0	-0.15	0	-0.05	1	-0,09
Japan																			
Residential properties	Prices			0	0.61*	0	0.64			1	0.35	0	0.35			0	0.42	0	0,55
Commercial properties	Prices			0	0.59	0	0.64			1	0.45	0	0.46			0	0.27	0	0,38
Industrial properties	Prices			0	0.69*	0	0.75			1	0.48	0	0.48			0	0.36	0	0,51
Stocks				0	0.03	0	0.09			0	-0.07	0	-0.14			1	-0.05	1	0,05
Govt. Bonds				0	0.24	1	0.33			0	-0.01	0	-0.01			1	0.29	0	0.14

Notes: The ***, ** and * indicates significance at the 1, 5 and 10% level. The effective sample size has been corrected in accordance with proposed method by Dawdy and Matalas (1964), by which first-order autocorrelation is taken into account.

2.5 Empirical Results

The regression described by Equation (2) tests the hedge characteristics of investments against expected and unexpected changes in consumer prices. The results of this research can be found in Table 5, which analyzes the overall period as well as each of the sub-periods. The expected changes in consumer prices for both countries were estimated for these models using the ARIMA (1,1,0) model.

For Hong Kong, the estimated coefficients of expected changes β_i are positive for all investments and range from 0.08 to 1.597. For the unexpected changes γ_i the estimated coefficients range from -0.199 to 1.756. The lowest coefficients are on stocks (β_i = 0.008 and γ_i = -0.199), but they are not statistically significant at any level. The coefficients for changes in real estate prices and changes in rents, with respect to expected inflation, all diverge from null at least at a 10% significance level. The magnitude of prices and rents varies dramatically, from slightly incompletely hedged to somewhat overhedged. The coefficients of unexpected changes are qualitatively very similar, although universally not statistically significant. Residential real estate prices offer the most strongly significant hedge against expected inflation; for unexpected inflation, it is industrial property prices. Japanese real estate also exhibits very good hedging behavior (β_i from 2.288 to 3.551; γ_i from 1,585 to 2.529); it strongly overhedges against expected as well as unexpected price changes. The coefficients all diverge from null at a 5% significance level. Japanese stocks act as a negative hedge (β_i = -0.309) for expected consumer price changes, but seem to react against unexpected inflation by overhedging (no statistical significance). Although government bonds offer high coefficients, the null hypothesis cannot be rejected. Thus, they do not offer protection against changes in consumer prices.

Just as with the cross-correlations, the lack of stability in the estimated results for both sub-periods is striking. The first-period values often differ markedly from those of the second period. In Hong Kong, the shift is unidirectional: Real estate strongly overhedges during deflationary periods, except for commercial properties. During the same period, stocks performed as a strong negative hedge with regard to expected deflation. Only the coefficients of housing rents diverged significantly from null in all periods (β_i = 0.949 and 1.268, or γ_i = 1.418 and 1.48) and tended to overhedge the changes in consumer prices. Hence, housing rents adjust upward during inflationary phases and downward during deflationary phases. Japanese real estate provided a significant overhedge against expected changes in consumer prices during both periods. No significant hedge characteristics are detected in stocks. Government bonds seem to have a particularly significant and negative relationship (γ_i = -4.679) to unexpected deflation. The coefficient concerning expected inflation is also negative, albeit not statistically significant. The government bond therefore offers unique opportunities in a deflationary environment. Note that the corrected coefficient of determination is relatively low for all models; only the strongly autocorrelated Japanese real estate indices attain values from 0.36 to 0.51. However, the Durbin-Watson test points to strong serial correlation in the disturbance terms for all the real estate indices. Apparently, the Fama and Schwert (1977) model does not capture the relevant factors for these returns. In addition, excluding key factors can lead to distorted estimators. Hence, we must interpret the estimator results for real estate in Table 5 cautiously.

Because of the low coefficients of determination and the DW (Durbin-Watson) statistics in the previous regression, we apply the extended regression model (3) to the complete set of real estate indices. Stocks and government bonds exhibit neither strong autocorrelation nor serial correlation in the residuals, so they are disregarded in the ARIMA model extension. The BIC proposed only an AR(1) for all Hong Kong cases. It proposed an AR(3) model for Japan, whose second AR term we did not consider meaningful, and so it was suppressed. In contrast to the pure Fama and Schwert (1977) model earlier, the DW statistics are now non-problematic (between 1.666 and 2.128), and all the adjusted coefficients of determination have increased substantially.

Table 5: Results of the regression analysis using the Fama and Schwert (1977) approach

	Predicted variables	1986-2010				1986-1997		1998-2010	
		β_i	γ_i	adj. R^2	DW-Stat.	β_i	γ_i	β_i	γ_i
Hong Kong									
Residential properties	Prices	1.597***	0.704	0.108	1.200	0.748	-0.818	1.232	1.233
	Rents	0.965***	1.349***	0.233	1.327	0.949*	1.418**	1.268*	1.48***
Office properties	Prices	1.307*	1.176	0.032	1.246	1.461	0.381	1.494	1.96*
	Rents	1.153**	1.486***	0.146	0.531	1.037	0.959	2.07**	2.406***
Commercial properties	Prices	1.015**	1.756***	0.082	1.361	1.22	0.505	0.253	2.31***
	Rents	0.844***	0.872***	0.198	1.525	1.058	0.365	0.229	0.964*
Industrial properties	Prices	0.81*	1.819***	0.098	0.867	1.962	1.378	2.236***	3.569***
	Rents	1.033***	0.638	0.124	1.492	1.119	-0.034	1.151**	1.256*
Stocks		0.08	-0.199	-0.021	2.189	1.156	-1.42	-3.291*	-0.91
Japan									
Residential properties	Prices	2.288**	1.585***	0.373	0.542	2.398	1.093	1.137***	1.057
Commercial properties	Prices	3.551***	2.529**	0.360	0.317	4.968**	1.994	1.42	2.345
Industrial properties	Prices	2.905***	1.744***	0.514	0.556	3.038**	1.145	1.382**	0.899
Stocks		-0.309	4.076	-0.032	1.793	-5.103	1.204	-3.034	-0.169
Govt. Bonds		1.06	0.824	0.011	1.987	-1.093	0.545	-0.902	-4.679*

Notes: The estimation method used is OLS. ***, ** and * indicates significance at the 1, 5 and 10% levels, respectively. The t-statistics of the regression coefficients have been adjusted for heteroscedasticity and autocorrelation using the method of Newey and West (1987). Expected inflation for both countries are estimated using the ARIMA(1,1,0) model.

Table 6: Results of the ARIMA regression on the real estate indexes for the entire sample period, 1986-2010

	Predicted variables	β_i	γ_i	AR(1)	AR(3)	ARIMA-Model	adj. R^2	DW-Stat.
		1986-2010						
Hong Kong								
Residential properties	Prices	1.483**	-0.177	0.445***	-	(1.1.0)	0.262	1.955
	Rents	0.982***	1.016***	0.364***	-	(1.1.0)	0.318	2.036
Office properties	Prices	1.206	-0.105	0.424***	-	(1.1.0)	0.182	2.005
	Rents	0.836	0.702*	0.765***	-	(1.1.0)	0.629	1.666
Commercial properties	Prices	1.048*	0.981**	0.358***	-	(1.1.0)	0.181	2.029
	Rents	0.798***	0.632**	0.248**	-	(1.1.0)	0.240	2.059
Industrial properties	Prices	1.162**	1.221**	0.603***	-	(1.1.0)	0.405	2.126
	Rents	1.057**	0.284	0.284**	-	(1.1.0)	0.179	2.128
Japan								
Residential properties	Prices	2.143*	0.338	0.914***	-0.202	(1.1.1)	0.763	1.804
Commercial properties	Prices	1.219	-0.002	1.06***	-0.172	(1.1.1)	0.877	1.961
Industrial properties	Prices	2.156*	0.136	0.983***	-0.187**	(1.1.1)	0.834	1.911

Notes: The estimation method used is OLS. ***, ** and * indicate significance at the 1, 5 and 10% levels. The t-statistics of the regression coefficients have been adjusted for heteroscedasticity and autocorrelation using the method of Newey and West (1987). The lag length of the ARIMA model was determined by using the BIC. The expected inflation is estimated with an ARIMA(1,1,0) or ARIMA(1,1,1) model as determined by the BIC.

Note further that the high autocorrelation of the real estate indices gives the first AR term a significant degree of influence; for Japanese industrial properties, only the third AR term diverges significantly from null. Hedging characteristics of Hong Kong real estate against expected inflation remained qualitatively the same. Only office properties do not diverge from null at any significance level. We observe similar results for the coefficient related to unexpected consumer price changes. Although the estimated coefficients are now universally smaller, they continue to have the same sign and to diverge significantly from null.

Japanese real estate behaves similarly relative to expected inflation: The terms estimated in the ARIMA regression are smaller in amount than in the pure Fama and Schwert (1977) regression, and they still overhedge. Commercial properties do not offer any significant hedge. Unlike our previous findings, the coefficients for unexpected inflation do not diverge significantly from null. Japanese real estate properties do not generally offer any protection here.

Because we are interested in the deflation hedging abilities of real estate, we also examine periods of inflation and deflation for both countries separately using the extended ARIMA model. Tables 7 and 8 show the estimated values for both phases.

With this model, we demonstrate a distinct change in the estimated coefficients from one period to the next. Because of the reduced sample size, the number of significant values drops considerably. Considering the overall inflationary phase, house prices in Hong Kong react very negatively (γ_i = -1.642) to unexpected inflation; rents, however, do so in a disproportionately positive way (β_i = 1.575). Rents of commercial properties appear to provide an overhedge against expected inflation. During deflationary periods, industrial property prices in Hong Kong overhedge against expected and unexpected deflation; in other words, nominal prices decline more steeply than the purchasing power of money increases.

Table 7: Results of ARIMA regression on the real estate indexes, sub-period 1986-1997 (inflation)

	Predicted variables	1986-1997 β_i	γ_i	AR(1)	AR(3)	ARIMA-Model	adj. R^2	*DW*-Stat.
Hong Kong								
Residential properties	Prices	0.847	**-1.642***	**0.537*****	-	(1.1.1)	0.220	1.605
	Rents	0.744	**1.575****	0.085	-	(1.1.1)	0.089	1.958
Office properties	Prices	0.219	-0.996	**0.474*****	-	(1.1.1)	0.129	1.923
	Rents	0.826	0.498	**0.779*****	-	(1.1.0)	0.584	1.497
Commercial properties	Prices	1.748	0.787	0.121	-	(1.1.0)	-0.036	1.650
	Rents	**1.472****	0.589	0.055	-	(1.1.0)	0.025	2.017
Industrial properties	Prices	2.34	1.241	**0.551*****	-	(1.1.0)	0.264	2.057
	Rents	1.278	-0.175	0.268*	-	(1.1.0)	0.039	2.137
Japan								
Residential properties	Prices	2.886	0.102	**0.922*****	-0.24	(1.1.1)	0.637	1.966
Commercial properties	Prices	**2.614***	-0.108	**1.064*****	-0.21	(1.1.1)	0.824	2.137
Industrial properties	Prices	**3.474****	-0.039	**0.979*****	**-0.256***	(1.1.1)	0.713	2.200

The results for the Fama and Schwert (1977) model are confirmed for Japanese real estate considering the inflationary period. All three categories overhedge (statistically significant only for commercial and industrial properties) against expected inflation. The estimators for unexpected inflation are not statistically significant. During the deflationary period, real estate appears to represent a negative hedge against expected and unexpected deflation. The coefficients for residential real estate in both cases diverge from null at the 10% significance level.

Table 8: Results of the ARIMA regression on the real estate indexes, sub-period 1998-2010 (deflation)

	Predicted variables	1998-2010 β_i	γ_i	AR(1)	AR(3)	ARIMA-Model	adj. R^2	*DW*-Stat.
Hong Kong								
Residential properties	Prices	1.18	0.011	**0.39***	-	(1.1.0)	0.111	2.110
	Rents	1.144	0.653	**0.522*****	-	(1.1.0)	0.306	1.876
Office properties	Prices	1.405	0.041	**0.403*****	-	(1.1.0)	0.127	2.019
	Rents	0.821	0.885	**0.746*****	-	(1.1.0)	0.598	1.821
Commercial properties	Prices	-1.231	1.244	**0.408*****	-	(1.1.1)	0.264	2.245
	Rents	-0.718	0.384	**0.494*****	-	(1.1.1)	0.267	1.793
Industrial properties	Prices	**1.987*****	**2.34*****	**0.606*****	-	(1.1.0)	0.523	2.184
	Rents	0.868	0.643	**0.342***	-	(1.1.0)	0.180	1.986
Japan								
Residential properties	Prices	**-0.83***	**-0.353***	**1.037*****	**-0.312****	(1.1.1)	0.783	1.624
Commercial properties	Prices	-0.135	-0.083	**1.176*****	**-0.328****	(1.1.1)	0.859	1.931
Industrial properties	Prices	-0.523	-0.402	**1.089*****	**-0.331*****	(1.1.1)	0.805	2.040

Notes: The estimation method used is OLS. ***, ** and * indicate significance at the 1, 5 and 10% levels. The t-statistics of the regression coefficients have been adjusted for heteroscedasticity and autocorrelation using the method of Newey and West (1987). The lag length of the ARIMA model was determined by using the BIC. The expected inflation is estimated with a ARIMA(1,1,0) or ARIMA(1,1,1) model as determined by the BIC.

2.6 Conclusion

This study examines historical asset returns for Hong Kong and Japan over the 1986 - 2009 period to determine whether real estate, stocks, and bonds can provide a hedge against inflation or deflation. We divide our dataset into two subsamples, an inflationary period and a deflationary period. We were able to show empirically that real estate prices and rents are strongly linked to consumer prices. This confirms the existent studies on this subject.

Our results show that real estate almost perfectly hedges, or even overhedges, against expected changes in consumer prices for both countries. By comparison, we find no statistically significant link to consumer prices for stocks or government bonds. Only considering Hong Kong, real estate partially provides a hedge against unexpected changes in consumer prices. In Hong Kong, residential real estate constitutes the best hedge against expected changes, and industrial real estate the best hedge against unexpected changes. For Japan, residential as well as industrial real estate properties equally overhedge against expected changes in consumer prices. Real estate in Japan does not provide a statistically significant hedge against unexpected changes in consumer prices.

The results for the 1986 - 1997 (inflationary phase) and 1998 - 2009 (deflationary phase) sub-periods are somewhat more difficult to interpret. We find that asset prices in the two countries behave differently. Using the Fama and Schwert (1977) model, enhanced with ARIMA specifications, the results in the first sub-period phase qualitatively resemble those of the overall period, although with parameters that are often not statistically significant. Hong Kong's residential real estate prices provided a negative or perverse hedge. Hence, residential real estate prices are higher when unexpected inflation is low and vice versa. This might be due to poor inflation forecasts and therefore market inefficiency. However, we can rule out poor forecasts since we predicted the expected inflation rate using a sound ARIMA model and used the CPI as a proxy for the actual inflation rate. Thus, equilibrium expected real housing prices are in fact negatively related to unexpected inflation. Residential rents work as a positive hedge, against unexpected inflation, while rents for commercial properties rose simultaneously disproportionately against expected inflation. In Japan, commercial and industrial real estate succeeded in overcompensating for expected inflation.

Statistically insignificant parameters are also pervasive in the second, or deflationary, phase. Here, we observe that industrial real estate prices in Hong Kong adjusted excessively for expected and unexpected deflation, and real values declined as a result. In contrast, Japanese housing prices provided a negative hedge against expected and unexpected inflation. We therefore conclude that, in a deflationary environment, real estate provides value stability in real terms (i.e., deflation protection). Our conclusions about stocks and government bonds are similar to those reached in other studies of diverse markets and time periods. Neither investment constitutes a significant hedge against consumer price changes. Only during the second sub-period in Japan we observe a significant negative relationship to unexpected deflation for government bonds. However, this result is economically questionable, given that interest rates on government bonds usually adjust gradually to prevailing inflation levels.

2.7 Bibliography

Adrangi, B., Chatrath, A., & Raffiee, K. (2004). REIT Investments and Hedging Against Inflation. Journal of Real Estate Portfolio Management, 10(2), pp. 97-112.

Bodie, Z. (1976). Common Stocks as a Hedge against Inflation. The Journal of Finance, 31(2), pp. 459-470.

Boudoukh, J., & Richardson, M. (1993). Stock Returns and Inflation: A Long-Horizon Perspective. American Economic Review, 83(5), pp. 1346-1355.

Brueggeman, W. B., Chen, A. H., & Thibodeau, T. G. (1984). Real Estate Investment Funds: Performance and Portfolio Considerations. Real Estate Economics, 12(3), pp. 333-354.

Dawdy, D., & Matalas, N. C. (1964). Statistical and Probability Analysis of Hydrologic Data in Handbook of Applied Hydrology. New York: McGraw-Hill.

Day, A. (2015). Investing in Resources: How to Profit from the Outsized Potential and Avoid the Risks. Hoboken: Wiley.

Dimson, E., Marsh, P., & Staunton, M. (2012). Global Investment Returns Yearbook 2012. Zürich: Credit Suisse AG.

Fama, E. F. (1975). Short-term Interest Rates as Predictor of Inflation. The American Economic Review, 65(3), pp. 269-282.

Fama, E. F., & Gibbons, M. R. (1982). Inflation, Real Returns and Capital Investment. Journal of Monetary Economics, 9(3), pp. 297-323.

Fama, E. F., & Schwert, W. G. (1977). Asset Returns and Inflation. Journal of Financial Economics, pp. 115-146.

Fisher, I. (1930). The Theory of Interest, as determined by Impatience to Spend Income and Opportunity to Invest it. Journal of Monetary Economics (4), pp. 2151-2197.

Fleckenstein, M., Longstaff, F. A., & Lustig, H. (2014). The TIPS-Treasury Bond Puzzle. The Journal of Finance, 69(5), pp. 2151-2197.

Ganesan, S., & Chiang, Y. H. (1998). The Inflation-Hedging Characteristics of Real and Financial Assets in Hong Kong. Journal of Real Estate Portfolio Management, 4(1), pp. 55-67.

Glascock, J. L., Fan, L., Feng, Lei, & Bao, H. X. (2008). Inflation Hedging Characteristics of Real Estate Assets in Hong Kong. Statistics & Decisions (301), pp. 1-25.

Gultekin, M. N., & Gultekin, B. N. (1983). Stock Market Seasonality: International Evidence. Journal of Financial Economics, 12(4), pp. 469-481.

Gyourko, J., & Linneman, P. (1988). Owner-occupied Homes, Income-producing Properties, and REITs as Inflation Hedges: Empirical Findings. The Journal of Real Estate Finance and Economics, 1(4), pp. 347-372.

Hamelink, F., Hoesli, M., & MacGregor, B. (n.d.). Inflation Hedging versus Inflation Protection in the US and the UK. Real Estate Finance, 14(2), pp. 63-73.

Hartzell, D., Hekman, J. S., & Miles, M. E. (1987). Real Estate Returns and Inflation. Real Estate Economics, 15(1), pp. 617-637.

Harvey, A. C. (1990). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.

Jaffe, J. F., & Mandelker, G. (1976). The "Fisher Effect" for Risky Assets: An Empirical Investigation. Journal of Finance, 31(2), pp. 447-458.

Limmack, R. J., & Ward, C. W. (1988). Property Returns and Inflation. Land Development Studies, 5(1), pp. 47-55.

Liu, C. H., Hartzell, D. J., & Hoesli, M. E. (1997). International Evidence on Real Estate Securities as an Inflation Hedge. Real Estate Economics, 25(1), pp. 193-221.

Lizieri, C., Hoesli, M., & MacGregor, B. (2008). The Inflation Hedging Characteristics of U.S. And U.K. Investments: A Multi-Factor

Error Correction Approach. Journal of Real Estate Finance and Economics, 38(2), pp. 183-206.

Maurer, R., & Sebastian, S. P. (2002). Inflation Risk Analysis of European Real Estate Securities. The Journal of Real Estate Research, 24(1), pp. 47-78.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), pp. 703-708.

Obereiner, D., & Kurzrock, B.-M. (n.d.). Inflation-hedging Properties of Indirect Real Estate investments in Germany. Journal of Property Investment & Finance, 30(3), pp. 218-240.

Park, J. Y., Mullineaux, D. J., & Chew, I.-K. (1990). Are REITs Inflation Hedges? The Journal of Real Estate Finance and Economics, 3(1), pp. 91-103.

Rubens, J., Bond, M., & Webb, J. (1989). The Inflation-Hedging Effectiveness of Real Estate. Journal of Real Estate Research, 4(2), pp. 45-55.

Yobaccio, E. R. (1995). The Inflation-Hedging Properties of Risk Assets: The Case of REITs. Journal of Real Estate Research, 10(3), pp. 279-296.

2.8 Appendix

$R_{i,t}^{n}$	Nominal investment return of asset i in time t
$R_{i,t}^{r}$	Real investment return
$R_{TBill,t}^{n}$	Return of U.S. treasury bill, used as the constant real interest rate
ϕ_t	Information on which expectations are based
π_t	Actual inflation rate
π_t^e	Expected inflation rate
$\pi_t - \pi_t^e$	Unexpected inflation rate
α_i	Intercept of regression model
β_i	Regression coefficient, signals hedging effectiveness concerning expected inflation
γ_i	Regression coefficient, signals hedging effectiveness concerning unexpected inflation
$\epsilon_{i,t}$	Residuals of the estimated model
a_j	Coefficient of the AR-component of the ARIMA model
b_k	Coefficient of the MA-component of the ARIMA model
$P_{i,t}$	Price of asset i in time t

3 AR-GARCH-EVT-Copula for Securitized Real Estate: An approach to improving risk forecasts?

Authors: Carsten Fritz, Cay Oertel

Submission to: The Journal of Property Research (JPR)

Status: published in 2021

Fritz, C., Oertel C. (2021). AR-GARCH-EVT-Copula for Securitized Real Estate: An approach to improving risk forecasts? Journal of Property Research, 38(1), pp. 71-98.

Abstract

This study presents a quantitative analysis of the so-called AR-GARCH-EVT-Copula model aimed at forecasting price risk metrics for multi-asset portfolios, including securitized real estate positions. The exposure of securitized real estate to price risk is well documented, and mainly driven by the dynamics of the underlying direct property markets. Thus, the statistical characteristics of securitized real estate returns are influenced by direct property markets behavior. In order to capture the exposure of securitized real estate investors, the model incorporates a non-linear dependence structure and time-varying volatility models for asset returns. Accordingly, an empirical study using data from six major global markets is carried out. An AR-GARCH-EVT-Copula model is applied in order to forecast price risk metrics, in comparison to classical methods

like historical simulation and variance-covariance models. Forecasts are then compared with realized returns, in order to calculate hit sequences and conduct statistical interference on the respective models. It is empirically shown that, the AR-GARCH-EVT-Copula model provides a superior forecast concerning price risk metrics. This is mainly due to the usage of copulae, allowing us to individually model the dependence structure of the return series. Back-testing and test results confirm the superiority of our model in terms of price risk forecasting accuracy. The decomposition of the univariate and multivariate models of the target model reveals the necessity to allow for high order and thus long-lasting autoregressive modelling as well as asymmetric tail dependence and rotated copulae across different portfolios.

3.1 Introduction

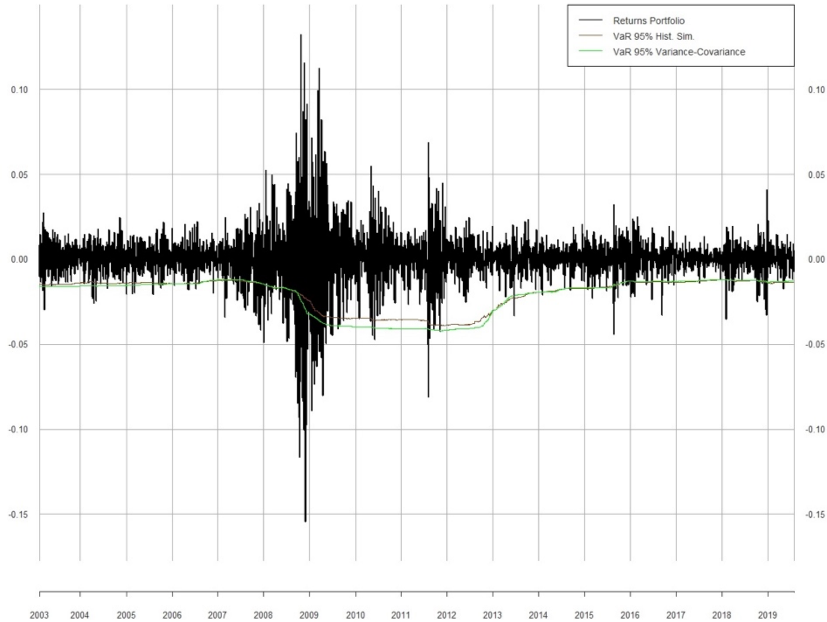
The risk modelling of investment positions has attracted larger interest, since the global financial crises (GFC) in 2008 painfully demonstrated the vulnerability of international financial markets. Central issues emerging from this turbulent period have included the contagion effects of direct real estate markets and corresponding mortgage positions towards other asset classes such as equities, which were affected by the triggered macroeconomic downturn (Hui and Chan, 2013).

Subsequently, institutional debt and equity investors have both experienced a tangible tightening of the regulatory framework, including Basel III (and its addendum know as Basel III reform package) as well as Solvency II. Especially banks and insurance companies are facing increased legal obligations relating to their internal price risk models, in case they are holding public equity positions, which are exposed to the risk of market price changes (Ergen, 2015). Risk measures for price risk such as the Value at Risk (VaR) or Conditional Value at Risk (CVaR) are typically underestimated and capital requirements insufficient, if they are computed on the assumption of normality and independence, while the returns of the multi-asset portfolios are in reality leptokurtic, and entail skew and autocorrelation (Liow, 2008). Rossignolo et al. (2012) advise an application of extreme value theory (EVT) as a potential response to the abovementioned regulatory challenge, including the correct modelling of skewed and fat-tailed returns.

Accordingly, the potential to model the data of non-normal returns, as well as the dependence structure of these positions so as to estimate joint extreme value losses, is of particular interest for the risk management of institutional investors. The classic approach of using linear concepts such as the Bravais-Pearson correlation coefficient for original time series data does not provide information about the structure and assumes an elliptical joint distribution of the assets (Wu and Lin, 2014).

With regard to the price risk of securitized real estate, however, dependence structure modelling towards other asset classes is scarce. Dependence modelling of securitized real estate either only discusses the structures within the specified asset class itself (Knight et al., 2005, Goorah, 2007), or with other asset classes, but without prior univariate AR-GARCH-EVT modelling (such as Dulgerov, 2009). Figure 2 shows the importance of a correct model to measure VaR as well as CVaR, especially during times of crises.

Figure 2: VaR (α = 0.95) estimates for Real Estate–Stocks portfolio (US)



Notes: Figures for VaR 99%, CVaR 95% and CVaR 99% for real estate-stocks and real estate-bonds portfolios are available upon request. All of these figures look similar to this figure concerning the hits of historical simulation and variance-covariance models with the return series.

Historical simulation and variance-covariance are two conventional tools for measuring VaR and CVaR. Figure 2 shows the problem of the currently applied models, because the models for the price risk of the portfolios are clearly failing to provide accurate price risk forecasts. Thus, the standard methodology is not able to make valid statements about the actual risk exposure, leading to potentially wrong risk bearing capabilities in terms of equity underlying. Hence, the need for a better model is obvious. This necessity seems to be even more important when considering, that extreme observations are particularly common in securitized real estate return series, due to the integration in direct markets and potential herding behavior due to drastically changing return expectations (Hoesli and Oikarinen, 2012).

The real estate literature has not been linked to the body of literature applying AR-GARCH-based univariate modelling and EVT to account for heteroscedastic and autocorrelated time series, as originally proposed by McNeil and Frey (2000). The connection between the abovementioned bodies of literature, which leads to the so-called *AR-GARCH-EVT-Copula* approach and its subsequent empirical study of the feasibility of enhancing price risk forecasting using the specified approach evaluation is among other preliminary results, the main motivation of the present study.

Accordingly, the central research question is whether the AR-GARCH-EVT-Copula approach can improve price risk forecasts for investors holding portfolios containing securitized real estate. Therefore, the study sets up the AR-GARCH-EVT-Copula model to account for the abovementioned statistical challenges associated with financial time series data. Subsequently, the study models the dependence structures, and forecasts the VaR and CVaR based on these univariate and multivariate models. Finally, a back-testing procedure compares forecasts with real returns to evaluate the model in comparison to known approaches such as variance-covariance and historical simulation.

Thus, the paper contributes to the existing real estate literature in several ways. Predominantly, a methodologically innovative application of the AR-GARCH-EVT-Copula technique including price risk metric forecasting is provided. To the best of the authors' knowledge, the approach has not yet been applied to multi-asset portfolios which include securitized real estate.

This study is structured as follows in order to make the contribution described above: Section 2 reproduces the most important related literature and derives the hypothesis. Section 3 explains the methodological approach. Section 4 describes the data and the ensuing section presents the results in terms of the risk forecast accuracy and model errors across various copula types. Section 6 concludes and outlines further research.

3.2 Literature Review

The following bodies of literature are relevant as framework for the present study, namely: Stylized facts of financial time series and the underlying economic drivers which cause the problematic statistical features (with special focus on securitized real estate, but also for stocks and bonds), univariate conditional mean and volatility modelling including EVT and the corresponding standardization procedures of the data, dependence modelling and the evaluation possibilities of risk forecast models, as well as the subsequent risk management implications. The present literature review is supposed to outline the actual problem set as well as existing studies in the field to illustrate the research gap. The subsequent methodology section will then outline the actual models in a more mathematical and formal way, including the equations of the approach.

The fundamental driver for the present study is the body of literature exploring stylized facts about the returns of stocks, bonds and securitized real estate. Primarily, stylized facts of daily securitized real estate returns are important for the present study. As shown by Hoesli and Oikarinen (2012), the specified returns are predominantly a function of the returns of the vehicles' underlying assets, and explicitly not only of the overall stock market. Accordingly, the features of direct property markets are highly relevant for the application of a price risk forecasting methodology for securitized real estate. Since direct real estate returns are widely known for non-normality (Byrne and Lee, 1997; Young et al., 2006; Richter et al., 2011), these underlying assets pass their statistical return characteristics through the securitizing vehicle.

Additionally, direct property markets also show autocorrelation of their returns. In this context, various studies have shown the autocorrelation and thus predictability of direct real estate returns, denying the classic assumption of market efficiency or random walk behavior (Wheaton et al., 1999; Payne and Sahu, 2004; Coleman and Mansour, 2005). Reasons for this are relatively high transaction costs, low turnover volumes, tax-related issues, asymmetric information and the heterogeneity of the commodity itself (Schindler, 2010). Just like the stylized fact of non-normality, autocorrelation is also passed through the securitizing vehicle, as empirically shown by e.g. Kuhle and Alvayay (2000). In this context, the authors differentiate between short- and long-term autocorrelation. The main reason for short-term autocorrelation in daily securitized real estate returns is assumed to arise mainly from differing information availability across investors. Long-term autocorrelation of daily returns is mainly caused by the long-lasting nature of cash flows from the leases of the underlying properties. Thus, a clear relationship between the characteristics of the held real estate assets and the resulting stylized facts of the return series of the securitizing vehicle can be identified.

At last and in addition to non-normality and autocorrelation, the volatility of securitized returns needs to be addressed. The central finding in the existent body of literature is the heavy volatility clustering with differing variance across time (Cotter and Stevenson, 2006; Jirasakuldech et al., 2009). From an economic point of view, Letdin et al. (2019) review the underlying mechanisms for this phenomenon. The authors name the low transparency and high capital volumes but also potentially suddenly changing information about property markets and investments as decisive driver for simultaneous investor decisions. These synchronic movements of investors are causing the volatility clustering. Based on these stylized facts, it can be concluded, that a feasible price risk forecasting model for any portfolio, containing securitized real estate positions needs to be able to account for the non-normality, autocorrelation and volatility clustering of the return series.

In addition to securitized real estate, stocks and bonds are the typical investment targets for multi-asset real estate investors seeking diversification (e.g. Hoesli et al., 2003). For market data concerning stocks, studies analyzing the distributional characteristics date back to the 1960, doubting classic Gaussian assumptions (Mandelbrot, 1963; Fama, 1965). Regarding normality, a large body of literature has empirically shown the existence of negative skew and leptokurtosis and additional fat tails (e.g. Officer, 1972; Bekaert and Harvey, 1998; Harris and Kücüközmen, 2001). Studies cite overreaction and herding behavior as a potential explanation (de Bondt and Thaler, 1985). Consigli (2002) also highlights the heterogeneity of financial markets, since they are especially prone to country- and period-specific risk, causing heavy intertemporal autocorrelation and the associated volatility clustering.

Bond return data is also known for skewed and leptokurtic returns (Rachev et al., 2003, Wu and Lin, 2014). Just as for the previously described equity returns, bond returns across various maturity levels are not normally distributed and are especially fat tailed, which are methodologically explored by means of highly significant kurtosis parameters of stable distributions (Gabriel and Lau, 2014).

Summarizing the existent literature for the cross-section of assets, return series are highly questionable regarding Gaussian assumptions. The named stylized facts cause biased statistical measures and lead to false asset allocation (Dittmar, 2002). In addition, falsely modelled tails, assuming perfectly elliptical asset returns, cause tail risk estimation and its hedging to fail which is highly important for strategic portfolio management.

Based on these obstacles of non-normality, autocorrelation and heteroscedasticity of financial time series data, McNeil and Frey (2000) have introduced the AR-GARCH-based standardization of returns, in order to account for the outlined problematic stylized facts.[1] Within the cited body of literature for univariate volatility modelling, two central methodical questions are of interest: degrees of autoregressive components of the conditional mean model and the distribution assumption of the error terms for the conditional volatility model.

First of all, the autoregressive and moving average components need to be specified. Interestingly, the literature agrees on an autoregressive component (see Rocco, 2014 for an overview). Regarding the distribution of the errors of the conditional volatility model, normally- and (skewed) t-distributed error terms are options. Skewed t-distributions have largely shown improvements in VaR predictions (Küster et al., 2006; Bali and Theodossiou, 2008; Mabrouk and Saadi, 2012).

Based upon the conditional volatility model of McNeil and Frey (2000), the decomposition of the distribution to model fat tails is necessary. Therefore, EVT is applied to model the observations over a threshold in the tails, assuming them to follow a Generalized Pareto Distribution (GPD). In combination with the univariate GARCH modelling, the resulting combined GARCH-EVT approach has been used by various studies (Bhattacharyya and Ritolia, 2008; Chan and Gray, 2006; Deng et

[1] As well as EVT application to the fat tails, which will be reproduced in detail below.

al., 2011). The GARCH-EVT-based univariate estimation of tail also entails two crucial advantages: It is based on well-established statistical theory and also enables a parametric estimation (Karmakar, 2017). With regard to the goal of the present study to forecast risk metrics more precisely, Bao et al. (2006), Küster et al. (2006), Bali (2007) are examples of enhanced risk metric forecasting performance, due explicitly to EVT application to the tails. The so-far described procedure accounts for standardizing the data, and generating independently, identically distributed observations. Classic approaches such as variance-covariance or historical simulation do not apply the named procedure to the original return series. Thus, the return series of these models still yield the specified issues and cause bias to the risk metrics.

Subsequently, the need to model the multi-asset dependence arises. The main economic reason for potentially non-linear tail dependence is the similarity of the underlying macroeconomic drivers for property market and returns of the broader stock market from industrial production etc. (Christoffersen et al., 2014). Traditional linear correlation models such as the widely adopted Bravais-Pearson's coefficient, however, only measure the degree of explicitly linear dependence. It needs to be highlighted that variance-covariance models for forecasting risk metrics exactly assume constant and linear dependence across time.

Accordingly, these models provide no information about the structure of the dependence.

From an empirical point of view, especially the additional proposition of Hoesli and Oikarinen (2012) regarding real estate's integration into the broader stock market has gained attention, because equity and securitized real estate returns are assumed to show heavy tail dependence for the outlined reasons. Empirical findings of various studies confirm this tail dependence of securitized real estate and stocks (Huang and Zhong, 2013; Yang et al., 2012).

The abovementioned reasons have motivated researchers to develop alternative concepts of dependence structure modelling, as firstly proposed by Sklar (1959), and introducing copula functions. Convening the usage of the correct copula, authors like Kole et al. (2007) and Hurd et al. (2007) find that the goodness-of-fit of an Archimedean Student-t as well as other copulae is superior to that of an elliptical Gaussian copula, for the reasons given of simultaneous heavy downturns and thus left tail dependence.[2] The described tail dependence for simultaneous extreme losses of securitized real estate and stocks in the same nation are a direct consequence of the similarity of risk factors such as fundamental macroeconomic drivers. This detail is especially crucial for

[2] Tail dependence of the individual copula families will be discussed in detail below.

risk management purposes, since the copula function is supposed to correctly explicitly model the lower tail or asymmetric tail dependence respectively.

The usage of copulae in real estate literature is scarce, although existing (Goorah, 2007; Dulgerov, 2009). Knight et al. (2005), as well as Chang et al. (2011), have adopted a non-linear modelling of multi-asset portfolios including real estate.[3] Nonetheless, the authors do not apply EVT to the univariate return distribution before modelling the bivariate tail dependence. However, they find heavy asymmetric tail dependence, especially in downturn markets. Since they find time-variant dependence, approaches which model dependence as constant across time are expected to perform worse in comparison. Hoesli and Reka (2013) found the same time-variance of the co-movement, especially for the tails of returns of securitized real estate and stocks. The associated asset class of infrastructure equities was analyzed in a closely-related study by Chakkalakal et al. (2018). It should be explicitly emphasized, that the named articles broadly assess parameters of the copulae, without any risk metric forecasting context.

Lastly, the methodical approach used to evaluate improvements to risk models is important. In this field, the literature has mainly focused on

[3] For a more technical approach on the details of the methodology in the broader stock market, we recommend the study of Wei and Zhang (2004).

back testing of risk metric forecasts (summarized by Du and Escanciano, 2017). Essentially, back-testing procedures estimate forecasts using the risk model and compare these values with true realizations, as conducted by Wu and Lin (2014) or Sahamkhadam et al. (2018). Whenever the model underestimates the risk metric for the period to be forecasted, a so-called “hit” occurs. These hits are collected in a binary vector and compared to the confidence level of the model (Kupiec, 1995 and Christoffersen, 2004). Normally, new approaches to forecast risk metrics are compared to benchmark models of historical simulation and variance-covariance. Based on the abovementioned literature, the following hypothesis is derived as the foundation for our own empirical study of the AR-GARCH-EVT-Copula: The AR-GARCH-EVT-Copula approach to estimating forecasts of risk metrics generates more accurate risk metric forecasts of portfolios containing securitized real estate, in comparison to classic approaches such as historical simulation or variance-covariance. This hypothesis is formulated, because risk models for multi-asset portfolios which account for autocorrelation, skew and fat tails, as well as non-linear dependence, are assumed to outperform their classic counterparts.

3.3 Methodology

The present study assesses the feasibility of the AR-GARCH-EVT-Copula approach to improving the forecasts for the VaR of multi-asset portfolios, which include securitized real estate. Based on the above-mentioned literature review, the methodology is supposed to describe the actual methodological translation to set up the AR-GARCH-EVT-Copula model. As benchmark methodologies, the study applies classic variance-covariance and historical simulation methods, which are not extensively discussed here. However, the basic idea of price risk forecasting for financial portfolios is the anticipation of future return changes based on available univariate or multivariate information such as past returns or co-movements of the portfolio constituents.

Essentially, our AR-GARCH-EVT-Copula methodology of interest is an algorithm, which refits univariate and multivariate models to rolling windows of time series data, in order to forecast the VaR for the day ahead of the analyzed part of the data by simulating return data for the profit-and-loss function (P&L). Since the forecasting of portfolio returns and subsequent risk metric calculation require univariate modelling of the individual return series as well as the dependence structure, these steps are presented in detail.

The univariate AR-GARCH modelling for each window of the return time series containing daily log returns $r_t, t \in [0, T]$ can be summarized by the following set of equations:

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \tag{6}$$

$$\mu_t = \mu + \sum_{i=1}^{s} \gamma_i r_{t-i} \tag{7}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{8}$$

$$z_t \sim skewed - t(0,1) \tag{9}$$

The return equation (6) is a function of the conditional mean and an error component ε_t, which can be rewritten as the product of the conditional volatility and the error z_t. The conditional mean Equation (7) for μ_t yields past returns r_{t-i}, and a constant term μ. Thirdly, the conditional variance σ_t^2 is modelled by equation (8) as a function of past variance σ_{t-i}^2 as well as a quadratic error term ε_{t-i}^2. Lastly, the error terms of the return Equation (6) are assumed to follow a skewed t-distribution for the outlined reasons of leptokurtic return behavior, as expressed by Equation (9).

The order for the AR models is adjusted for each rolling window of 1000 observations by testing for the minimum Akaike Information Criterion (AIC) up to order 5. For the conditional variance model, the study follows

Hansen and Lunde (2005) or Wang et al. (2010) by applying a GARCH(1,1) model.

Conditional on the available information of each rolling window, the model parameters of $\hat{\lambda} = (\hat{\mu}, \hat{\gamma}, \hat{\omega}, \hat{\alpha}, \hat{\beta})$ are estimated. Additionally, the one day ahead conditional mean, as well as conditional volatility are estimated, denoted by $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ for $t \epsilon [1000, T-1]$. Thus, the first 1000 days of the data set represents the burn-in sample, for which no risk metrics are calculated. The first day for which the study estimates risk forecasts is the 1001st day. The estimates are saved for the later simulation of the P&L function of the one-day-ahead returns.

More importantly, the estimated standardized residuals are extracted to model the dependence, since they are expected to satisfy the assumption of independent and identical distribution, so as to produce unbiased estimates:

$$\hat{z}_t = \frac{r_t - \hat{\mu}_t}{\sigma_t} \tag{5}$$

Nonetheless, these standardized residuals may still exhibit fat tails, which can be modeled directly by EVT, in particular by the peak-over-threshold method assuming tails to follow a GPD, proposed by McNeil et al. (2005). In choosing the correct threshold, there is a trade-off that should be noted. If selected too low, there may not be enough data points in the tails to ensure an unbiased estimation, and some data points which are relatively far from the actual kernel distribution are not considered in the tails. By introducing a lower threshold, more observations from the center of the distribution are introduced into the series we want to cut off which makes the estimator less volatile but increases the bias of a tail distribution which should consist of extreme observations. Following DeMelo Mendes (2005), we set the threshold to the 10% quantile for the left part and to the 90% quantile for the upper part of the distribution. For a further assessment of the correct threshold selection, mean excess functions and so-called Hill plots were considered, as well. (Wang et al., 2010).

Modelling the standardized residuals is achieved by using the GPDs for the marginal distributions of the tails, in combination with the empirical distribution for the interior kernel. Equation (10) illustrates the newly created distribution:

$$F(z)\begin{cases} \frac{n_L}{n}\left\{1+\xi_L\frac{(z-u_L)}{\psi_L}\right\}^{-1/\xi_L} & z<u_L \\ \phi(z) & u_L<z<u_R \\ 1-\frac{n_R}{n}\left\{1+\xi_R\frac{(z-u_R)}{\psi_R}\right\} & z>u_R \end{cases} \tag{10}$$

where u_L, and u_R are the lower and upper threshold respectively. n denotes the overall number of observations of z and n_L, n_R represent the number of observations that are in excess of the thresholds. Scale

(ψ) and shape (ξ) are then estimated via maximum likelihood. $\phi(z)$ represents the empirical distribution of the Gaussian kernel. The specified procedure completes the univariate modelling and preparation of standardized residuals for each of the $T - 1000$ rolling windows.

Based upon the described univariate modelling, the second step is to model the multivariate relationship, since the simulation of portfolio returns requires information about the dependence structure of the assets. As outlined, various economic factors cause dependence structures to be non-linear.

Thus, the present study uses copula functions to model the dependence, in contrast to classic measures like the Bravais-Pearson correlation coefficient. A copula is a function that couples a multivariate distribution function to its univariate marginal distributions, and does not require any assumptions on the selection of the distribution function, as introduced by Sklar (1959) and Sklar (1973).

Formally, a copula function C can be expressed as a link of the marginal distributions between the random variables, or in this specific case as a bivariate copula of the standardized residuals, $\hat{z}_{i,t}$, for real estate positions and stocks or bonds:

$$F\left(\hat{z}_{1,t}, \hat{z}_{2,t} \vee \hat{z}_{3,t}\right) = C\left(F_1\left(\hat{z}_{1,t}\right), F_2\left(\hat{z}_{2,t} \vee \hat{z}_{3,t}\right)\right), \text{ for } \hat{z}_1, \hat{z}_2, \hat{z}_3 \quad (11)$$

C is a bivariate distribution function containing the marginals $F_1$and F_2 of the two assets in the portfolio of interest. If F_i is the joint distribution function of a random vector with continuous marginals of F_1 and F_2, then C is unique and given by:

$$C(p_1, p_2) = F\left(F_1^{-1}(p_1), F_2^{-1}(p_2)\right), \text{for all } (p_1, p_2) \epsilon [0,1]^n \qquad (12)$$

There are mainly two main families of copulae, namely elliptical and Archimedean, containing a variety of parametric copula types. As noted by Nelsen (1999), Archimedean copulae allow for asymmetry in the tail dependence. Typical examples of such asymmetric couplae are the Frank, Gumbel, BB1, BB2 and BB7. As described above, the literature has shown the potential of asymmetric dependence, especially in the tails between securitized real estate and stocks or bonds respectively. Accordingly, Archimedean copulae are incorporated into the modelling.

In this context, different copulae also allow for different tail dependence. Each copula family has its own formula to derive the lower and upper tail dependence. If the two tail dependences are equal, there is symmetrical behaviour, which, due to the abovementioned reasons is rather unexpected.

For example, the common Gaussian copula has zero tail dependence, whereas the asymmetrical Gumbel copula has right or upper tail dependence, but zero left or lower tail dependence. Other copulae like

BB1 and BB7 have tail dependence of different, non-zero, strength (e.g. Gumbel (1960), Clayton (1978), Frank (1979), Joe (1993 & 1997) and Nelsen (1999)).

For each rolling window and its pair of standardized residuals, the named copulae are fitted to estimate the model parameters following the inference-for-margins (IFM) approach proposed by Joe and Xu (1996). The copula, which shows the lowest AIC for the respective window is chosen as dependence structure model. In this study, copulae with one and two parameters as well as their 90°, 180° and 270° rotated peers are tested (see Table 14 for the full list of the 28 copula types).

Based on the above pattern, the methodology models the individual asset returns as well as the dependence structure between them for each window, as a foundation for the simulation of the one-day-ahead forecast of returns. Nonetheless, one may wonder how the dependence modelling of standardized residuals translates into return forecasts. Here, the decisive methodological step is carried out: Probability integral transformation (PIT) of the standardized residuals.[4] This transformation uses random numbers from the multivariate distribution. Subsequently, the correlation matrix of the copula is disintegrated, and the residuals

[4] The transformation methodology differs across the copula families; nonetheless, the basic idea is consistent. See Wang et al. (2010) for more details on differences for elliptical and Archimedean copulae.

for the univariate model are then generated using the inverse of the joint distribution, namely F_i^{-1}.This simulation of the one-day-ahead residuals out of the named distribution is carried out M times to generate the simulated residuals, $z_{i,t+1}$. For the simulation, 10,000 return scenarios from the estimated AR-GARCH-EVT-Copula model are generated. Notably, to check whether the number of simulations is sufficient enough, the simulation was also performed 50,000 times. The results were not significantly different, in fact, they were the same. These residuals are then incorporated into Equation (6), as expression for the individual returns of the two assets of interest:

$$\hat{r}_{i,t+1} = \mu_i + \hat{z}_{i,t+1}\hat{\sigma}_{i,t+1}, \qquad i = 1,2. \tag{13}$$

Based on the simulated returns, the equal portfolio weights are introduced to calculate the portfolio returns of the hypothetical two-asset portfolios. Given these weights, risk metrics for the simulated portfolio-return P&L distribution can be calculated and compared with the actually observed returns, in order to measure the accuracy of the forecast. For each portfolio, the forecasted VaR_{t+1}^{α} and $CVaR_{t+1}^{\alpha}$ for any confidence level α can be derived from the P&L of the simulated returns.

Lastly, the specified risk metrics are back-tested. Since, in comparison with the $CVaR^{\alpha}$, the VaR^{α} is known to be elicitable, different procedures have to be applied.[5] For the VaR_{t+1}^{α}, violation and independence-based tests are carried out. In order to conduct these tests, the VaR_{t+1}^{α} values from the AR-GARCH-EVT-Copula model and the classic historical simulation and variance-covariance model and are compared to the actual return series, so as to calculate so-called hit sequences (or "violations"), I_{t+1}. These sequences represent the model violations, namely the negative exceedance of realized returns over risk-metric forecasts:

$$I_{t+1}(\alpha) = \begin{cases} 1, if\ r_{t+1} < -VaR_{t+1}^{\alpha} \\ 0, if\ r_{t+1} > -VaR_{t+1}^{\alpha} \end{cases} \tag{14}$$

Firstly, a binominal test is applied to $I_{t+1}(\alpha)$. The abovementioned hit sequence should be a Bernoulli-distributed random variable with probability (α) and the number of observations for which risk forecasts are calculated (n):

$$I_{t+1}(\alpha) \sim\ B(n, \alpha) \tag{15}$$

Additionally, the Kupiec test is conducted. In order to conduct statistical inference on the specified distributional property of the hit sequence and its accuracy, the test statistic $K \sim \chi(1)$ is calculated in order to

[5] Accordingly, the approaches to back-test the $CVaR^{\alpha}$ are still subject to debate. See Nolde and Ziegel (2017) and Acerbi and Szekely (2017) for a detailed discussion.

conduct a two-sided test of the null hypothesis, regarding whether the hit sequence follows the specified distribution (Kupiec, 1995):

$$K = -2\ln[(1-p)^{n-m}p^m] + 2\ln\left[(1-{}^{m}\!/_{n})^{n-m}({}^{m}\!/_{n})^m\right] \quad (16)$$

In Equation (16), p denotes the assumed probability of occurrence, or α respectively, m the number of hits of the model and n the number of tests. Thus, the methodologies outlined above test whether the AR-GARCH-EVT-Copula model or historical simulation and variance-covariance provide a statistically sound modelling of the hit sequence for the VaR_{t+1}^{α} forecasts.

Additionally, the independence-based test of Christoffersen (1998) is applied. In contrast to the violation-based Bernoulli and Kupiec tests, this procedure not only measures the number of hits, but also their occurrence across time. Since the null hypotheses address specific properties of independence like exceedances not clustering, or loss quantiles not being autocorrelated, independence tests are more relevant for deciding whether the corresponding model is superior. Therefore, the null hypothesis states that the occurrence of violations $I_{t+1} = 1$ cannot be described by a first-order Markov Chain:

$$P(I_{t+1} = 0|I_t = 0) = P(I_{t+1} = 0|I_t = 1) = 1 - \alpha \quad (17)$$

For the $CVaR_{t+1}^{\alpha}$ on the other hand, a zero mean test is conducted, as proposed by McNeil et al. (2005). The test essentially assesses whether the excess loss component, given that a hit of the VaR_{t+1}^{α} occurred ($I_{t+1} = 1$), has a mean of zero. The procedure can be interpreted as a standard t test under the assumption of i.i.d.:

$$S = (r_{t+1} - CVaR_{t+1}^{\alpha} | I_{t+1} = 1) \qquad (18)$$

Here, the statistic S is expected to have a zero mean (under the null hypothesis), implying that the $CVaR_{t+1}^{\alpha}$ is under- and overestimating the tail risk for the next day to an exactly similar extent, if the VaR_{t+1}^{α} forecast generates a hit. A violation to the null hypothesis of a mean of zero showed a divergence from this assumption and thus structural under- or overestimation of the risk exposure in the tail of the return simulations. Since the present study is particularly interested in extreme risk and tail-risk estimation of coherent measurements in line with the axioms of Artzner et al. (1999), the analysis of the $CVaR_{t+1}^{\alpha}$ is of greater interest than the analysis of the VaR_{t+1}^{α}. Nonetheless, since the VaR_{t+1}^{α} is a widely used measurement in the banking industry for example, its importance for market participants is obvious.

3.4 Data and descriptive statistics

The data covers daily log return observations derived from total return data for securitized real estate, equity and debt indices between January 4th 1999 and July 31st 2019.[6] Due to data availability issues for some indices, we restricted our whole sample to start in January 1999. The inclusion of a trading day depends on the opening of all three asset class markets in the respective country. Thus, the number of observations can differ across countries due to differing public holidays or other specific distractions and market closure (such as 9/11 in the US). However, this heterogeneity does not affect the results, since the test statistics themselves are dependent on the number of observations. Time-series are denominated in the respective local currency in order to rule out any effects due to currency risk.

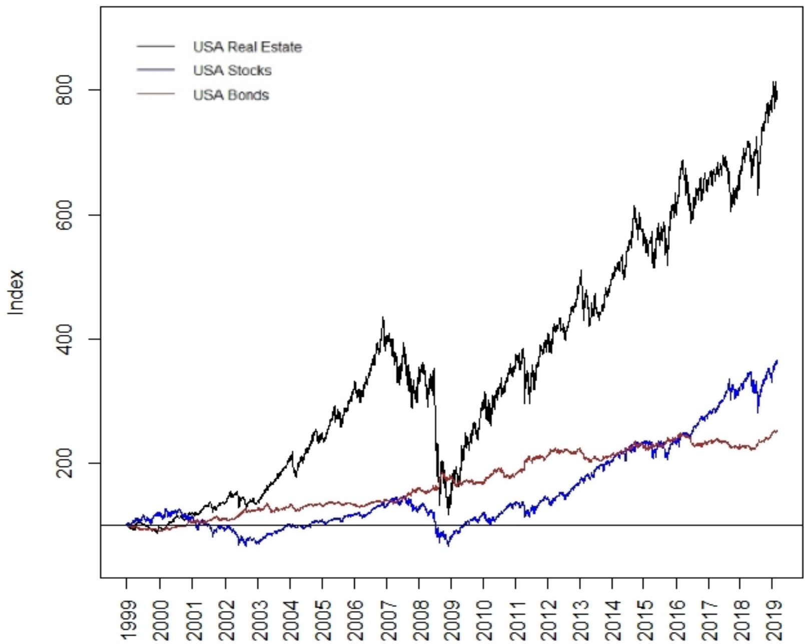
The included markets are Australia, France, Germany, Japan, the UK and the US. These nations were chosen, since they represent the class of mature securitized real estate markets (e.g. as proposed by Liow, 2008). For this study however, Hong Kong and Singapore were excluded, because these countries do not provide a debt index of sufficient length. This sufficiency is defined as a time span which covers several prominent

[6] Fritz and Oertel (2021) originally published their paper based on daily log returns derived from prices instead of total returns. Results and implications hardly differ.

critical market phases (most importantly the GFC in 2008 and the Dot-com bubble in the late 1990s). Inclusion of these is important since the risk model is supposed to be tested and stressed through multiple periods of intense downturns. Cross-country dependencies were excluded for two reasons: Firstly, a clear market separation is supposed to be isolated to identify potential differences across national borders. Secondly, the idea to use data per country provokes simultaneous heavy downturns, since national markets are heavily integrated, causing additional stress on the risk forecasting.

As a first insight, the following figure shows the performance of our three main asset classes for the US market over the full sample period (see Figure 3):

Figure 3: Cumulated return series for real estate, stocks and bonds (US)



Notes: The graphic shows the cumulated log returns of the real estate, stocks and bond series for the USA. Each series is starting at 100. Figures and graphics for the other countries in the sample are available upon request. Due to limited space we do not present those graphics here. Further descriptive statistics concerning these countries are showcased later on.

The variety of markets is introduced for two reasons. Firstly, a larger number of markets and thus dependencies of securitized real estate and the two other asset classes is intended to ensure robustness of the model. A market study on a single market appears to be insufficient to derive valid statements about global market behavior and in order to proof that our model may be eligible for more than just one specific market. Secondly, country specifics may be of interest, since the abovementioned crises are expected to be globally heterogeneous (e.g. especially extreme losses during the GFC in the US). For the securitized real estate, EPRA NAREIT All Equity indices are used. The equity data sets are the leading national indices, namely the ASX100 (Australia), CAC40 (France), DAX30 (Germany), Nikkei (Japan), FTSE100 (UK) and the S&P500 (US). The debt returns are from the countries' government bonds with ten-year maturity. For the outlined dataset constituents, the following table summarizes the descriptive statistics (see Table 9):

Table 9: Descriptive statistics for country sample

	AUS n = 5028	**GER** n = 5132	**FRA** n = 5134	**JAP** n = 4838	**UK** n= 5159	**USA** n = 5080
Panel A: Real Estate						
Mean	0.73	3.20	6.99	5.73	2.14	5.38
Std. Dev.	20.43	24.94	20.19	30.82	20.82	27.73
25th percentile	-75.24	-82.19	-76.66	-90.46	-73.78	-75.02
75th percentile	354.72	530.51	439.45	1100.99	349.26	426.55
Skewness	-0.70	0.05	-0.07	0.18	-0.57	-0.22
Kurtosis	29.22	8.70	4.38	4.84	10.35	22.25
JB	179309	16215	4111	4755	23323	104896
Q(16)	269	66	57	96	43	407
$Q^2(16)$	3651	3823	3894	3944	3892	11249

Panel B: Stocks						
Mean	12.66	4.48	1.68	2.32	1.24	4.48
Std. Dev.	25.03	23.36	22.64	24.05	18.53	19.14
25th percentile	-83.39	-81.69	-81.18	-83.84	-74.63	-70.88
75th percentile	744.81	557.95	532.42	696.63	338.85	330.53
Skewness	0.37	-0.06	-0.02	-0.36	-0.16	-0.25
Kurtosis	10.57	4.52	5.09	6.18	6.10	8.08
JB	23516	4385	5546	7811	8032	13901
Q(16)	269	66	57	96	43	407
$Q^2(16)$	3651	3823	3894	3944	3892	11249
Panel C: Bonds						
Mean	1.40	2.45	2.53	1.81	1.91	1.01
Std. Dev.	7.71	5.54	5.63	3.96	6.19	7.50
25th percentile	-49.78	-37.72	-38.03	-23.09	-42.73	-50.29
75th percentile	112.71	75.94	74.27	38.14	85.04	109.67
Skewness	-0.14	-0.21	-0.23	-0.56	0.04	-0.05
Kurtosis	2.95	1.75	2.59	6.90	1.86	2.52
JB	1843	695	1485	9854	746	1348
Q(16)	59	63	70	80	58	60
$Q^2(16)$	772	677	1465	3786	575	1040

Notes: The table presents descriptive statistics of the three asset return series for each of the six countries in our sample. The figures for mean, standard deviation, the 25th as well as the 75th percentile are annualized under the assumption of 252 (trading) days per year and reported in percent. For Jarque-Bera, Q(16) and $Q^2(16)$, we state the individual test statistic.

The dispersion of the data is highest for four of the real estate time series (except for Australia and France). Skewness is mainly close to zero and positive for each return series, showing skew towards the right. Hence, the observed skew justifies the application of the skewed-t errors for the univariate models. The large kurtosis of all returns indicates leptokurtic distributions. In addition to the statements about return series volatility for securitized real estate, the minima reveal the largest downturns for the specified asset class (except for France). In the context of risk management and metric forecasting, these extreme values are of particular interest, since these returns are the most likely observations to cause violations of the price risk forecast of the VaR and the CVaR in comparison to the real return.

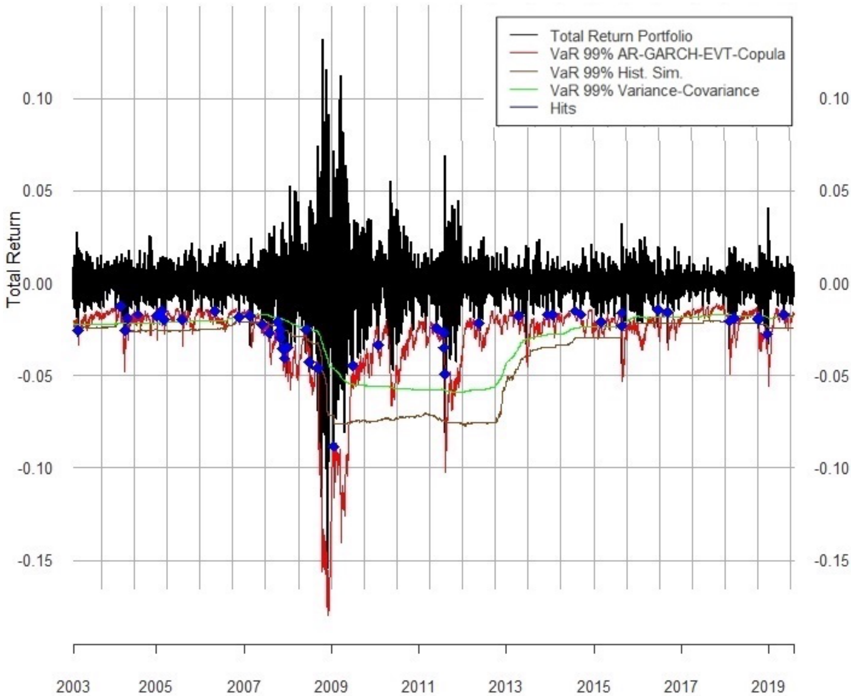
The descriptive statistics also reveal evidence of other statistical issues mentioned in the literature review. These indicate a compelling need for the application of the AR-GARCH-EVT-Copula approach. This applies especially to the securitized real estate data. Additionally, the Jarque-Bera tests yield very strong empirical evidence of the violation of normality for each time series of the dataset. In addition, serial correlation can be detected due to the findings of the Q(16) and $Q^2(16)$ statistics. Thus, the application of statistical procedures to account for these issues is needed to ensure unbiased univariate and multivariate modelling.

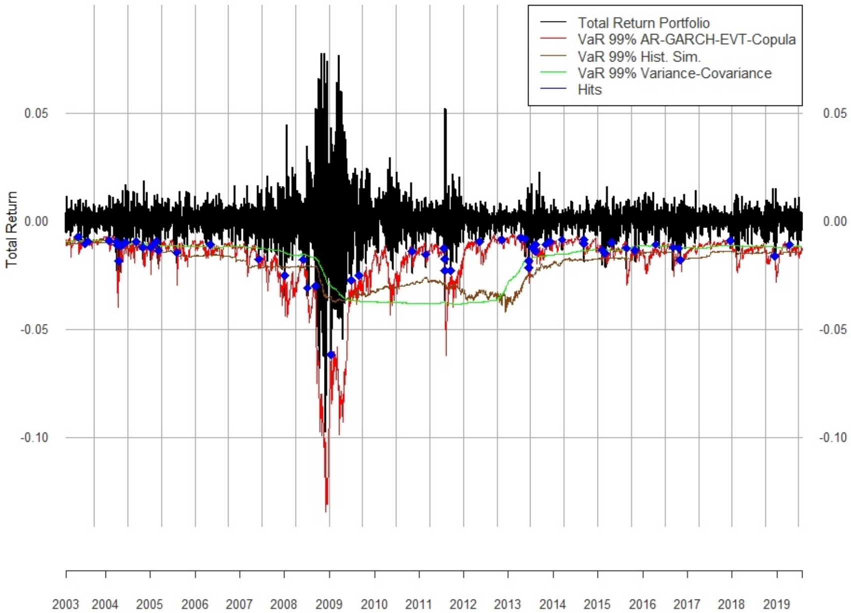
3.5 Empirical results

The empirical analysis covers the results of the back testing for the AR-GARCH-EVT-Copula and the two benchmark methodologies, namely variance-covariance and historical simulation for the specified return series.[7] For each approach, risk forecasts for the equally weighted portfolio are calculated and compared to the actual portfolio return. Graphically, the figures below display the VaR ($\alpha = 0.99$) from the AR-GARCH-EVT-Copula model as well as both benchmark methodologies for both portfolios from the US (see Figure 4):

[7] Implementing the AR-GARCH-EVT-Copula model leads to a load of typical estimates. Since the estimates change over time due to the usage of rolling windows, those estimates can only be illustrated in figures. These figures for AR-GARCH estimates, scale and shape as well as copula parameters are available upon request.

Figure 4: VaR (α=0.99) estimates for Real Estate - Stocks & Real Estate - Bonds portfolio (US)





From the graphical representation of the observed returns (black lines) and the risk forecasts, the primary difference between the benchmark models and the AR-GARCH-EVT-Copula model is the relative responsiveness of the latter approach to differing levels of market volatility, and especially extreme losses (as displayed by the red lines). In comparison, the benchmark methods do not provide this flexibility and react to periods of increasing volatility and heavy downturns (e.g. the GFC) and also to decreased volatility too reluctantly and late (e.g. the brown and green graphs both respond in early 2013 by indicating significantly lower risk forecasts). This finding applies to both portfolio scenarios alike as well as across all countries in our sample.

One explanation may be the increased correlation between asset classes especially in downturn markets (Case et al., 2012), which can cause extreme simultaneous asset losses, contradicting heavily with the assumption of constant correlation of the benchmark methodologies. Accordingly, refitting the dependence structure appears to be a key element of appropriate risk metric forecasting, due to breakdowns in correlation patterns during increased volatility.

Beside optical inspection, numerical measures provide deeper insight into the model accuracy. By back-testing the methodologies through the data sample, the absolute and relative number of hits, the corresponding Bernoulli as well as the Kupiec statistics are displayed for the $VaR_{t+1}^{\propto}$ (see Table 10). Additionally, for the $CVaR_{t+1}^{\propto}$, the zero mean test results are shown on Table 11.[8]

[8] The results for the portfolios containing stocks and bonds are available upon request.

Table 10: Empirical results for VaR forecasts

VaR								
Country	Port-folio	Risk Metric	Model	Hits	Rel. Hits	Bern-oulli	Kupiec	Christof-fersen
Australia (n=4026)	Real Estate - Stocks	$VaR_{t+1}^{0.01}$	Var.-Cov.	96	2.38	0.00	0.00	0.00
			Hist. Sim.	59	1.47	0.53	0.55	0.09
			AR-GARCH-EVT-Copula	42	1.04	75.12	78.43	20.77
		$VaR_{t+1}^{0.05}$	Var.-Cov.	215	5.34	32.88	34.30	0.44
			Hist. Sim.	217	5.39	26.24	27.81	0.65
			AR-GARCH-EVT-Copula	225	5.59	8.91	9.23	23.88
	Real Estate - Bonds	$VaR_{t+1}^{0.01}$	Var.-Cov.	93	2.31	0.00	0.00	0.00
			Hist. Sim.	57	1.42	1.10	1.26	0.00
			AR-GARCH-EVT-Copula	56	1.39	1.71	1.86	0.92
		$VaR_{t+1}^{0.05}$	Var.-Cov.	214	5.32	34.73	36.31	0.00
			Hist. Sim.	233	5.79	2.49	2.52	0.00
			AR-GARCH-EVT-Copula	217	5.39	26.21	26.20	1.74
France (n=4134)	Real Estate - Stocks	$VaR_{t+1}^{0.01}$	Var.-Cov.	90	2.18	0.00	0.00	0.00
			Hist. Sim.	41	0.99	100.00	95.76	72.87
			AR-GARCH-EVT-Copula	53	1.28	7.19	8.07	9.10
		$VaR_{t+1}^{0.05}$	Var.-Cov.	190	4.60	25.33	22.72	0.00
			Hist. Sim.	221	5.35	30.06	30.92	0.00
			AR-GARCH-	186	4.50	15.32	13.50	0.12

			EVT-Copula					
	Real Estate - Bonds	$VaR_{t+1}^{0.01}$	Var.-Cov.	89	2.15	0.00	0.00	0.00
			Hist. Sim.	42	1.02	87.57	91.80	3.19
			AR-GARCH-EVT-Copula	47	1.14	34.87	38.66	58.25
		$VaR_{t+1}^{0.05}$	Var.-Cov.	227	5.49	15.33	15.36	0.00
			Hist. Sim.	199	4.81	61.73	58.04	0.00
			AR-GARCH-EVT-Copula	220	5.32	33.53	34.74	0.09
Germany (n=4132)	Real Estate - Stocks	$VaR_{t+1}^{0.01}$	Var.-Cov.	83	2.01	0.00	0.00	0.00
			Hist. Sim.	43	1.04	75.44	79.41	0.03
			AR-GARCH-EVT-Copula	39	0.94	81.43	71.42	64.48
		$VaR_{t+1}^{0.05}$	Var.-Cov.	184	4.45	10.83	10.05	0.00
			Hist. Sim.	200	4.84	66.84	63.58	0.00
			AR-GARCH-EVT-Copula	197	4.77	52.05	49.00	9.13
	Real Estate - Bonds	$VaR_{t+1}^{0.01}$	Var.-Cov.	85	2.06	0.00	0.00	0.00
			Hist. Sim.	44	1.06	63.89	67.84	0.04
			AR-GARCH-EVT-Copula	47	1.14	34.84	38.48	4.00
		$VaR_{t+1}^{0.05}$	Var.-Cov.	183	4.43	9.34	8.61	0.00
			Hist. Sim.	222	5.37	26.85	27.72	0.00
			AR-GARCH-EVT-Copula	187	4.53	17.48	15.54	0.15
		$VaR_{t+1}^{0.01}$	Var.-Cov.	71	1.85	0.00	0.00	0.00

Japan (n=3838)	Real Estate - Stocks		Hist. Sim.	50	1.30	6.19	7.17	1.89
			AR-GARCH-EVT-Copula	42	1.09	51.66	56.29	3.29
		$VaR_{t+1}^{0.05}$	Var.-Cov.	161	4.19	2.16	1.87	0.00
			Hist. Sim.	170	4.43	11.11	9.84	0.00
			AR-GARCH-EVT-Copula	191	4.98	100.00	94.68	52.40
	Real Estate - Bonds	$VaR_{t+1}^{0.01}$	Var.-Cov.	51	1.33	5.07	5.13	0.27
			Hist. Sim.	30	0.78	19.38	15.75	18.20
			AR-GARCH-EVT-Copula	34	0.89	56.93	46.87	8.74
		$VaR_{t+1}^{0.05}$	Var.-Cov.	148	3.86	0.09	0.07	0.00
			Hist. Sim.	127	3.31	0.00	0.00	0.00
			AR-GARCH-EVT-Copula	172	4.48	14.85	13.38	0.30
United Kingdom (n=4159)	Real Estate - Stocks	$VaR_{t+1}^{0.01}$	Var.-Cov.	91	2.19	0.00	0.00	0.00
			Hist. Sim.	52	1.25	11.81	11.84	11.62
			AR-GARCH-EVT-Copula	46	1.11	48.24	49.92	21.78
		$VaR_{t+1}^{0.05}$	Var.-Cov.	203	4.88	74.89	72.37	0.00
			Hist. Sim.	212	5.10	77.59	77.39	0.00
			AR-GARCH-EVT-Copula	213	5.12	72.19	72.04	75.15
	Real Estate - Bonds	$VaR_{t+1}^{0.01}$	Var.-Cov.	105	2.52	0.00	0.00	0.00
			Hist. Sim.	64	1.54	0.10	0.12	0.01
			AR-GARCH-	53	1.27	8.56	8.81	21.63

			EVT-Copula					
		$VaR_{t+1}^{0.05}$	Var.-Cov.	228	5.48	15.47	15.98	0.00
			Hist. Sim.	231	5.55	10.17	10.69	0.00
			AR-GARCH-EVT-Copula	220	5.29	39.31	39.55	0.20
USA (n=4080)	Real Estate - Stocks	$VaR_{t+1}^{0.01}$	Var.-Cov.	105	2.57	0.00	0.00	0.00
			Hist. Sim.	60	1.47	0.44	0.48	0.08
			AR-GARCH-EVT-Copula	47	1.15	30.67	34.08	19.12
		$VaR_{t+1}^{0.05}$	Var.-Cov.	206	5.05	88.57	88.59	0.00
			Hist. Sim.	221	5.42	22.19	22.80	0.00
			AR-GARCH-EVT-Copula	212	5.20	56.54	56.79	2.95
	Real Estate - Bonds	$VaR_{t+1}^{0.01}$	Var.-Cov.	114	2.79	0.00	0.00	0.00
			Hist. Sim.	66	1.62	0.03	0.03	0.00
			AR-GARCH-EVT-Copula	57	1.40	1.44	1.62	2.86
		$VaR_{t+1}^{0.05}$	Var.-Cov.	267	6.54	0.00	0.00	0.00
			Hist. Sim.	263	6.45	0.00	0.00	0.00
			AR-GARCH-EVT-Copula	248	6.08	0.20	0.22	0.00

Notes: The number of observations (n) equals to the number of total observations for each country less the burn-in sample of 1000 observations. Relative hits are calculated as the number of actual hits divided by total observations. Relative hits as well as p-values are given in percent. Null hypotheses for Bernoulli, Kupiec and Christoffersen tests are described in detail in the methodology section (Formulas 15, 16 & 17). In short, for the Bernoulli test, the null hypothesis is that the results do not differ significantly from the expected number of hits. The null hypothesis for the Kupiec test states that the observed failure rate is equal to the failure rate suggested by the confidence interval. Finally, the null hypothesis of the Christoffersen test describes the correct number of exceedances and the independence of failures. Further results, back-tests and graphics for all Stocks-Bonds pairs are available upon request.

For the VaR_{t+1}^{α}, the results provide numerical proof of improvements in the violation-based figures at both levels of significance. The absolute as well as the relative number of hits provide some initial but rather sparse insight into the quality of the proposed model. For example, for the $VaR_{t+1}^{0.01}$, the expected number of relative hits should be exactly one percent, as it is the case for the Australian real estate & stocks portfolio according to the AR-GARCH-EVT-Copula model. Additionally, one would rather underestimate the number of violations than overestimate it. Tendencies to underestimate the number of violations do not involve such grave consequences for portfolio holders as overestimation. For the $VaR_{t+1}^{0.05}$, the p-values suggest partial missing improvements of the AR-GARCH-EVT-Copula model, for example for the Real Estate & Bond portfolio in Japan. However, the majority of the results confirm the superiority of the model in comparison to the benchmarks. Even more significant confirmation can be found for the $VaR_{t+1}^{0.01}$. Here, all countries and portfolios show improved violation-based figures for the Bernoulli and Kupiec tests. Thus, it can be stated that the AR-GARCH-EVT-Copula model is especially feasible for tail-risk estimation, since the model outperforms the benchmark more clearly with an increased level of confidence.

With regard to the Christoffersen test and thus the temporal dimension of the model hits, the p-values show temporal independence of the AR-GARCH-EVT-Copula model. For $VaR_{t+1}^{0.01}$, the unanimous approval of independence for the AR-GARCH-EVT-Copula model is supported, whereas the benchmark models fail to generate hits without temporal dependence at the one percent level. The results of the $VaR_{t+1}^{0.05}$, however, contain some p-values which indicate temporal dependence, especially for the real estate & bond portfolios (e.g., in the US and the UK). It becomes apparent that the benchmark models produce hits with a clear timely pattern.

In sum, the violation-based and the independence tests yield similar results, in favour of the AR-GARCH-EVT-Copula model. Turning to the $CVaR_{t+1}^{\propto}$, the following table summarizes the back-testing and the especially the zero mean test results for the respective CVaR and country sample (see Table 11):

Table 11: Empirical back testing result for the CVaR

CVaR						
				Hits	Rel. Hits	Zero Mean Test
Australia (n=4026)	Real Estate - Stocks	$CVaR_{t+1}^{0.01}$	Var.-Cov.	67	1.66	0.00
			Hist. Sim.	21	0.52	54.11
			AR-GARCH-EVT-Copula	13	0.32	99.78
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	135	3.35	0.00
			Hist. Sim.	94	2.33	1.44
			AR-GARCH-EVT-Copula	64	1.59	100.00
	Real Estate - Bonds	$CVaR_{t+1}^{0.01}$	Var.-Cov.	66	1.64	0.00
			Hist. Sim.	19	0.47	74.80
			AR-GARCH-EVT-Copula	12	0.30	99.71
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	135	3.35	0.00
			Hist. Sim.	96	2.38	7.55
			AR-GARCH-EVT-Copula	65	1.61	100.00
France (n=4134)	Real Estate - Stocks	$CVaR_{t+1}^{0.01}$	Var.-Cov.	53	1.28	0.00
			Hist. Sim.	19	0.46	13.57
			AR-GARCH-EVT-Copula	12	0.29	98.56
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	121	2.93	0.00
			Hist. Sim.	74	1.79	11.68
			AR-GARCH-EVT-Copula	68	1.64	100.00
	Real Estate - Bonds	$CVaR_{t+1}^{0.01}$	Var.-Cov.	57	1.38	0.00
			Hist. Sim.	20	0.48	11.64
			AR-GARCH-EVT-Copula	16	0.39	98.79
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	137	3.31	0.00
			Hist. Sim.	71	1.72	70.32
			AR-GARCH-EVT-Copula	67	1.62	100.00
Germany (n=4132)	Real Estate	$CVaR_{t+1}^{0.01}$	Var.-Cov.	56	1.36	0.00
			Hist. Sim.	20	0.48	10.88

	- Stocks		AR-GARCH-EVT-Copula	6	0.15	100.00
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	115	2.78	0.00
			Hist. Sim.	80	1.94	19.60
			AR-GARCH-EVT-Copula	74	1.79	100.00
	Real Estate - Bonds	$CVaR_{t+1}^{0.01}$	Var.-Cov.	64	1.55	0.00
			Hist. Sim.	21	0.51	22.58
			AR-GARCH-EVT-Copula	12	0.29	99.95
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	114	2.76	0.00
			Hist. Sim.	75	1.82	79.95
			AR-GARCH-EVT-Copula	72	1.74	100.00
Japan (n=3838)	Real Estate - Stocks	$CVaR_{t+1}^{0.01}$	Var.-Cov.	53	1.38	0.00
			Hist. Sim.	19	0.50	25.41
			AR-GARCH-EVT-Copula	18	0.47	42.66
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	88	2.29	0.00
			Hist. Sim.	72	1.88	6.32
			AR-GARCH-EVT-Copula	54	1.41	100.00
	Real Estate - Bonds	$CVaR_{t+1}^{0.01}$	Var.-Cov.	53	1.38	0.00
			Hist. Sim.	19	0.50	25.49
			AR-GARCH-EVT-Copula	17	0.44	43.40
		$CVaR_{t+1}^{0.05}$	Variance-Covariance	88	2.29	0.00
			Hist. Sim.	72	1.88	6.32
			AR-GARCH-EVT-Copula	55	1.43	100.00
United Kingdom (n=4159)	Real Estate - Stocks	$CVaR_{t+1}^{0.01}$	Var.-Cov.	66	1.59	0.00
			Hist. Sim.	27	0.65	3.26
			AR-GARCH-EVT-Copula	13	0.31	63.13
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	123	2.96	0.00
			Hist. Sim.	85	2.04	3.29
			AR-GARCH-EVT-Copula	54	1.30	100.00
	Real Estate	$CVaR_{t+1}^{0.01}$	Var.-Cov.	72	1.73	0.00
			Hist. Sim.	28	0.67	11.44

	- Bonds		AR-GARCH-EVT-Copula	15	0.36	82.00
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	141	3.39	0.00
			Hist. Sim.	92	2.21	3.59
			AR-GARCH-EVT-Copula	52	1.25	100.00
USA (n=4080)	Real Estate - Stocks	$CVaR_{t+1}^{0.01}$	Var.-Cov.	76	1.86	0.00
			Hist. Sim.	36	0.88	1.02
			AR-GARCH-EVT-Copula	16	0.39	91.60
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	128	3.14	0.00
			Hist. Sim.	94	2.30	5.22
			AR-GARCH-EVT-Copula	66	1.62	100.00
	Real Estate - Bonds	$CVaR_{t+1}^{0.01}$	Var.-Cov.	82	2.01	0.00
			Hist. Sim.	27	0.66	20.82
			AR-GARCH-EVT-Copula	24	0.59	73.49
		$CVaR_{t+1}^{0.05}$	Var.-Cov.	157	3.85	0.00
			Hist. Sim.	92	2.25	15.24
			AR-GARCH-EVT-Copula	74	1.81	100.00

Notes: The number of observations (n) equals the number of total observations for each country less the burn-in sample of 1000 observations. Relative hits are calculated as the number of actual hits divided by total observations. Relative hits as well as p-values are given in percent. As a reminder, the null hypothesis for the zero mean test is that the excess conditional shortfall, is i.i.d. and has zero mean. See Formula 18 and methodology section for detailed information. Further results, back-tests and graphics for all Stocks-Bonds pairs are available upon request.

The results for the $CVaR_{t+1}^{\propto}$ indicate at both levels a clear superiority of the AR-GARCH-EVT-Copula approach in comparison to the benchmarks. The variance-covariance method produces hit sequences, which do not exhibit a mean of zero. The historical simulation approach shows a superior hit sequence compared to the variance-covariance method. Nonetheless, the AR-GARCH-EVT-Copula method shows the highest p-values across all markets and portfolios. Thus, the $CVaR_{t+1}^{\propto}$ results suggest outperformance of the benchmark at both levels of significance. The results can be interpreted as confirmation for structural under- and overestimation of the $CVaR_{t+1}^{\propto}$, if $VaR_{t+1}^{\propto}$ is violated. This does not apply to the variance-covariance method, and only partially to the historical simulation. Taking the graphical inspection into account, especially the heavy underestimation of losses during the GFC may cause these results for the benchmarks.

In sum, the results provide empirical evidence, both graphically and numerically, of an improved risk measurement of the AR-GARCH-EVT-Copula in comparison to the benchmark methodologies. In particular, the re-estimation of dependence patterns appears to be a key feature for correctly modelling its time-variance. Since the results also show greater improvements for risk measurements of the tail (e.g., the larger confirmation of the $VaR_{t+1}^{0.99}$ than the $VaR_{t+1}^{0.95}$), the dependence patterns of the analysed asset classes may also reveal a need to model non-linear relationships in contrast to the strictly linear correlation measurement. This applies especially to critical market phases, since the graphical inspection revealed heavy underestimation of the risk exposure in these periods (e.g., during the GFC). Since the $CVaR_{t+1}^{\alpha}$ represents a coherent risk measurement in accordance with Artzner et al. (1999), as proposed by Rockafellar and Uryasev (2000), the results imply the feasibility of the AR-GARCH-EVT-Copula model especially for the named figure.

Based upon the empirical results of the back-testing, questions arise regarding the underlying univariate and multivariate models. Since varying models are used for each rolling window of the AR-GARCH-EVT-Copula approach, a deeper look into the results for the autoregressive and dependence models for each portfolio may provide additional information. Therefore, Table 12 displays the results of the goodness of fit for the autoregressive models for each time series. More precisely, Table 12 reports the discrete distribution for the highest fit of each autoregressive order across the respective data series for the rolling windows (see Table 12):

Table 12: Results of the autoregressive modelling

AR Order	US Real Estate	UK Real Estate	DE Real Estate	FR Real Estate	AUS Real Estate	JP Real Estate	Sum
0	989	611	310	989	752	108	3759
1	967	1086	889	295	74	585	3896
2	251	387	385	511	255	255	2044
3	112	320	264	672	936	612	2916
4	469	636	598	420	401	858	3382
5	1292	1119	1686	1247	1608	1421	8373
	US Stocks	UK Stocks	DE Stocks	FR Stocks	AUS Stocks	JP Stocks	Sum
0	0	265	747	239	960	1066	3277
1	618	449	532	317	214	734	2864
2	455	289	108	350	141	293	1636
3	131	790	133	568	255	214	2091
4	434	932	146	363	1364	783	4022
5	2442	1434	2466	2297	1092	749	10480
	US Bonds	UK Bonds	DE Bonds	FR Bonds	AUS Bonds	JP Bonds	Sum
0	814	676	450	1219	460	606	4225
1	417	408	435	297	601	210	2368
2	838	469	1018	712	1370	128	4535
3	273	684	272	411	558	1113	3311
4	406	416	649	656	446	587	3160
5	1332	1506	1308	839	591	1195	6771

Notes: The table displays the number of occurrences of the highest fit for the respective autoregressive order by asset class and country as well as the sum across each row.

Firstly, the univariate results reveal the highest percentage of best fitting models for the autoregressive order of five (35.05% of the overall number of windows across all asset classes). The distribution across the remaining five orders yields homogenous results between 11.24% - 15.40% of the overall number of windows across all-time series.

Considering the cross-section of asset classes, a pattern can be observed for real estate and stocks. Firstly, the goodness of fit for the lower orders yields percentages in double figures, decreasing through the second and third order, regaining fit in the lags four and five.[9] For bonds in comparison, the autoregressive models in the middle of the tested orders are more accurate and reveal a significantly higher percentage of fits for order two. Thus, with regard to the autoregressive character of the data, bonds are the asset class, which behave more balanced across the autoregressive orders than its peers in the sample. For securitized real estate and stocks, 34.36%, or 43.00% respectively, of the overall windows are modelled best by an autoregressive model of order five. In comparison, only 27.78% of the bond windows are showing the highest fit for the longest autoregressive order. Since public equity positions such as securitized real estate and stocks are known for their long-lasting

[9] With an interesting outlier of US stocks, with a total number of zero times for the highest fit of order zero.

and heavy serial autocorrelation, the univariate results are in line with expectations based on the literature review above.

Within the asset classes, the results also reveal a certain extent of heterogeneity across the markets. For securitized real estate, Germany and Japan show extremely low number of occurrences for order zero. In contrast, the time series for Germany also displays the highest number for the longest order. Another notable unusualness within the securitized real estate data is the extremely low number of order one models for the Australian time series. In addition to the low number of order zero models, Australia appears to be a market with more long-lasting autoregressive effects. The same applies to Japan. The US and the UK on the other hand are markets with more occurrences (1,956 and 1,697) of short autoregressive effects (zero and one).

For the other asset classes, a surprising finding is the missing occurrence for order zero. Thus, the US stock time series entails autoregressive effects for every other window. In fact, the data for the US stocks time series is heavily long-lasting autoregressive (2,442 observations for the highest order). For the bond data, the results reveal the highest occurrences for the middle orders (especially two and three), as outlined above.

Nonetheless, from a methodological point of view, an extension to even higher autoregressive orders could be thinkable for further model improvements especially for securitized real estate and stocks. With regard to the implications for the price risk forecasting model of interest, the general necessity to allow for individual order selection based on the respective goodness of fit can be extracted from the results, since the asset classes and markets of the study show largely differing results and thus individual specifics.

Based on these univariate findings, the multivariate results are assessed. In particular, the question is, what types of copulae are providing the highest fit overall and for which specific portfolios or markets. From the chosen type of copula alone, insights about the symmetry of the co-movements of the portfolio constituents can be derived. Therefore, Table 13 summarizes the discrete distribution of the copulae with the highest fit among the tested ones for each rolling window across all asset classes and countries.

The major finding of the multivariate modelling is the clear dominance of the Student-t-Copula. Out of the total number of windows, more than 51.53% of the dependence models reveal the highest goodness of fit for the named copula, implying a symmetric but existing tail dependence. This finding, however, is largely driven by the dependence of securitized real estate and stocks, since 63.23% of the Student-t-Copula models apply the named portfolio constituents. This finding is in line with the expectations based on the literature review, because previous studies have repeatedly shown this simultaneous market behaviour of securitized real estate and stocks.

Interestingly, out of the entire data set the second-best fitting copula family is the Gaussian. 7.71% of the relationships are modelled by a Gaussian copula, which is contradicting the assumption of non-normality of the joint marginal. This finding, however, is largely impacted by the results of the US real estate – bond portfolio, on its own already accounting for 2,240 out of the 3,723 total windows, which are modelled by the Gaussian copula. This finding can be interpreted as a sign for no tail dependence of the specified portfolio constituents. Other markets do not support the application of the Gaussian copula, which reduces the finding to a market specific phenomenon.

Thirdly, the survival BB1 copula models 3,191 dependencies out of the sample. The named copula type also shows an entirely skewed distribution across the portfolio constituents, since only real estate – stocks portfolios are displayed. This 180-degree rotated copula, with lower tail dependence but higher variance in the empirical density in the named tail reveals the potential for simultaneous but also more less dense realizations in the tail. Thus, portfolios tend to show higher variance in the tail observations, but still existing asymmetric dependence.

Furthermore, the rotated copulae (both, 90 and 270°) are only used by securitized real estate and bonds. This finding is highly important for securitized real estate investors, who seek multi-asset diversification and correct portfolio modelling. Generally, these rotated copulae symbolize opposing price movements, because they are used to model data, which explicitly shows positive (negative) returns of one asset, when the other asset moves in the opposite direction. Thus, this finding implies the strict necessity to apply the specified copulae, when securitized real estate and bond positions are gathered in a portfolio to fully capture the nature of the data. This finding is especially of interest from a strategic risk management point of view, because these occurrences imply the possibility to hedge price risk movements of the named asset classes. Whereas this finding is not new to the real estate literature, the empirical dependence modelling by using rotated copulae has not been extensively studied

Table 13: Results for the copulae estimation

Copula Type	Australia		France		Germany		Japan		UK		US		Sum		
	RE-Stocks	RE-Bonds	RE-Stocks	RE-Bonds	RE-Stocks	RE-Bonds	RE-Stocks	RE-Bonds	RE-Stocks	RE-Bonds	RE-Stocks	RE-Bonds	all	RE-Stocks	RE-Bonds
Gaussian	5	227	202	92	301	0	436	55	68	169	0	0	1555	1012	543
Student t	1875	656	1590	2065	2587	1995	1866	1988	2435	2571	3960	2178	25766	14313	11453
Clayton	0	7	0	73	0	0	0	0	0	0	0	12	92	0	92
Gumbel	0	293	0	0	0	0	0	0	0	0	0	0	293	0	293
Frank	32	0	0	11	153	0	0	1017	0	133	0	0	1346	185	1161
BB1	1402	0	293	0	182	0	440	0	4	0	59	0	2380	2380	0
BB7	0	128	403	0	0	0	0	0	0	0	0	0	531	403	128
BB8	0	0	0	0	2	0	0	0	0	0	0	0	2	2	0
survival Clayton	0	345	0	0	0	0	0	0	0	0	0	0	345	0	345
survival Gumbel	84	6	416	0	374	0	0	0	777	0	0	0	1657	1651	6
survival Joe	0	98	0	0	0	11	0	0	0	0	0	291	400	0	400
survival BB1	226	0	1086	0	460	0	1096	0	875	0	61	0	3804	3804	0
survival BB7	0	39	144	0	20	0	0	0	0	0	0	0	203	164	39
survival BB8	402	0	0	0	53	0	0	0	0	0	0	182	637	455	182
rotated Clayton (90°)	0	0	0	470	0	0	0	50	0	84	0	0	604	0	604
rotated Gumbel (90°)	0	475	0	167	0	147	0	57	0	133	0	1069	2048	0	2048
rotated Joe (90°)	0	300	0	1	0	4	0	0	0	0	0	121	426	0	426
rotated BB1(90°)	0	0	0	371	0	463	0	162	0	0	0	198	1194	0	1194
rotated BB7(90°)	0	5	0	118	0	29	0	0	0	0	0	29	181	0	181
rotated BB8 (90°)	0	1424	0	21	0	192	0	175	0	191	0	0	2003	0	2003
rotated Clayton (270°)	0	4	0	258	0	609	0	116	0	39	0	0	1026	0	1026
rotated Gubel (270°)	0	0	0	31	0	0	0	152	0	87	0	0	270	0	270
rotated Joe (270°)	0	19	0	14	0	0	0	0	0	0	0	0	33	0	33
rotated BB1 (270°)	0	0	0	134	0	541	0	0	0	746	0	0	1421	0	1421
rotated BB7 (270°)	0	0	0	308	0	141	0	0	0	6	0	0	455	0	455
rotated BB8 (270°)	0	0	0	0	0	0	0	66	0	0	0	0	66	0	66

3.6 Conclusion

The present study contributes to the existing body of real estate literature by extending the stream of publications on copula dependence modelling with the empirical study, not only of the parameters of the dependence structures and fit assessment, but by the actual application of nonlinear dependence modelling to price risk metric forecasting. Therefore, the dependence modelling is extended and enriched by univariate modelling and the Monte Carlo simulation, based on copula dependence using the so-called AR-GARCH-EVT-Copula approach. After describing the conceptual construction of the risk model, the empirical study reveals improvements in the specified methodology across different risk metrics and levels of significance.

The study also reveals that the VaR based on AR-GARCH-EVT-Copula provides better one-day-ahead estimates, compared to the traditional VaR/CVaR estimation methods (variance-covariance and historical simulation). The results of simple violation ratios and additional test statistics like Kupiec, Christoffersen and zero mean for our model at different significance levels, were within the range of a superior estimation model. A detailed decomposition of the model revealed the necessity for univariate modelling of high autoregressive orders. Additionally, the multivariate analysis showed the predominant symmetric and negative tail dependence mainly for securitized real estate and stocks, but also towards bonds. In addition, the results of the multivariate modelling of securitized real estate and bonds showed evidence to incorporate rotated copulae at both levels of rotation to fully capture the dependence correctly. A limitation on classic elliptical and Archimedean copulae does not provide the necessary range of dependence structures.

The practical implications are the viable implementation of the presented approach and the replacement of variance-covariance or historical simulation methods for the specified asset classes. Especially in periods of extreme volatility and accordingly heavy negative daily returns, investors can benefit from improved risk metric forecasts in comparison to classic models. VaR and CVaR have also been widely used as risk measures by many financial institutions and regulators, such as the Basel Committee on Banking Supervision. Hence, our results also provide further insight into the correct approach of estimating these risk measures for those market participants.

Future research may also incorporate cross-country dependencies, which were not studied in this article. This could be especially useful for investors who diversify their portfolios across geographical borders. An extensive focus on securitized real estate could be thinkable, by analyzing portfolios of indirect property investment indices from different countries. Furthermore, an extension towards different types of equity securities could be beneficial, such as small or medium cap or debt positions like high or low yield. Additionally, it should be mentioned that only mature securitized real estate markets were analyzed in the present study. An extension to less mature markets can be useful so as to compare the feasibility of the model between mature and immature markets, although potential data limitations may occur. Potentially interesting studies would include those on the underlying copulae and a comparative study of them. Since the present approach uses switching copulae for each window, the fixation of a copula type and subsequent simulation out of each copula across the entire sample may be beneficial in detecting differences across varying dependence models. In this context, the investigation of a true time-varying parameter model with Bayesian updates could be of interest. Lastly, the option to investigate the ability to use the AR-GARCH-EVT-Copula approach as portfolio optimization tool, as applied for example by Chakkalakal et al. (2018) could be subject to future research.

3.7 Bibliography

Acerbi, C., & Szekely, B. (2017). General Properties of Backtestable Statistics. SSRN Electronic Journal.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent Measures of Risk. Mathematical Finance, 9(3), pp. 203–228.

Bali, T. G. (2007). An Extreme Value Approach to Estimating Interest-Rate Volatility: Pricing Implications for Interest-Rate Options. Management Science, 53(2), pp. 323–339.

Bali, T. G., & Theodossiou, P. (2008). Risk Measurement Performance of Alternative Distribution Functions. The Journal of Risk and Insurance, 75(2), pp. 411–437.

Bao, Y., Lee, T.-H., & Saltoglu, B. (2006). Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check. Journal of Forecasting, 25(2), pp. 101–128.

Bekaert, G., Erb, C. B., Harvey, C. R., & Viskanta, T. E. (1998). Distributional Characteristics of Emerging Market Returns and Asset Allocation. The Journal of Portfolio Management, 24(2), pp. 102–116.

Bhattacharyya, M., & Ritolia, G. (2008). Conditional VaR using EVT – Towards a Planned Margin Scheme. International Review of Financial Analysis, 17(2), pp. 382–395.

Bondt, W., & Thaler, R. (1985). Does the Stock Market Overreact? The Journal of Finance, 40(3), pp. 793–805.

Byrne, P., & Lee, S. (1997). Real Estate Portfolio Analysis under Conditions of Non-Normality: The Case of NCREIF. The Journal of Real Estate Portfolio Management, 3(1), pp. 37–46.

Case, B., Yang, Y., & Yildirim, Y. (2012). Dynamic Correlations Among Asset Classes: REIT and Stock Returns. The Journal of Real Estate Finance and Economics, 44(3), pp. 298–318.

Chakkalakal, L., Hommel, U., & Li, W. (2018). Transport infrastructure Equities in Mixed-Asset Portfolios: Estimating Risk with a Garch-Copula CVaR Model. Journal of Property Research, 35(2), pp. 117–138.

Chan, K., & Gray, P. (2006). Using Extreme Value Theory to Measure Value-at-Risk for Daily Electricity Spot Prices. International Journal of Forecasting, 22(2), pp. 283–300.

Chang, M.-S., Salin, V., & Jin, Y. (2011). Diversification Effect of Real Estate Investment Trusts: Comparing Copula Functions with

Kernel Methods. Journal of Property Research, 28(3), pp. 189–212.

Christoffersen, P. (1998). Evaluating Interval Forecasts. International Economic Review, 39(4), pp. 841–862.

Christoffersen, P. (2004). Backtesting Value-at-Risk: A Duration-Based Approach. Journal of Financial Econometrics, 2(1), pp. 84–108.

Christoffersen, P., Errunza, V., Jacobs, K., & Jin, X. (2014). Correlation Dynamics and International Diversification Benefits. International Journal of Forecasting, 30(3), pp. 807–824.

Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease. Biometrika, 65(1), pp. 141–151.

Coleman, M. S., & Mansour, A. (2005). Real Estate in the Real World - Dealing with Non-Normality and Risk in an Asset Allocation Model. Journal of Real Estate Portfolio Management, 11(1), pp. 37-54.

Consigli, G. (2002). Tail Estimation and Mean–VaR Portfolio Selection in Markets Subject to Financial Instability. Journal of Banking & Finance, 26(7), pp. 1355–1382.

Cotter, J., & Stevenson, S. (2006). Multivariate Modeling of Daily REIT Volatility. The Journal of Real Estate Finance and Economics, 32(3), pp. 305–325.

de Melo Mendes, B. V. (2005). Asymmetric Extreme Interdependence in Emerging Equity Markets. Applied Stochastic Models in Business and Industry, 21(6), pp. 483–498.

Deng, L., Ma, C., & Yang, W. (2011). Portfolio Optimization via Pair Copula-GARCH-EVT-CVaR Model. Systems Engineering Procedia, 2, pp. 171–181.

Dittmar, R. F. (2002). Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns. The Journal of Finance, 57(1), pp. 369–403.

Du, Z., & Escanciano, J. C. (2017). Backtesting Expected Shortfall: Accounting for Tail Risk. Management Science, 63(4), pp. 940–958.

Dulguerov, M. (2009). Real Estate and Portfolio Risk: An Analysis based on Copula Functions. Journal of Property Research, 26(3), pp. 265–280.

Ergen, I. (2015). Two-Step Methods in VaR Prediction and the Importance of Fat Tails. Quantitative Finance, 15(6), pp. 1013–1030.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. The Journal of Business, 38(1), pp. 34–105.

Frank, M. J. (1979). On the Simultaneous Associativity of F (x, y) and x + y − F (x, y). Aaequationes Mathematicae, 19(1), pp. 194–226.

Fritz, C., & Oertel, C. (2021). AR-GARCH-EVT-Copula for Securitised Real Estate: An Approach to improving Risk Forecasts?. Journal of Property Research, 38(1), pp. 71-98.

Gabriel, C., & Lau, C. (2014). On the Distribution of Government Bond Returns: Evidence from the EMU. Financial Markets and Portfolio Management, 28(2), pp. 181–203.

Goorah, A. (2007). Real Estate Risk Management with Copulas. Journal of Property Research, 24(4), pp. 289–311.

Gumbel, E. J. (1960). Bivariate Exponential Distributions. Journal of the American Statistical Association, 55(292), pp. 698–707.

Hansen, P. R., & Lunde, A. (2005). A Forecast Comparison of Volatility Models: Does Anything beat a GARCH(1,1)? Journal of Applied Econometrics, 20(7), pp. 873–889.

Harris, R. D., & Kucukozmen, C. C. (2001). The Empirical Distribution of UK and US Stock Returns. (Journal of Business Finance Accounting, 28(5-6), pp. 715–740.

Hiang Liow, K. (2008). Extreme Returns and Value at Risk in International Securitized Real Estate Markets. Journal of Property Investment & Finance, 26(5), pp. 418–446.

Hoesli, M., & Oikarinen, E. (2012). Are REITs Real Estate? Evidence from International Sector Level Data. Journal of International Money and Finance, 31(7), pp. 1823–1850.

Hoesli, M., & Reka, K. (2013). Volatility Spillovers, Comovements and Contagion in Securitized Real Estate Markets. The Journal of Real Estate Finance and Economics, 47(1), pp. 1-35.

Hoesli, M., Lekander, J., & Witkiewicz, W. (2003). Real Estate in the Institutional Portfolio. The Journal of Alternative Investments, 6(3), pp. 53–59.

Huang, J.-Z., & Zhong, Z. (2013). Time Variation in Diversification Benefits of Commodity, REITs, and TIPS. The Journal of Real Estate Finance and Economics, 46(1), pp. 152–192.

Hui, E. C., & Chan, K. K. (2013). The European Sovereign Debt Crisis: Contagion Across European Real Estate Markets. Journal of Property Research, 30(2), pp. 87–102.

Hurd, M., Salmon, M., & Schleicher, C. (2007). Using Copulas to Construct Bivariate Foreign Exchange Distributions with an Application to the Sterling Exchange Rate Index (Vol. 334). Bank of England Working Paper.

Jirasakuldech, B., Campbell, R. D., & Emekter, R. (2009). Conditional Volatility of Equity Real Estate Investment Trust Returns: A Pre- and Post-1993 Comparison. The Journal of Real Estate Finance and Economics, 38(2), pp. 137–154.

Joe, H. (1993). Parametric Families of Multivariate Distributions with Given Margins. Journal of Multivariate Analysis, 46(2), pp. 262–282.

Joe, H. (1997). Multivariate Models and Dependence Concepts (1. CRC reprint ed., Vol. 73). Boca Raton, Fla.: Chapman & Hall/CRC.

Joe, H., & Xu, J. J. (1996). The Estimation Method of Inference Functions for Margins for Multivariate Models.

Karmakar, M. (2017). Dependence Structure and Portfolio Risk in Indian Foreign Exchange Market: A GARCH-EVT-Copula Approach. The Quarterly Review of Economics and Finance, 64, pp. 275–291.

Kole, E., Koedijk, K. C., & Verbeek, M. (2007). Selecting Copulas for Risk Management. Journal of Banking & Finance, 31(8), pp. 2405–2423.

Kuester, K., Mittnik, Stefan, & Paolella, M. (2006). Value-at-Risk Prediction: A Comparison of Alternative Strategies. Journal of Financial Econometrics, 4(1), pp. 53–89.

Kuhle, J. L., & Alvayay, J. R. (2000). The Efficiency of Equity REIT Prices. Journal of Real Estate Portfolio Management, 6(4), pp. 349-354.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. The Journal of Derivatives, 3(2), pp. 73–84.

Mabrouk, S., & Saadi, S. (2012). Parametric Value-at-Risk analysis: Evidence from Stock Indices. The Quarterly Review of Economics and Finance, 52(3), pp. 305–321.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. The Journal of Business, 36(4), p. 394.

McNeil, A. J., & Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. Journal of Empirical Finance, 7(3), pp. 271–300.

McNeil, A. J., Frey, R., Embrechts, P., & Frey, R. D. (2005). Quantitative Risk Management. New Jersey: Princeton University Press.

Nelsen, R. B. (1999). An Introduction to Copulas (Second Edition ed.). New York, NY: Springer Science+Business Media Inc.

Nolde, N., & Ziegel, J. F. (2017). Elicitability and Backtesting: Perspectives for Banking Regulation. The Annals of Applied Statistics, 11(4), pp. 1833–1874.

Officer, R. R. (1972). The Distribution of Stock Returns. Journal of the American Statistical Association, 67(340), pp. 807–812.

Payne, J., & Sahu, A. (2004). Random Walks, Cointegrtion, and the Transmission of Shocks across Global Real Estate and Equity Markets. Journal of Economics and Finance, 28(2), pp. 198-210.

Rachev, S. T., Schwarz, E. S., & Khindanova, I. (2003). Stable Modeling of Market and Credit Value-at-Risk. In S. T. Rachev, Handbook of

Heavy Tailed Distributions in Finance (pp. 249–328). Amsterdam: Elsevier.

Richter, J., Thomas, M., & Füss, R. (2011). German Real Estate Return Distributions: Is There Anything Normal? Journal of Real Estate Portfolio Management, 17(2), pp. 161–179.

Rocco, M. (2014). Extreme Value Theory in Finance: A Survey. Journal of Economic Surveys, 28(1), pp. 82–108.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. Journal of Risk, 2(3), pp. 21–41.

Rossignolo, A. F., Fethi, M. D., & Shaban, M. (2012). Value-at-Risk Models and Basel Capital Charges. Journal of Financial Stability, 8(4), pp. 303–319.

Sahamkhadam, M., Stephan, A., & Östermark, R. (2018). Portfolio Optimization based on GARCH-EVT-Copula Forecasting Models. International Journal of Forecasting, 34(3), pp. 497–506.

Schindler, F. (2010). Further Evidence of the (In-)Efficiency of the U.S. Housing Market. Discussion paper / ZEW - Centre for European Economic Research (10-004).

Sklar, A. (1959). Fonctions de Répartition à n dimensions et leurs Marges. Publications de l'Institut Statistique de l'Université de Paris, 8, pp. 229–231.

Sklar, A. (1973). Random Variables, Joint Distribution Functions, and Copulas. Kybernetika, 9(6), pp. 449–460.

Wang, Z.-R., Chen, X.-H., Jin, Y.-B., & Zhou, Y.-J. (2010). Estimating Risk of Foreign Exchange Portfolio: Using VaR and CVaR based on GARCH–EVT-Copula model. Physica A: Statistical Mechanics and its Applications, 389(21), pp. 4918–4928.

Wei, Y.-h., & Zhang, S.-y. (2004). Dependence Analysis of Finance Markets: Copula-GARCH Model and its Application. Systems Engineering(4).

Wheaton, W. C., Torto, R. G., Southard, J. A., Sivitanides, P. S., & Richard, C. (1999). Evaluating risk in real estate. Real Estate Finance(16), pp. 15-22.

Wu, C.-C., & Lin, Z.-Y. (2014). An Economic Evaluation of Stock–Bond Return Comovements with Copula-Based GARCH Models. Quantitative Finance, 14(7), pp. 1283–1296.

Yang, J., Zhou, Y., & Leung, W. K. (2012). Asymmetric Correlation and Volatility Dynamics among Stock, Bond, and Securitized Real

Estate Markets. The Journal of Real Estate Finance and Economics, 45(2), pp. 491–521.

Young, M. S., Lee, S. L., & Devaney, S. P. (2006). Non-Normal Real Estate Return Distributions by Property Type in the UK. Journal of Property Research, 23(2), pp. 109–133.

3.8 Appendix

Table 14: List of applied copulae

Bivariate Copula family

One parameter	Two parameters	90°-rotated	180°-rotated	270°-rotated
Gaussian	Clayton-Gumbel (BB1)	Clayton	Clayton	Clayton
Student-t	Joe-Gumbel (BB6)	Gumbel	Joe	Gumbel
Clayton	Joe-Clayton (BB7)	Joe	Joe-Gumbel (BB6)	Joe
Gumbel	Joe-Frank (BB8)	Clayton-Gumbel (BB1)	Joe-Frank (BB8)	Clayton-Gumbel (BB1)
Frank		Joe-Gumbel (BB6)		Joe-Gumbel (BB6)
Joe		Joe-Clayton (BB7)		Joe-Clayton (BB7)
		Joe-Frank (BB8)		Joe-Frank (BB8)

4 The Performance Gap: When is average Investor Performance poor?

Authors: Carsten Fritz, René-Ojas Woltering, Steffen Sebastian

Submission to: The Journal of Empirical Finance

Status: revise and resubmit, under revision since April 2021

Abstract

This study examines the market timing ability of mutual fund investors, measured by the difference between money-weighted and buy-and-hold returns. We find an annual performance gap of 0.52% versus 0.20% when using a more realistic reinvestment assumption via modified internal rates of return (MIRR) versus traditionally used IRRs. Rolling-window calculations enable us to track when actual investor performance is good or poor. The time-variant performance gap demonstrates spikes in the performance gap around crises as well as money-weighted returns being susceptible to the length of rolling windows. Fama-MacBeth regressions show that the performance gap is larger for retail investors and increases with underlying return volatility. Passive investment approaches appear to reduce the performance gap.

4.1 Introduction

The typical investor tends to accept the imperfect choices and high fees imposed by active mutual funds, and compounds those liabilities by buying and selling at the wrong times. Economists often like to implement the concept of rationality when talking about the behavior of investors. The mutual fund literature, together with that on behavioral finance, provide evidence that this concept is of limited use, since investors tend to behave suboptimal. Hirshleifer and Shumway (2003), for example, demonstrate that fully rational price setting is not present at all. They show that even the weather, as a proxy for the mood of an investor, has a significant influence on stock returns. Thus, the investment decision is prone to irrationalities such as emotions and other psychological biases like overconfidence, excessive risk taking as well as framing, just to name a few (e.g. Kahneman and Tversky, 1979; Black, 1986; Shiller, 2016). While institutional investors are generally considered more sophisticated than retail investors, (Keim and Madhavan, 1995), Fisher and Statman (2002) point out that institutional investors are also subject to behavioral biases. Apparently, institutional investors do not necessarily implement rational investment strategies and exhibit irrational behaviors like momentum strategy, herding behavior and anchoring effects, just like retail investors (e.g. Grinblatt and Keloharju, 2000; Luo and Li, 2008; Dichtl and Drobetz, 2011; Freiburg and Grichnik, 2013).

This paper examines the market timing ability of mutual fund investors to answer the research question when investor performance is poor. To track relative investment performance, we calculate the difference between money-weighted and buy-and-hold returns experienced by mutual fund investors. We find an annual performance gap of 0.52% versus 0.20% when using a more realistic reinvestment assumption via modified internal rates of return (MIRR) versus traditionally used IRRs.

We are the first to analyze fund-level rolling performance gaps in a Fama-MacBeth regression framework. Our regression analyses are based on 7,480 U.S. mutual equity funds in the period from the beginning of 1999 up to the end of 2019 This setting allows us to test for new determinants of the performance gap over time, while including proven control variables already used in the cross-sectional regression models of Friesen and Sapp (2007), as well as Hsu et al. (2016). We extend the literature by exactly tracking the volatility of timing performance and demonstrating the existence of bad timing-performance in times of financial turmoil (Hypothesis 1). The regression results show that this dummy variable is highly significant. Thus, the performance gap is more prominent (timing-performance of the individual investor is worse) when there is a financial crisis. To the best of our knowledge, we are the first to implement such a crisis-dummy, since this is only feasible within a panel regression approach and performance gaps based on rolling-window returns.

Our results also show that investors in index and retirement funds experience a significantly lower performance gap, whereas retail investors demonstrate worse timing-performance, irrespective of the underlying calculation approach.

Our paper contributes to several strands of the literature. First, we contribute to a growing literature which distinguishes between the investment acumen of fund managers via alpha or time-weighted return (e.g. Bollen and Busse, 2001; Dellva, DeMaskey, and Smith, 2001; Cuthbertson and Nitzsche, 2010) and literature that analyzes the actual investment skill of the average investor (e.g. Dichev, 2007; Friesen and Sapp; 2007, Hsu, Myers, and Whitby; 2016). Comparing the returns reported by mutual funds to the returns actually achieved by investors, the result is somewhat sobering. Trading in and out of mutual funds as opposed to a simple buy-and-hold strategy is the reason why investors experience a different return to that of the underlying security. There are two discriminative concepts to estimate the returns actually experienced by the average mutual fund investors. Time-weighted return (TWR, the geometric mean return or simple buy-and-hold return) and money-weighted return (MWR, the internal rate of return as well as modified internal rate of return). The delta of these two return measures results into the performance gap. Hence, a positive gap is a sign of investor return falling short of buy-and-hold return. For a sample of all funds, we find an annualized rolling twelve-month performance gap of 2.3% according to IRR, and up to 4.2% based on MIRR calculations during crisis periods. Considering the full period, the average performance gap is 0.20%, respectively 0.52%. We also find that the performance gap increases with longer periods of rolling windows. This indicates that calculating a performance gap for the full available history of a fund's data, as Dichev (2007) and others did, does not shed light on the complete picture of the performance gap and thus on the actual investment experience of an average investor. Additionally, it supports the findings of Keswani and Stolin (2008), mentioned in the subsequent literature review.

Second, our paper contributes to the literature on illustrating why and especially when the timing-performance of the average investor is poor or good. We build up the analyses of Clare and Motson (2010), who introduced a rolling window performance gap at an aggregate level, but the authors only provided univariate statistics. In contrast, our panel regression analyses include important control variables such as fund age, fund size, the fund's expense ratio, net cash flow, or standard deviation. To the best of our knowledge, we are the first authors that calculate the performance gap for various rolling-return windows and combine these time-variant performance gaps with a regression analysis that is, contrary to Friesen and Sapp (2007) or Hsu et al. (2016), not based on simple means. In particular, calculating rolling time-weighted and money-weighted returns enables us to reexamine the performance gap over certain periods of time and to analyze it's change over time. This provides much more valuable detail and insight than previous studies on this topic. Thus, we show where the performance was actually lost or gained and in which periods the average investor was more or less able to put cash in and take cash out of his investment at the right time and with the appropriate magnitude. No published study so far tracks the impact of rolling multiple-period returns on a single cash flow of a fund, and therefore measures the full consequence of average investor cash-flow timing decisions.

Third, we introduce an alternative measure of fund-flow weighted returns calculating a modified internal rate of return (MIRR) that is based on a more realistic reinvestment assumption compared to the classic approach via internal rate of return (IRR). Our analyses show that the average investor experiences, money-weighted returns that are significantly lower than buy-hold-returns. We find that poor timing is not unique to retail or non-index fund investors. We find a similar pattern for all fund sub-categories in our sample of index, non-index, institutional and retail fund shares. The IRR and MIRR is almost always significantly lower than the buy-hold-return. Comparing these sub-categories, reveals that on average, retail investors have a higher performance gap than institutional investors (0.21% versus 0.15% in the case of twelve-months rolling window returns) according to calculations using IRR. Considering index and non-index funds, the latter yield a performance gap that is two to three times higher than the former, regardless of the underlying calculation method of money-weighted returns.

The remainder of the paper is organized as follows. Section 2 discusses the literature on the performance gap. In Section 3, we discuss the methodology as well as the partly unrealistic assumptions underlying the IRR calculation approach. Additionally, we provide descriptive statistics. Section 4 consists of a quintile-based analysis of standard deviation and fund size, in order to demonstrate their influence on the performance gap in more detail. Section 5 presents the empirical results and implications concerning the performance gap and its drivers. Finally, Section 6 concludes.

4.2 Related Literature and Hypothesis

Considering the aggregate fund level, Nesbitt (1995) finds an annualized performance gap of 1.08% for U.S. mutual funds for the period of 1984 to 1994. A similar study conducted by Braverman, Kandel, and Wohl (2005) concludes that the difference between time-weighted and money-weighted returns must be due to investor sentiment or time-varying expected returns. In his influential paper, Dichev (2007) finds a gap of 1.3% between dollar-weighted and buy-and-hold returns for a sample of NYSE and AMEX indices. One of his corollaries is that the risk premium an investor expects for investing in a particular asset, is upwardly biased. Shortly after Dichev's study, Keswani and Stolin (2008) published a paper contradicting almost all the findings. Using the same data, they find that the results are not actually robust. Instead of calculating a performance gap for the full sample period, they reveal that building shorter subsamples leads to different results, even to a negative performance gap. Considering shorter periods is more realistic when thinking about the typical investment horizon of an average investor. Cutting or extending the investment horizon might influence the overall performance gap, inter alia due to the dependence of dollar weighted returns on periods with increased flows into or out of the investment. Using the same technique as Dichev (2007), Dichev and Yu (2009) investigate the returns achieved by investors in hedge funds and find a much larger performance gap of 4% up to 9% per year.

Research at the individual fund level sheds additional light on the timing-performance of the average investor. Zweig (2002) is one of the first to identify that the return investors receive from mutual funds is less than the actual real return proclaimed by the same fund. He analyzes 100 U.S. stock funds right after the bursting of the dot-com bubble and finds a positive performance gap. Thus, he concludes that investors should abandon the hope of ever investing into a fund at the bottom and selling out right at the top. Hence, he conjectures that the poor timing skills of the investors themselves are the reason for the performance gap. Friesen and Sapp (2007) assess the timing ability of investors for 7,125 U.S. equity mutual funds for the period 1991 to 2004. The authors find an overall performance gap of 1.56% annually. Once again, the performance gap is calculated considering the full sample period of each fund. Their key finding is that the difference between dollar-weighted and time-weighted returns largely offsets any risk-adjusted outperformance generated by a well-performing fund. Hence, if one is lucky enough to find and invest in a fund that generates a positive alpha, the potential surplus in return does not actually materialize for the average investor. Friesen and Sapp (2007) also conduct a basic multivariate regression analysis based on the cross-section of simple means for the full time-series of the respective variable, and conclude that a greater performance gap results from return-chasing investor behavior, since underperformance due to timing is positively associated with momentum-style funds as well as those with higher returns. Variables which influence the performance gap include volatility, volume or size of fund, fund loads, turnover and the length of fund history. Thus, it seems that bigger and more costly funds attract unsophisticated investors which leads to a more significant difference between time-weighted and money-weighted returns. In their working paper, Clare and Motson (2010) were the first to show a time-varying performance gap calculated on a rolling basis for retail, institutional and bond mutual funds. They analyze market-level data for UK mutual funds as well as a subsample of individual fund data. The figures show that retail investors experienced poor money-weighted returns in the period from 2000 – 2004, whilst institutional investors received higher IRR-returns compared to buy-and-hold during the Great Financial Crisis. Clare and Motson do not elaborate further on their rolling performance gap measure, which shows when the actual performance loss happened. Hence, we conduct our analysis based on their approach and do elaborate further.

Together with the findings of Keswani and Stolin (2008) mentioned above, this leads us to our first hypothesis concerning the period used to calculate money- and time-weighted returns:

Hypothesis 1: *The performance gap increases with higher volatility and thus during crisis periods.*

Hsu, Myers, and Whitby (2016) also conduct a multivariate regression analysis based on a cross-section of simple means for the full time-series concerning U.S. mutual funds over the period from 1991 through 2013. They find an underperformance, as measured by IRR, of almost 2% per year for a sample of all funds and a remarkable 2.72% for index funds, which rather tend to attract investors with a passive, buy-and-hold type of investment strategy, and thus demonstrate better timing-performance. This underperformance varies, when considering different fund categories and investor cohorts like institutional and retail investors. Following their results, the authors hypothesize that investors who time poorly tend to be unsophisticated. They also reason that a simple buy-and-hold approach ignores the fact that investors tend to trade on a regular basis and therefore, money-weighted returns are a better measure of average investor performance.[10] The authors classify retirement share classes as institutional funds, although they state that these share classes are dominated by small investors following a simple buy-and-hold strategy. Therefore, their assumptions provide latitude for

[10] Further studies at the individual fund level, which find poor timing-performance include: Chieh-Tse Hou (2012), Navone and Pagani (2015), Muñoz (2016).

our second hypothesis concerning different investor clienteles following varying investment strategies, therefore revealing differences concerning their timing-performance:

Hypothesis 2: *The performance gap is smaller for investors following a buy-and-hold like strategy.*

As the literature shows, a consensus has emerged that the aggregate effects of poor investor timing are substantial. Nonetheless, almost all of these studies calculate a performance gap, at the aggregate or individual level, for the full available data history and sample period of a fund. As Keswani and Stolin (2008) pointed out when questioning the results of Dichev (2007), the performance gap is highly sensitive to aggregation across time and across the cross-section. We therefore adopt the approach of a performance gap calculated on a rolling basis, as proposed by Clare and Motson (2010).

Additionally, the literature so far focuses on the internal rate of return (IRR) as a calculation approach of money-weighted returns, although a modified internal rate of return (MIRR) provides solutions to partly unrealistic assumptions, such as flows into the investment vehicle being reinvested at the rate of the IRR itself, underlying the IRR approach. Reinvesting free cash flows at the rate of return of the fund itself, or parking cash at least for the risk-free rate, is a more realistic approach. Therefore, the research objective of this article is to examine the timing ability of the average mutual fund investor by illustrating the performance gap as well as its change throughout the research period for U.S. domiciled mutual funds investing in U.S. equity, based on three-, six-, and twelve-month rolling-window calculations of returns.

We further show the determinants of the performance gap at the individual fund share-class level using Fama-MacBeth regressions for our panel data set of rolling window variables, instead of regressions based on simple means as deduced so far in the literature. As mentioned above, introducing the MIRR as an alternative measure of money-weighted return enables us to correct for typical drawbacks of the IRR, like the reinvestment assumption. Therefore, and in addition to the literature so far, we work with a measure of average investor returns which is more realistic, considering the investment experience of an individual. For comparison, we provide all results calculated with time-weighted returns based on IRR as well as MIRR.

4.3 Measuring the Performance Gap

Our empirical study is based on a sample of 7,480 mutual equity funds domiciled in the U.S., with an investment focus on U.S. equities for the period January 1999 to December 2019. The data is obtained from the Morningstar mutual fund database. We exclude so-called fund of funds, exchange traded funds as well as those with a prospectus objective that does not fit (e.g. Real Estate, World Stock etc.), as the benchmarks employed in our analyses may otherwise be inappropriate. The sample is cross-checked and complemented where necessary via Datastream, the SEC Edgar file search and the respective fund prospectuses. Fund share classes are treated as distinct funds. Only share classes with at least 75% of the asset allocation in U.S. equity are considered. Special share classes marked as load waived are excluded due to a lack of data. The initial sample covers the period from January 1999 – December 2019, since the data coverage before January 1999 was insufficient. The sample contains total net assets (TNA) on a monthly basis, as well as monthly fund return and other typical fund characteristics (e.g. category, inception date etc.). In accordance with Hsu, Myers, and Whitby (2016), funds with fewer than 24 monthly observations and average TNAs of less than 10 million USD, considering the full lifetime of the fund in the sample, are excluded from the dataset.

Concerning possible outliers in our sample, we follow the approach of Cashman et al. (2014) and eliminate observations of monthly net flows with a net cash flow ratio (compared to beginning TNA) greater than 50%, or less than -20% of assets. Additionally, every time-series is trimmed at the 1%/99% level. These observations are likely to result from data entry issues. After full treatment of the data, this leads to a sample of 7,480 mutual equity funds. Table 15 reports descriptive statistics for the sample of funds.

Table 15: Sample statistics for U.S. equity funds

	Mean	Median	25th perc.	75th perc.	Stand. Dev.
Panel A: All Funds (n = 7480)					
TNA ($ millions)	335.10	319.61	277.92	403.84	80.58
Monthly net cash flow ($ millions)	-0.80	-0.79	-1.72	0.14	1.57
Expense ratio (% per year)	1.19	1.20	1.10	1.28	0.10
Monthly return (%, annualized)	8.45	16.02	-19.07	52.68	15.25
Panel B: Non-index funds (n = 6925)					
TNA ($ millions)	307.63	298.54	259.90	368.90	68.48
Monthly net cash flow ($ millions)	-0.88	-0.83	-1.76	0.01	1.50
Expense ratio (% per year)	1.24	1.26	1.15	1.34	0.10
Monthly total return (%, annualized)	8.48	16.08	-19.07	52.14	15.27
Panel C: Index funds (n = 555)					
TNA ($ millions)	815.39	701.24	557.57	1013.97	347.47
Monthly net cash flow ($ millions)	1.13	0.53	-0.99	2.55	4.00
Expense ratio (% per year)	0.50	0.50	0.48	0.51	0.03
Monthly total return (%, annualized)	8.07	15.23	-18.90	51.37	15.08
Panel D: Institutional funds (n = 1881)					
TNA ($ millions)	345.64	320.46	273.90	430.76	99.38
Monthly net cash flow ($ millions)	-0.03	0.24	-0.61	0.91	1.49
Expense ratio (% per year)	0.89	0.89	0.86	0.91	0.03
Monthly total return (%, annualized)	8.76	17.25	-18.97	52.03	15.28
Panel E: Retail funds (n = 4705)					
TNA ($ millions)	353.71	342.04	296.38	428.65	82.94
Monthly net cash flow ($ millions)	-1.18	-1.31	-2.39	-0.37	1.83
Expense ratio (% per year)	1.31	1.33	1.21	1.40	0.10
Monthly total return (%, annualized)	8.30	15.76	-19.18	52.90	15.25

Notes: For TNA (total net assets), net cash flow and total return, we compute the monthly cross-sectional averages. The reported statistics are computed as the mean from each time-series of these monthly averages. The expense ratio is based on the cross-sectional average of 20 equal-weighted annual observations (1999 - 2019) for each fund in the respective panel.

The table is divided into panels in order to provide detailed insight into the sub-samples we will be analyzing. Panel A of Table 15 shows that the average fund in our sample suffers overall from outflows during the analyzed period. This suggests a trend to move from actively managed vehicles to, for example, exchange traded funds. We also retrieved data about the last recorded front-load fees of a fund. Unfortunately, this data was not available as a monthly or yearly time-series for our sample. It shows that around 16% of all funds charge or charged front-load fees. Comparing Panel B with Panel C shows that index funds are larger in terms of total net assets, which is confirmed when taking a look at the current list of the largest mutual U.S. equity funds. Index funds are the only group in our sample that, on average, received inflows. Their annual expense ratio is typically lower, which is in line with the fact that they are not as actively managed as non-index funds. Having said that, their average annualized total return is typically lower. 16.5% of non-index funds and only 7% of index funds charge a front load. Panel D and Panel E confirm the advantage of institutional funds compared to retail funds considering total return. Average outflows for institutional funds were not as bad as for retail funds. In contrast to institutional funds, retail funds charge a higher expense ratio. In our sample, only 0.13% of institutional funds have front-loads, whereas 26% of retail funds require an upfront payment.

The following criteria were used to separate institutional funds from retail funds. Frist of all, if a fund has the word “institutional” in its name, we looked into greater detail at that fund. If a fund share class has an initial purchase limit of at least 100.000 USD or states in its prospectus that it is designed for institutional investors, or only sells shares to institutional investors, we classified that share class as an institutional fund. A similar approach has also been applied by other authors, see James and Karceski (2006) or Ammann et al. (2018), for example. Z-share classes were marked as institutional funds. Z-shares are those that only employees of the fund are allowed to own. Hence, we assume that these individuals are relatively investors. Retirement share classes are neither counted as institutional nor as retail share class. As mentioned in the literature review, we disagree with Hsu et al. (2016), who classify retirement share classes as a part of institutional funds. Although the average investor could decide on the timing and magnitude of the investment in a retirement share class on his own, inflows for these share classes typically tend to occur in repeated patterns without following an actual decision of the investor. Thus, we rather do not treat these flows from defined contribution plans as cash flows from retail or institutional investors.

Following Friesen and Sapp (2007), we calculate the monthly delta of time-weighted and money weighted return. The difference between these two rates is then used as a measure of the effect that the timing of these flows exerts on average investor return. This delta is called a performance gap.

$$Performance\ Gap_i = TWR_i - MWR_i \tag{19}$$

where $TWR_{i,t}$ is calculated as the geometric return of each fund for the respective period. It provides the return that would be earned if investors followed a strict buy-and-hold strategy, immediately reinvesting any dividends. $MWR_{i,t}$ represents the money-weighted return and can either be calculated as IRR or MIRR of fund i. Therefore, a positive performance gap demonstrates poor timing performance of the average investor, compared to the buy-and-hold return. Hence, more investors participated in downside returns and less in upside returns. A negative performance gap shows that the average investor outperformed the actual time-weighted return due to the timing and magnitude of flows, so that good timing-skill is present. The IRR is defined as the rate of return at which the value of TNA at the end of the sample period equals the accumulated value of initial TNA plus the accumulated value of all net cash flows (NCF):

$$TNA_{i,0}(1 + IRR_i)^T + \sum_{t=1}^{T} NCF_{i,t}(1 + IRR_i)^{(T-t)} = TNA_{i,T} \tag{20}$$

In contrast to IRR based calculations, the MIRR assumes that cash flows are reinvested at an appropriate or more realistic rate of return. Cash flows withdrawn the fund by the investor are compounded by a risk-free rate, the one-month US T-Bill rate, IA SBBI US 30 Day TBill TR USD, provided by Ibbotson Associates.[11] These cash flows can always be reinvested at the risk-free rate until the end of the analyzed period. For flows invested into the fund, we assume that the monthly total return of the particular fund is a more realistic rate of return. We elaborate further concerning this approach in the next section. The following formula shows the calculation of the MIRR for the n periods considered:

$$MIRR_i = \sqrt[n]{\frac{FV\ of\ positive\ Cash\ Flows_i}{PV\ of\ negative\ Cash\ Flows_i}} - 1 \tag{21}$$

For the future value of cash flows, the positive flows and therefore those coming out of the fund or inflows for the investor, are compounded at the risk-free rate until the end of the period and summed. Negative cash flows and therefore flows into the fund, which are outflows for the investor, are discounted to the beginning of the period and also summed as present value. The variable n is the horizon period over which projects are evaluated.

As mentioned earlier, we do not derive means for the full time-series for each fund. Instead, we create a rolling performance gap for twelve, six and three months. Thus, in the case of a twelve months rolling window, the performance gap for December 2019 consists of the difference between time-weighted and money-weighted returns from December 2018 until December 2019. The respective cash flow for the time-

[11] We also confirmed the robustness of our results using three-month and twelve-month US T-Bill rates.

weighted return measures accordingly begins with initial TNA from November 2018, since previous month TNA also start TNA for the next month. These one-month-ahead rolling windows yield detailed insight into when and where performance was actually lost or gained.

Calculating a performance gap using money-weighted returns based on IRR has been the main focus of the literature so far. However, the concept of IRR exhibits a number of drawbacks. It is a relative measure of value creation, multiple answers are possible if cash flows go from negative to positive more than once, so that it is difficult to calculate, the average IRR is different to the IRR of aggregated cash flows, and it makes an unrealistic reinvestment rate assumption.[12] IRR is the average rate of return that will be earned if the external cash flows are financed or reinvested using an implicit reinvestment assumption, where cash inflows and outflows are reinvested at an interest rate that is identical to the IRR itself. However, it is known that the actual rate of return experienced by investors differs from IRR. Therefore, we also calculate money-weighted returns according to the MIRR which uses explicit (dynamic) reinvestment assumptions. Modified Internal Rate of Return is used to account for the fact that cash flows are reinvested and compounded at rates different from and more realistic than the IRR. However, we are aware of introducing subjectivity into the measurement of money-weighted returns by using unique finance or

[12] See Phalippou (2008) for more information on the biasedness of IRR as a return measure.

reinvestment rates. Implementing this concept of MIRR provides a theoretical and, compared to the IRR, potentially more realistic investment strategy for the average investor. An actual investor might not use this strategy, however, investing free cash-flow in an asset that returns at least the risk-free rate is always a possibility. Additionally, Balyeat et al. (2013) argue that MIRR is a more accurate measure of return than IRR, which compensates for the major drawbacks of IRR.

Imagine an investor selling a security and therefore having cash at disposal. The cash can either be reinvested at the monthly return of the respective fund or at the rate of return of a corresponding risk-free investment. Reinvesting the cash flow at the internal rate of return thus seems highly unrealistic, regardless of the actual magnitude of IRR. Before being able to invest in a fund, the full amount of cash flow intended for investment and gained from prior disinvestments might already be available for longer periods of time, being held on the account of the investor until the actual cash inflow eventually takes place. Hence, a more realistic assumption relating to these flows is that for the duration of their holding, they are invested at least at the risk-free rate. Hurley et al. (2014) state that the internal rate of return assumes that intermediate cash flows can be reinvested (or borrowed) at the IRR, and thus, at the same return as the initial investment. However, always being able to reinvest or borrow at the same rate of return throughout the full history of the investment is rather unrealistic. As shown with this example, the discussion about an (implicit) underlying reinvestment assumption considering IRR in this case should not confound the ongoing debate within the private equity and project finance literature as to whether there is an unrealistic assumption about the rate of return or not.[13] We are working with cash flows that are free for investment and therefore need to be (re)invested at a more realistic rate of return.

Concerning the susceptibility of IRR to the length of the investment horizon as Keswani and Stolin (2008) point out, Phalippou (2008) for example, shows that amongst other possibilities, one can boost IRRs by modifying the time horizon of the investments, returning cash to investors earlier for successful projects, and further keeping projects with poor performance alive. The volatility of IRR as a return measure is also exaggerated. According to Ingersoll et al. (2007), the best strategy in order to bias performance measures upwards is to quit whilst ahead, but gamble more following poor outcomes.

We compute monthly rolling return measures (TWR and MWR according to IRR and MIRR) on the basis of twelve, six and three months for all of our return measures. Thus, a rolling performance gap for these windows is calculated. Table 16 shows the overall results for the full sample of all funds.

[13] For more literature analyzing the reinvestment assumption of the IRR from the viewpoint of project finance see, for example, Keef and Roush (2001); Johnston et al. (2002); Kierulff (2008); Walker et al. (2011); Cheremushkin (2012); Ross et al. (2013); Rich and Rose (2014).

Table 16: Return measures and performance gap for Panel A

		Mean	Median	25th perc.	75th perc.	Stand. Dev.
Panel A: All Funds						
12-months rolling	TWR	6.72	11.24	-1.42	18.79	4.90
	IRR	6.37	10.96	-1.46	18.13	4.92
	MIRR	5.93	10.12	-1.46	16.80	4.61
	Gap IRR	0.20	0.11	-0.03	0.37	0.13
	Gap MIRR	0.52	0.69	0.16	1.07	0.35
6-months rolling	TWR	7.03	11.05	-2.92	20.97	6.85
	IRR	6.83	10.81	-3.45	21.05	6.86
	MIRR	6.72	10.09	-3.02	20.59	6.63
	Gap IRR	0.11	0.04	-0.04	0.20	0.09
	Gap MIRR	0.18	0.30	-0.06	0.61	0.25
3-months rolling	TWR	7.50	11.97	-7.10	29.63	9.24
	IRR	7.40	11.91	-7.23	29.77	9.23
	MIRR	7.39	11.50	-7.05	29.59	9.11
	Gap IRR	0.05	0.02	-0.04	0.11	0.06
	Gap MIRR	0.06	0.11	-0.13	0.34	0.14

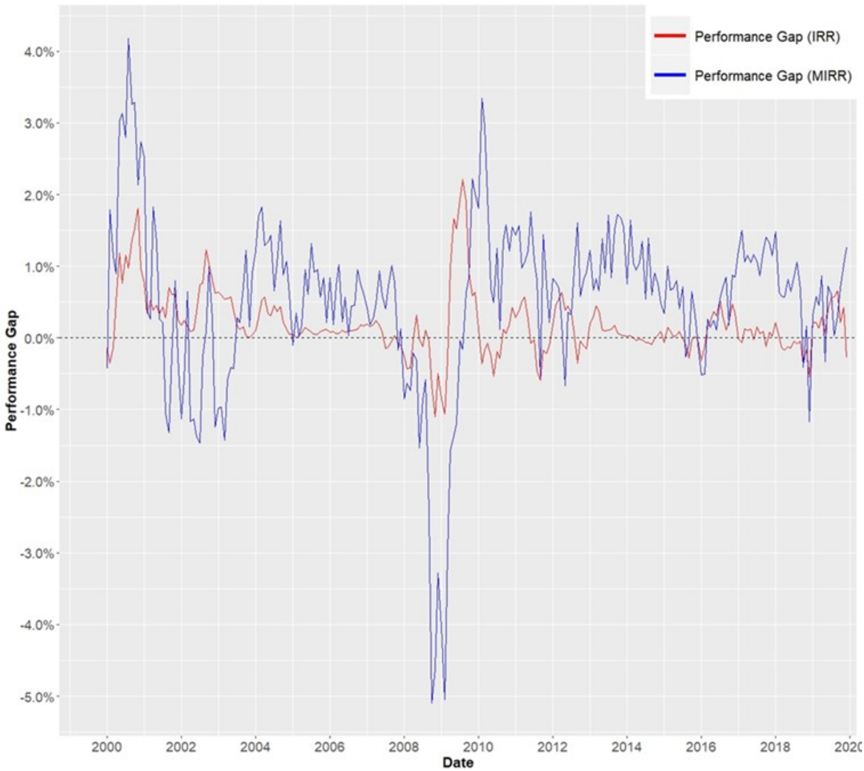
Notes: Results are based on a time-series of monthly cross-sectional averages. The reported statistics are computed as the mean from each time-series of these monthly averages. All figures are annualized and stated in percent. t-statistics for the mean performance gap show that these figures are significantly different from zero at a 5% level.

This first analysis reveals that on average, there is a positive performance gap regardless of the length of the rolling period. For example, the twelve-months rolling-window calculations according to IRR and MIRR yield a performance gap of 0.20%, respectively 0.52%. Thus, the return achieved by the average investor is always lower than simple buy-and-hold return, with MIRR being smaller compared to IRR, and therefore resulting in a larger performance gap considering an MIRR-based calculation. Directly comparing our results with those of Dichev (2007) or Hsu et al. (2016) is not possible due to the time frame underlying the calculations. Throughout the following analyses, we will focus on the results of the twelve-month rolling windows. Additionally, we report findings for other lengths of the rolling window if peculiarities occur. Table 16 shows that the performance gap increases with the length of the rolling period, since there are less possibilities for in- or outflows, and therefore less potential for deviating money-weighted and time-weighted returns and the accumulation of losses. Thus, funds with longer data history will automatically have a greater influence on mean statistics of the performance gap. The performance gap calculated using MIRR is on average larger than the gap according to IRR. This is due to the unrealistic reinvestment assumption of money-weighted returns calculated according to IRR. For example, consider the beginning of a crisis with the reinvestment rate according to IRR still being (unrealistically) high, whilst calculations according to MIRR already use the lower monthly return of the respective fund as the reinvestment rate. Considering free cash flows being held by the investor, these flows are thought to be invested at the risk-free rate instead of the IRR.[14]

[14] We also confirmed the robustness of our results by building 24- and 36-month rolling-windows. Figures for these longer rolling-windows resemble the results for shorter rolling-windows. However, they exhibit a smoothing effect of overall results since the time-span for the analysis is wider. The figures are available from the authors upon request.

For a better comparison of both, the performance gap calculated via IRR or MIRR, one has to have insight into where the actual performance was gained or lost, since the figures in Table 16 are only averages. Positive returns in one period might nullify negative returns in other periods and vice versa. Therefore, Figure 5 shows both performance gap measures for the twelve-month rolling window over the full sample period. Bear in mind that due to the rolling window, the performance gap, for example, in October 2009 results from total net assets, flows and fund returns for the prior twelve months.

Figure 5: Rolling twelve-months performance gap based on IRR or MIRR



Notes: The performance gaps are annualized and stated in percent. Bear in mind that due to the twelve-months rolling window, a performance gap for the time shown in the graph actually consists of TNA, fund flows and returns (money-weighted and time-weighted) for the previous twelve months. A positive performance gap shows that the average investor return was less than the buy & hold return (Performance Gap = TWR – MWR).

Figure 5 clearly shows that the performance gap, no matter what the actual basis of calculation, is more prominent during crises. The average investor's strategy can be described as follows during crises. Due to diminishing returns and the financial downturn itself, investors withdraw their money from funds, which might, at first glance, seem like a good decision in the event of further falling prices. However, after the crisis has come to an end and prices start to rise, investors are no longer fully invested and start reinvesting too late which leads to a larger performance gap. As seen in Figure 5, there is an annual performance gap of around 3.35% for January 2010, which includes all time- and money-weighted returns for the prior twelve months. The periods of the dot-com bubble as well as the great financial crisis are clearly visible. A wide performance gap for the early 2000s is in line with the yearly performance gap measurement of Friesen and Sapp (2007). These detailed insights are not available when considering only averages or calculating a performance gap over the full history of the share class. Both time-series show poor timing skills of the average investor during the mentioned crises. However, they also show an outperformance of money-weighted returns compared to time-weighted returns, especially for the months prior to the great financial crises of 2008/2009. Thus, the average investor was able to participate in the upswing right before the crisis with good timing and magnitude of flows. Considering both performance gaps, there are more positive observations than negative ones. Thus, the average investor seems to have difficulty achieving at least reasonable buy-and-hold returns. Periods demonstrating great timing skills are seldom. Figures for the six- and three- months rolling windows are shown in the appendix (Figures 10 & 11). Thus, the findings from Table 16 and Figure 5 confirm our Hypothesis 1. A look at Figure 5 reveals that especially during crisis periods, there are peaks revealing the poor timing-performance of the average investor. The following quintile analysis concerning the standard deviation of returns will shed further light on this issue.

Comparing index funds with non-index funds, the average statistics show that investors in index funds experienced a smaller performance gap. Table 17 illustrates the statistics for Panel B (non-index funds) and Panel C (index funds), whilst Table 18 shows the figures for Panel D (institutional funds) and Panel E (retail funds). Table 17 demonstrates that concerning time-weighted return, non-index funds outperform index funds, on average. However, their standard deviation is higher. Money-weighted returns show a similar performance. The performance gap, regardless of the length of the rolling-window, is bigger for non-index funds. This is due to their active nature leaving more room for actual trading decisions from the investor. A smaller time-span for the rolling-window leads to a smaller performance gap, supporting the findings of Keswani and Stolin (2008). Considering the average performance gap for institutional versus retail funds, the latter have a larger average performance gap according to IRR calculations. This is not always the case for money-weighted returns based on MIRR, which hints towards institutional investors also showing overall poor timing-performance. Furthermore, Table 17 and Table 18 highlight that a positive performance gap and thus poor timing-performance is apparent within all different groups of investors. Retail investors and non-index funds tend to show the worst underperformance whilst passive investment approaches definitely proof to produce investor return close to a classic buy-and-hold approach.

Table 17: Return measures and performance gap for Panels B & C

		Mean	Median	25th perc.	75th perc.	Stand. Dev.
Panel B: Non-Index Funds						
12-months rolling	TWR	6.73	11.04	-1.49	18.86	4.91
	IRR	6.35	10.83	-1.54	18.24	4.93
	MIRR	5.91	10.00	-1.63	16.89	4.60
	Gap IRR	0.21	0.12	-0.04	0.39	0.14
	Gap MIRR	0.53	0.71	0.17	1.11	0.37
6-months rolling	TWR	7.04	11.00	-3.06	21.09	6.87
	IRR	6.83	10.73	-3.46	21.15	6.87
	MIRR	6.72	10.16	-3.07	20.76	6.63
	Gap IRR	0.11	0.05	-0.04	0.22	0.09
	Gap MIRR	0.18	0.31	-0.05	0.63	0.26
3-months rolling	TWR	7.52	11.96	-7.06	29.67	9.26
	IRR	7.41	11.93	-7.02	29.93	9.25
	MIRR	7.40	11.66	-6.75	29.96	9.12
	Gap IRR	0.05	0.02	-0.04	0.11	0.06
	Gap MIRR	0.06	0.11	-0.13	0.35	0.14
Panel C: Index Funds						
12-months rolling	TWR	6.48	11.10	-0.52	17.66	4.87
	IRR	6.36	11.07	-0.41	17.51	4.88
	MIRR	5.97	10.68	-0.73	16.80	4.72
	Gap IRR	0.07	0.03	-0.04	0.10	0.08
	Gap MIRR	0.39	0.50	0.03	0.83	0.24
6-months rolling	TWR	6.75	10.76	-2.32	21.01	6.75
	IRR	6.65	10.48	-2.62	21.15	6.75
	MIRR	6.57	10.70	-2.52	20.59	6.61

	Gap IRR	0.05	0.01	-0.02	0.08	0.06
	Gap MIRR	0.12	0.20	-0.08	0.44	0.18
3-months rolling	TWR	7.19	11.26	-7.31	30.03	9.02
	IRR	7.12	11.26	-7.29	29.81	9.02
	MIRR	7.11	11.05	-7.13	29.54	8.95
	Gap IRR	0.02	0.01	-0.03	0.06	0.04
	Gap MIRR	0.03	0.07	-0.13	0.24	0.09

Notes: Results are based on a time-series of monthly cross-sectional averages. The reported statistics are computed as the mean from each time-series of these monthly averages. All figures are annualized and stated in percent. t-statistics for the mean performance gap show, that these figures are significantly different from zero at a 5% level except for the three-month rolling Performance Gap according to MIRR. This figure is significantly different from zero at a 10% level for both panels (p-value of 0.054 for Panel B, and p-value of 0.098 for Panel C).

Table 18: Return measures and performance gap for Panels D & E

		Mean	Median	25th perc.	75th perc.	Stand. Dev.
Panel D: Institutional Funds						
12-months rolling	TWR	7.03	11.22	-1.27	19.05	4.89
	IRR	6.77	11.28	-1.59	18.88	4.90
	MIRR	6.12	10.18	-1.92	17.38	4.68
	Gap IRR	0.15	0.09	0.02	0.22	0.07
	Gap MIRR	0.68	0.74	0.12	1.22	0.29
6-months rolling	TWR	7.33	11.23	-2.51	21.33	6.85
	IRR	7.20	11.08	-2.57	21.33	6.84
	MIRR	7.01	10.58	-2.41	20.97	6.65
	Gap IRR	0.08	0.03	-0.01	0.12	0.06
	Gap MIRR	0.22	0.30	-0.10	0.60	0.21
3-months rolling	TWR	7.81	12.34	-7.02	30.60	9.24
	IRR	7.74	12.37	-6.80	30.77	9.22

	MIRR	7.70	12.44	-6.62	30.34	9.11
	Gap IRR	0.03	0.01	-0.03	0.06	0.04
	Gap MIRR	0.06	0.12	-0.12	0.30	0.12
Panel E: Retail Funds						
12-months rolling	TWR	6.56	11.06	-1.53	18.54	4.91
	IRR	6.21	10.75	-1.71	17.89	4.93
	MIRR	5.84	9.95	-1.59	16.62	4.58
	Gap IRR	0.21	0.12	-0.05	0.40	0.17
	Gap MIRR	0.47	0.66	0.12	1.09	0.39
6-months rolling	TWR	6.87	10.95	-3.11	20.86	6.86
	IRR	6.66	10.53	-3.69	20.93	6.87
	MIRR	6.58	9.98	-3.22	20.56	6.62
	Gap IRR	0.11	0.05	-0.06	0.24	0.11
	Gap MIRR	0.17	0.32	-0.07	0.62	0.27
3-months rolling	TWR	7.35	11.93	-7.20	29.47	9.24
	IRR	7.25	11.84	-7.18	29.50	9.24
	MIRR	7.24	11.40	-7.02	29.56	9.11
	Gap IRR	0.05	0.02	-0.06	0.12	0.08
	Gap MIRR	0.06	0.12	-0.17	0.35	0.15

Notes: Results are based on a time-series of monthly cross-sectional averages. The reported statistics are computed as the mean from each time-series of these monthly averages. All figures are annualized and stated in percent. t-statistics for the mean performance gap show that these figures are significantly different from zero at a 5% level expect for the three-month rolling Performance Gap according to MIRR. This figure is significantly different from zero at a 10% level for Panel E (p-value of 0.08).

Analyzing the performance gap in detail for retail and institutional funds, reveals that both showed poor timing skills due the dot-com bubble and the great financial crisis, comparable to Figure 5. Hence, neither of them shows any peculiarities in the course of the IRR or MIRR-performance gap. The same applies to non-index funds. Index funds show peaks and troughs at the same time, but with a lower magnitude as in Figure 5. Thus, an investment in index funds would not have resulted in severe timing-underperformance for the average investor. In general, the magnitude of the performance gap is not as large for index funds compared to all other subsamples, which is in line with our expectations of more passive investment strategies being closer to a plain buy-and-hold approach. This finding is in line with Hsu, Myers and Whitby (2016), assuming that investors in index funds focus on a buy-and-hold-like investment strategy.

4.4 Determinants of the Performance Gap

As specified in Hypothesis 2, we will test for the timing-performance of different investor clienteles following varying investment strategies like index fund investors, plain retail investors as well as professional investors, using dummy variables in a Fama-MacBeth regression model for each of these categories. Our next step, is to check for other potential determinants of the performance gap. As demonstrated, the performance gap seems to be more prominent during crises or expressed differently, in phases of high risk. Thus, we calculate the standard deviation of each funds' monthly return on the rolling basis for all of our panels. Nonetheless, we focus on analyzing the twelve-month rolling results, since this time span enables a detailed look inside the movement of the rolling performance gap, whilst still representing a reasonably long investment horizon for the average investor. In order to analyze whether the standard deviation is a potential driver of the performance gap, we sort all funds into quintiles according to their standard deviation for each individual month in the sample. Table 19 shows the results for those quintiles for the panel of all funds on a twelve-months rolling basis. The results for all other panels and time-frames follow the same trend. Due to limited space, figures for other panels and lengths of rolling windows are available upon request.

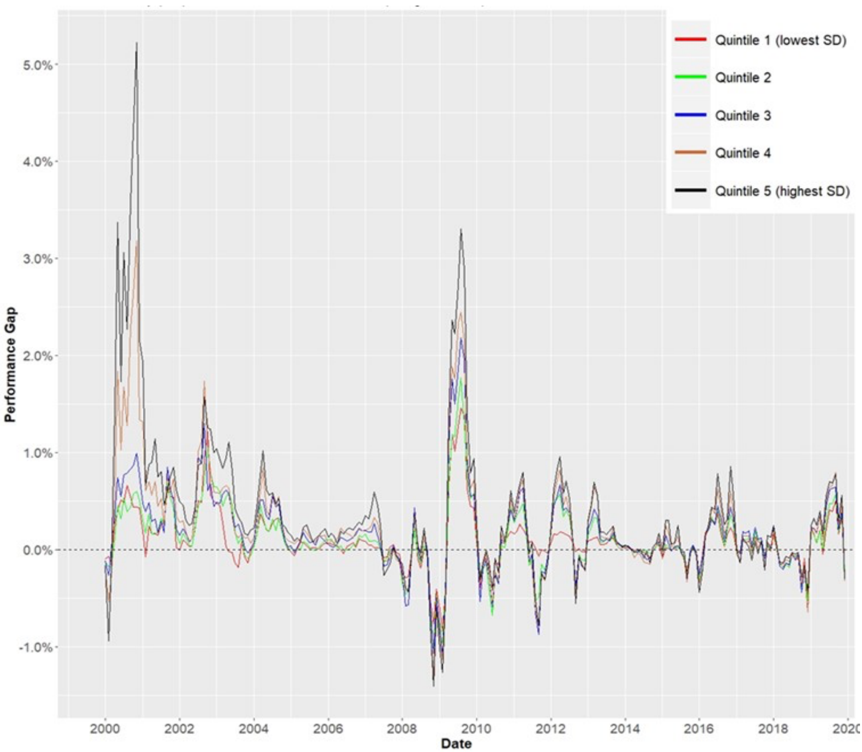
Table 19: Performance gap by fund standard deviation of monthly return

	Quintile 1 (lowest)	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5 (highest)**
Panel A: All Funds					
Avg. standard deviation	11.53	13.38	14.84	16.86	21.02
Avg. Geometric return	8.27	7.65	7.76	8.27	8.77
Avg. IRR	8.05	7.41	7.45	7.94	8.11
Avg. MIRR	7.57	6.90	6.92	7.25	7.17
Avg. Performance Gap IRR	0.10	0.12	0.18	0.26	0.38
Avg. Performance Gap MIRR	0.40	0.43	0.48	0.61	0.80

Notes: All figures are annualized and in percent. All funds are sorted into monthly quintiles based on their standard deviation. Thus, we know the funds that are in the particular quintiles for each month, including their returns and performance gap. The average for each of these monthly quintiles is taken, which results in a monthly time-series for each quintile for the respective measure. We report the mean of these time-series in this table. Performance gap figures are significantly positive at a 5% level for all quintiles.

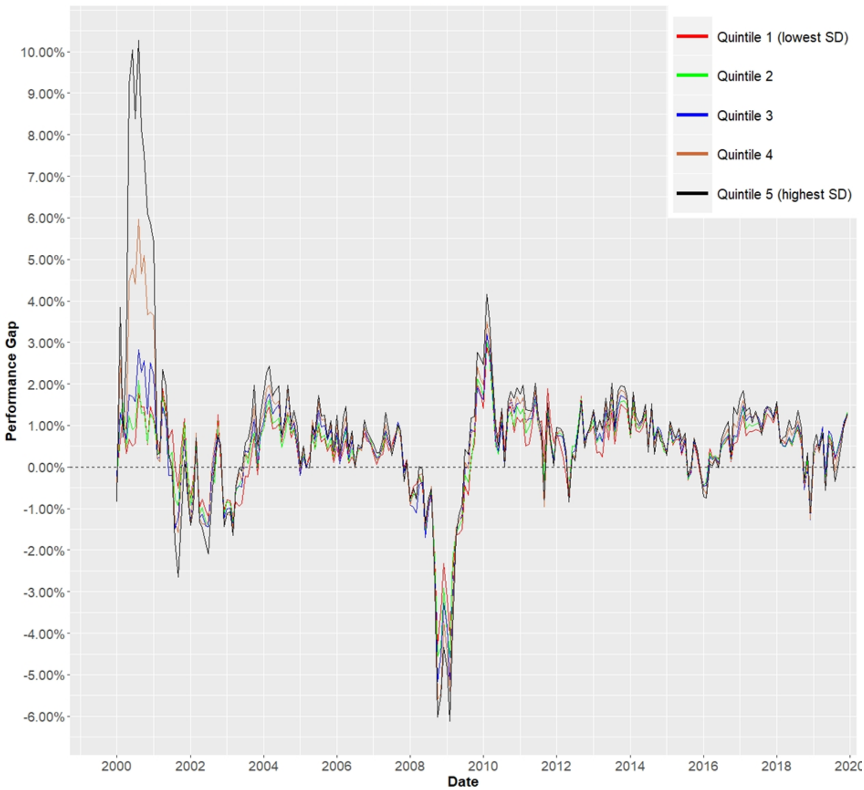
Table 19 showed that a higher standard deviation leads to a greater performance gap, regardless of the calculation basis of money-weighted returns. Quintile 5, consisting of the funds with the highest standard deviation in each respective month, also has the most positive performance gap on average (0.38% according to IRR and 0.8% according to MIRR). By contrast Quintile 1 yields an average performance gap of 0.1% and 0.4% according to calculations based von IRR, respectively MIRR. Considering time-weighted and money-weighted returns, the analysis is not as clear as for the performance gap. The lowest quintile, for example, shows higher or almost equal return measures compared to the following three quintiles. Only the fifth quintile definitely reveals that a higher risk leads to higher returns, on average. As mentioned before, simple averages do not provide full insight into all possible information. Thus, Figures 6 and 7 show the IRR- and MIRR-based performance gaps for each quintile over time.

Figure 6: Performance gap (IRR) for quintiles of standard deviation (twelve-months rolling)



Notes: The figure shows the rolling twelve-month performance gap calculated via IRR time-weighted returns for quintiles of standard deviation. Each observation contains the TNA, fund flows and monthly returns for the previous twelve months.

Figure 7: Performance gap (MIRR) for quintiles of standard deviation (twelve-months rolling)



Notes: The figure shows the rolling twelve-month performance gap calculated via MIRR time-weighted returns for quintiles of standard deviation. Each observation contains the TNA, fund flows and monthly returns for the previous twelve months.

Again, the dot-com and Great Financial Crisis are the periods in which the performance gap has its most prominent peaks and troughs. This seems reasonable, since these periods of financial turmoil are traditionally accompanied by higher volatility. This further confirms

Hypothesis 1 from the introduction. Following the indications in Table 19, the funds within higher quantiles of standard deviation tend to lead to a performance gap measure that fluctuates more. Thus, funds with the highest standard deviation tend to exhibit a higher performance gap. Especially for the early 2000s, investors return from funds with high standard deviation were not as high as simple time-weighted returns. Hence, the timing performance of the average investor, mainly for these funds with higher volatility, was poor. By contrast, the average investor did well, considering the performance gap right before the great financial crisis. Analyzing negative deviations of the performance gap and therefore investors outperforming simple buy-and-hold return, funds with higher risk seem to have greater potential for generating an even smaller performance gap. Hence, the average investor is rewarded for taking higher risks. However, Quintile 5 shows the most positive performance gap, on average. Therefore, simply investing in the funds with the highest volatility would not have been beneficial for the average investor considering the full sample period, since there were more positive deviations than negative deviations from zero. In summary, a visual inspection suggests that funds within the highest quintile of standard deviation lead to poor timing performance especially during crises. This very same quintile also shows the best timing performance in phases of upswing. Considering the other quintiles, the same pattern holds for almost all periods. Hence, we conclude that investors tend to be unsure about their investment decisions due to high volatility, which leads to them being more likely to avoid a simple buy-and-hold strategy ending in suboptimal behavior.

Analyzing an MIRR-based approach, Figure 7 reveals a similar pattern compared to Figure 6. Nevertheless, it should be pointed out that the performance gap calculated with money-weighted returns based on MIRR shows a higher magnitude. Especially, considering the negative performance gap for the upswing phase immediately before the Great Financial Crisis reveals that good timing-performance for the average investor is higher when money-weighted returns are calculated using MIRR. This is due to the (realistic) reinvestment rate of the respective fund return itself. In terms of standard deviation analysis, the funds with the highest volatility demonstrate the best (the most negative) performance gap and therefore the highest money-weighted return considering the pre-GFC-phase. The same pattern can be seen for all other sub-samples and calculation methods. Thus, and in the interest of readability, we do not report these figures here.

According to Friesen and Sapp (2007), the performance gap is greatest among the largest funds. Since there is always the possibility that our overall results may be driven by large or by small funds, we conducted the same quintile analysis as for standard deviation on the funds' total net assets. Additionally, we use total net assets as a variable of control for the following regressions. Table 20 gives an initial impression of average figures and shows the TNA quintile analysis for all of our panels, based on a twelve-month rolling window. Six- and three-month rolling windows show similar overall results.

For all panels, except for index funds, higher TNA leads to lower time-weighted returns and money-weighted returns based on IRR calculation, on average. Considering money-weighted returns derived via MIRR, there is no clear trend except for the index fund panel. Panel A on its own suggests that the performance gap is the lowest among the largest funds, which contradicts the finding of Friesen and Sapp (2007). Nonetheless, one has to keep in mind, that our figures are not directly comparable with those of Friesen and Sapp, mainly due to the length of the period used for calculations. Taking a look at the sub-samples, the same trend is apparent for all panels, except for institutional funds. Thus, fund size seems to have no influence on the performance gap of an institutional fund. This is due to the fact that regardless of the size of an institutional fund, the average investor in these funds is sophisticated or simply follows other investment strategies like plain buy-and-hold or holding on to the same investment for a longer period. Once again, we also provide the twelve-month rolling performance gap over time based on IRR and MIRR in Figures 8 and 9 to provide detailed insights.

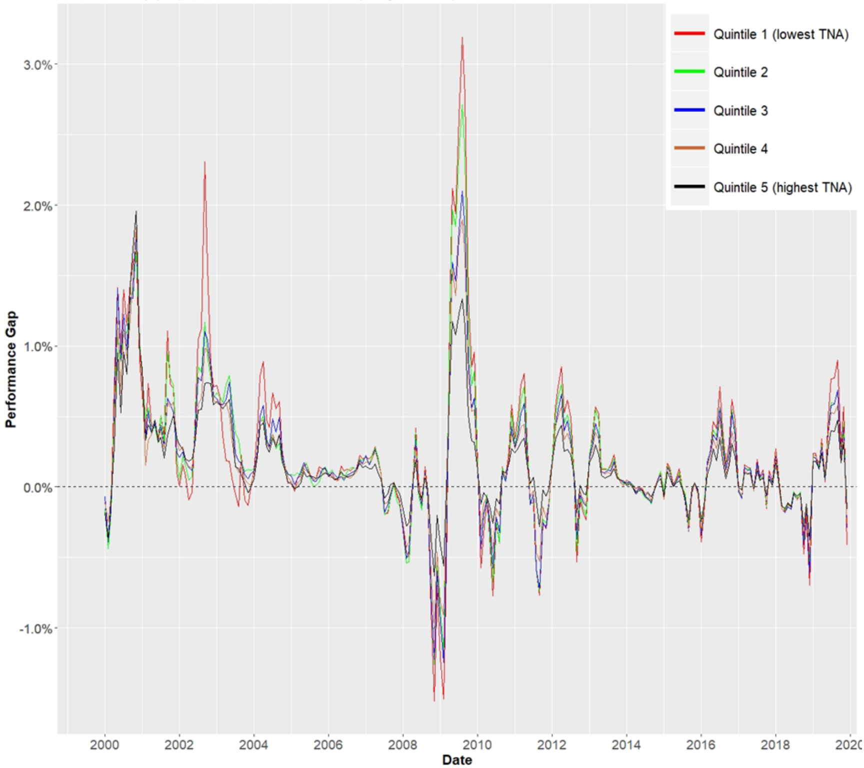
Table 20: Performance gap by funds' total net assets

	Quintile 1 (lowest)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (highest)
Panel A: All Funds					
Avg. TNA ($ millions)	15.37	51.56	135.41	374.73	3560.35
Avg. Geometric return	8.30	8.08	8.01	7.93	7.81
Avg. IRR	7.92	7.70	7.70	7.63	7.59
Avg. MIRR	7.16	7.04	7.12	7.08	7.09
Avg. Performance Gap IRR	0.24	0.22	0.19	0.18	0.17
Avg. Performance Gap MIRR	0.64	0.57	0.50	0.49	0.45
Panel B: Non-index funds					
Avg. TNA ($ millions)	14.94	49.34	127.60	348.72	2979.12
Avg. Geometric return	8.31	8.10	8.03	7.94	7.81
Avg. IRR	7.92	7.70	7.70	7.62	7.56
Avg. MIRR	7.16	7.04	7.11	7.05	7.05
Avg. Performance Gap IRR	0.24	0.22	0.20	0.19	0.18
Avg. Performance Gap MIRR	0.64	0.58	0.51	0.50	0.46
Panel C: Index funds					
Avg. TNA ($ millions)	25.51	107.39	307.09	873.16	10195.56
Avg. Geometric return	8.01	7.62	7.86	7.79	7.94
Avg. IRR	7.86	7.45	7.77	7.70	7.83
Avg. MIRR	7.14	6.99	7.28	7.32	7.49
Avg. Performance Gap IRR	0.15	0.12	0.09	0.09	0.08
Avg. Performance Gap MIRR	0.63	0.40	0.42	0.39	0.30
Panel D: Institutional funds					
Avg. TNA ($ millions)	18.07	66.69	171.33	431.22	2816.77
Avg. Geometric return	8.46	8.38	8.28	8.02	7.92
Avg. IRR	8.00	8.01	8.01	7.80	7.71
Avg. MIRR	7.28	7.32	7.35	7.16	7.15
Avg. Performance Gap IRR	0.21	0.17	0.18	0.16	0.14
Avg. Performance Gap MIRR	0.67	0.67	0.60	0.63	0.55
Panel E: Retail funds					
Avg. TNA ($ millions)	15.19	49.72	130.94	376.70	4013.30

Avg. Geometric return	8.18	7.99	7.89	7.87	7.77
Avg. IRR	7.84	7.62	7.57	7.56	7.52
Avg. MIRR	7.07	6.97	7.00	7.02	7.06
Avg. Performance Gap IRR	0.24	0.23	0.19	0.19	0.18
Avg. Performance Gap MIRR	0.63	0.54	0.48	0.45	0.40

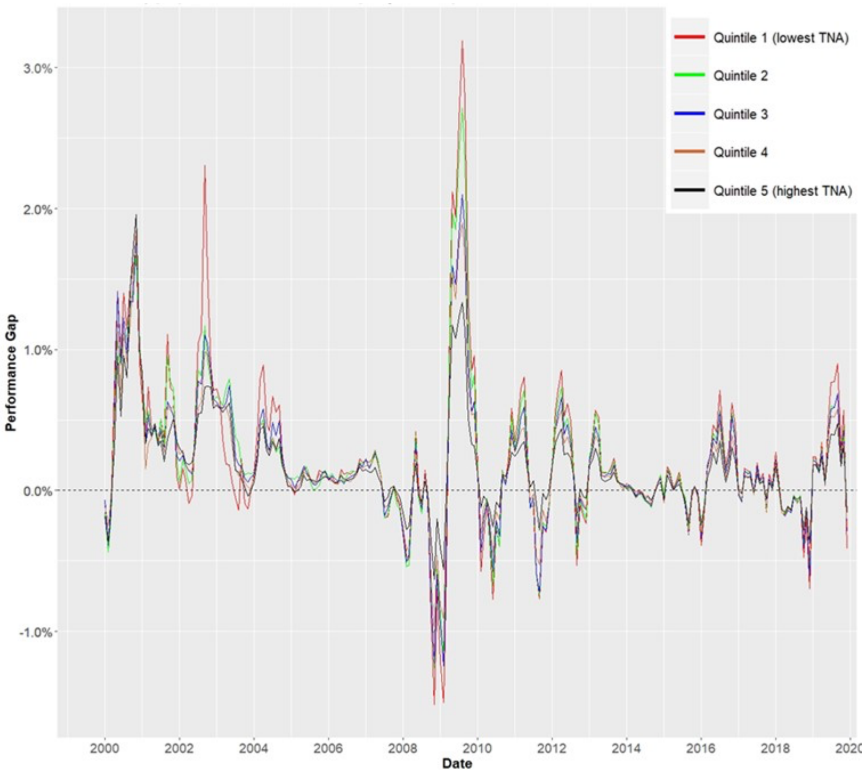
Notes: All return measures as well as the performance gap are annualized and in percent. All funds are sorted into monthly quintiles based on their fund size (TNA). Thus, we know the funds that are in the particular quintiles for each month, including their returns and performance gap. The average for each of these monthly quintiles is used which results into a monthly time-series for each quintile for the respective measure. We report the mean of these time-series in this table. Performance gap figures are significantly positive at a 5% level for all quintiles.

Figure 8: Performance gap (IRR) for quintiles of TNA (twelve-months rolling)



Notes: The figure shows the rolling twelve-months performance gap calculated via IRR time-weighted returns for quintiles of TNA. Each observation contains the TNA, fund flows and monthly returns for the previous twelve months.

Figure 9: Performance gap (MIRR) for quintiles of TNA (twelve-months rolling)



Notes: The figure shows the rolling twelve-months performance gap calculated via MIRR time-weighted returns for quintiles of TNA. Each observation contains the TNA, fund flows and monthly returns for the previous twelve months.

Again, the crises of the last decades are clearly evident in both figures. Overall, smaller funds demonstrate a performance gap measure that fluctuates more, so that an investment in larger funds led to a smaller gap, considering only positive observations. By contrast, smaller funds are the quintile with the most positive performance gap. This pattern has been very consistent since the early 2000’s. Taking a look at negative observations, the opposite effect is apparent. The average investor had better timing performance considering a fund within smaller quintiles of TNA. The following regression analysis yields detailed insights into actual drivers of the performance gap.

4.5 Empirical Results

This section contains the Fama-MacBeth regressions results which help to identify potential drivers of the performance gap. Considering the behavior of the performance gap during times of crises, we included a dummy variable in our model, indicating when a crisis took place. Therefore, we based the indication of a crisis on the official National Bureau of Economic Research (NBER) business cycle data, stating the month of turning points. Thus, during phases of contraction, our dummy variable indicates a crisis. There are two crises within the full sample period, the bursting of the dot-com bubble and the Great Financial Crisis.

As mentioned in Hypothesis 2, we test whether the average investor in the categories of our sample such as index, institutional and retirement funds shows better or worse timing-performance using dummy variables. Additionally, we control in our model for typical performance characteristics like fund age, size, expense ratio, net cash flow and standard deviation. A typical issue resulting from investor irrationality leads to buying the funds with the highest past returns (Ippolito, 1992; Sirri and Tufano, 1998; Del Guerico and Tkac, 2002) whilst not withdrawing money from the worst performing funds, to a degree that would minimize their losses. Hence, we include the total return of the fund as control variable. Barber et al. (2005) demonstrate that investment costs such as load fees and operating expenses also affect mutual fund flows. According to Jain and Wu (2000), advertising of funds may be considered as an influence of these flows as well. Thus, we also analyze whether the total expense ratio has an influence on the performance gap. Navone and Pagani (2015) state that investors in larger and older funds demonstrate better timing-skills. Thus, they reason that investors with more information make better timing decisions concerning their fund flows. Therefore, we include age and size of the fund as dependent variables and check for their influence on timing-performance. Since we provide results for the twelve-month rolling window, all of these (time-varying) independent variables are calculated on the same rolling basis.[15]

The Fama and MacBeth (1973) approach accounts for some of the typical issues concerning financial data. Since heteroskedasticity, serial correlation as well as cross-sectional dependence are an issue in our sample, we have to implement ways to fix these typical problems.[16] The model is set up in three steps. In a first step, time-series regressions are conducted for each fund in the sample, including the independent as

[15] Results for the three- and six-month rolling windows yield results that are in line with the twelve-month rolling window. These results are available upon request.

[16] In accordance with Baltagi (2005), the Pesaran CD-Test for panels with $T < N$ was conducted to test for cross-sectional dependence. In addition, the Breusch-Godfrey test was conducted to test for serial correlation in the idiosyncratic errors, as illustrated in Wooldridge (2010).

well as the dependent variables. The coefficients of each of these regressions are stored. In the second step, for each single time-period a cross-sectional regression, is performed using the coefficients from step one. Doing so produces slope coefficients as well as the coefficient for the intercept. Then, in the third step, the final coefficient estimates are obtained as the respective averages of the second step coefficient estimates. According to Petersen (2009), the Fama and MacBeth approach is designed to account for time effects. Thus, the approach already controls for cross-sectional dependence, but not for time-series dependence which might, in its original form, lead to biased estimates of standard errors. Fama-MacBeth make assumptions about a lack of serial correlation in the standard errors, which is legitimate concerning, especially for stock and fund data with short durations between observations. In order to circumvent these issues, heteroskedasticity and autocorrelation-consistent Newey and West (1987) standard errors are provided. These standard errors handle autocorrelation up to and including a specific lag. Following Greene (2012), we set the lag order to the integer part of $T^{1/4}$, where T represents the maximum order of months in the panel data set.[17]

Tables 21 and 22 show the regression results for the rolling window of twelve months, considering the performance gap as the dependent variable. Model 1 is the basic model without dummy variables

[17] Checks with lag order of two and six are qualitatively robust concerning standard errors.

concerning Hypothesis 2, distinguishing between different investor types and their respective performance gap. Each further model includes a dummy variable for these respective investor clienteles. Furthermore, Table 21 shows the regression results for the performance gap based on IRR calculations, whereas Table 22 shows the results based on the MIRR approach.

Table 21: Fama-MacBeth regression results with performance gap based on IRR

	Performance Gap based on IRR				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	0.1157***	0.1156***	0.1155***	0.1156***	0.1157***
	(3.89)	(3.88)	(3.88)	(3.99)	(3.87)
TNA	-0.0325	-0.0328	-0.0322	-0.0326	-0.0323
	(-0.43)	(-0.45)	(-0.42)	(-0.43)	(-0.42)
Stand. Dev.	0.0354*	0.0354*	0.0357**	0.0353*	0.0355**
	(1.94)	(1.96)	(1.98)	(1.97)	(1.98)
Crisis	0.0462**	0.0463**	0.0465**	0.0464**	0.0465**
	(2.11)	(2.12)	(2.13)	(2.11)	(2.13)
Flows	0.0248*	0.0247*	0.0243	0.0249*	0.0242
	(1.67)	(1.66)	(1.64)	(1.66)	(1.64)
Return	0.0256**	0.0258**	0.0256**	0.0256**	0.0257**
	(1.98)	(2.00)	(2.01)	(2.02)	(1.99)
TER	0.0227*	0.0230*	0.0231*	0.0229*	0.0228*
	(1.73)	(1.76)	(1.77)	(1.76)	(1.75)
Age	0.0019*	0.0023*	0.0020*	0.0021*	0.0022**
	(1.88)	(1.90)	(1.89)	(1.89)	(1.90)
Index		-0.0424**			
		(-1.99)			
Retail			0.0317**		
			(2.06)		
Institutional				-0.0032	
				(-0.68)	
Retirement					-0.0217*
					(-1.82)
R^2	0.1128	0.1130	0.1132	0.1130	0.1129
adj. R^2	0.1112	0.1113	0.1118	0.1112	0.1114

Notes: The table shows the results for the Fama-MacBeth regressions conducted on the twelve-month rolling performance gap as dependent variable. Performance gap was calculated with money-weighted returns according to IRR. Performance gap, standard deviation, return and TER are annualized and in percent. TNA and flows are given in Mio

\$. Standard errors are corrected according to Newey-West (1987) with a lag order of four, and t-statistics are reported in parenthesis below the respective estimation coefficient. The superscripts $***$, $**$ and $*$ denote the rejection of the null hypothesis regarding parameter insignificance at 1%, 5% and 10% significance levels, respectively.

The models in Table 21 demonstrate that investors in index funds as well as retirement funds experience a performance gap that is significantly negative, indicating overall better timing-performance of the average investor compared to for example retail funds. This finding is in line with the overall statistics presented above. Since index funds are a more passive kind of investment, it seems reasonable that the average investor rather follows a passive approach, without plenty of actions involving market-timing et cetera, at least after the initial investment in these passive vehicles. Hence, there are fewer possibilities for receiving a money-weighted return that differs greatly from a simple buy-and-hold return. Though the negative performance gap of these funds is significantly different from zero, it is not of large magnitude, which also supports the thought of index and retirement funds following a buy-and-hold like strategy. Considering the proxy for retail investors, this type of investor is associated with a higher performance gap. This finding fits into the picture of rather unsophisticated retail investors being more prone to irrational behavior and therefore investing at the wrong time. However, contradicting Hsu et al. (2016), we do not find statistically significant evidence of institutional fund investors having superior timing-performance. This might be due to our separation of retirement share classes and the sample of institutional funds. As mentioned, investments in retirement funds tend to follow a plain buy-and-hold strategy, thus being influenced less by investment judgement. Therefore, including these funds in the sample of institutional funds might bias the results. The dummy-variable for crisis shows that the average investor timing-performance is significantly poor during crisis periods. The standard deviation of fund returns also hints at this relationship, indicating that higher volatility of fund returns leads to poor timing-performance. This is in line with the quintile analyses conducted above and confirms Hypothesis 1, which states that the performance gap increases in phases with higher volatility and thus is especially prominent during times of financial turmoil. Such a performance can be justified by the average investor reacting rather irrational when volatility is high.

In accordance with the quintile analysis of fund size, the estimate of TNA is rather small with an effect indistinguishable from zero. Thus, bigger funds do not necessarily attract unsophisticated investors leading to worse timing-performance. Our quintile fund size analysis showed that in phases of financial turmoil like the GFC, smaller funds tend to have the poorest timing performance. Conversely, in phases of financial upswing, smaller funds show the best timing outperformance. Considering the other control variables, better performing funds, according to their monthly return as well as older funds seem to have a higher performance gap calculated via IRR, on average. These results confirm aspects of the findings of Friesen and Sapp (2007). Thus, funds with better performance as well as older and therefore well-known funds might attract rather unsophisticated investors which results in a larger performance gap. Analyzing from a mathematical point of view, investing in a fund with high returns is associated with more potential for a larger performance gap. Additionally, a higher total expense ratio seems to attract an investor clientele with poor timing-performance. The estimates indicate that more expensive funds are associated with timing underperformance. Therefore, the average investor is basically penalized twice, first from higher total costs for the respective fund and second, from poor cash flow timing and choice of magnitude of flows, which all ultimately leads to lower performance compared to simple buy-and-hold return. Typically, funds with higher total expense ratio tend to spend more money on marketing, which might also attract rather unsophisticated investors. The flows of a fund are at times positively correlated with the performance gap, which contradicts Friesen and Sapp (2007) in part. According to them, the fact that flow has no significant influence on the performance gap suggests that the overall rate of non-investment growth of the fund is irrelevant to the timing performance of the average investor.

Fama-MacBeth regression results for the performance gap calculated, with money-weighted returns based on MIRR, yield partially different indications. Table 22 illustrates the findings.

Considering the highly significant and positive crisis dummy, the performance gap based on MIRR calculations also reveals that the performance gap is more prominent in times of crisis. This finding is supported by investments in funds with greater standard deviations in returns, which are positively correlated with poor timing-performance. The empirical evidence is thus consistent with our Hypothesis 1, as in the case of IRRs.

Once again, the dummy variables for the different investor clienteles are consistent with Hypothesis 2 that the performance gap is smaller for investors following a more passive investment strategy. Investors in index funds display better overall timing-performance. The same relationship is apparent when looking at the coefficient for the retirement dummy. However, the negative correlation with the performance gap is not statistically significant concerning retirement funds. Nevertheless, these findings are in favor of passive investment strategies leading to overall better returns for the investor. Looking at the average performance of retail and institutional investors, evidence indicates that these groups tend to attract an investor clientele with a higher performance gap. As mentioned in the introduction, even institutional investors are prone to irrational investment decisions, which might lead to a positive correlation between investors in institutional funds and the performance gap based on time-weighted returns according to MIRR. There are several behavioral patterns like overconfidence, overestimation of skill and selective judgement (e.g. Griffin and Tversky, 1992; Odean, 1998) that lead to investors being return-chasing and taking action at the wrong time concerning their investment. For example, Greenwood and Shleifer (2014) argue that the expectations of future returns of an average investor are correlated with past returns and with the current level of the stock market. As Jegadeesh and Titman (2011) demonstrate, a momentum-effect still prevails and is used by investors. According to Grinblatt and Han (2005) as well as the prospect theory, investors hold on to losing stocks for too long because they do not want to realize losses. Having said that, the average investor also tends to sell highly performing stocks too early in order to take profits. Our findings concerning the performance gap clearly reflect this behavior.

Table 22: Fama-MacBeth regression results with performance gap based on MIRR

	Performance Gap based on MIRR				
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Constant	0.2433***	0.2431***	0.2431***	0.2433***	0.2432***
	(4.86)	(4.87)	(4.86)	(4.87)	(4.87)
TNA	-0.0453	-0.0455	-0.0453	-0.0454	-0.0456
	(-0.71)	(-0.73)	(-0.72)	(-0.71)	(-0.73)
Stand. Dev.	0.0422**	0.0421**	0.0425**	0.0423**	0.0423**
	(2.01)	(1.99)	(2.03)	(2.04)	(2.04)
Crisis	0.0552**	0.0549**	0.0554**	0.0553**	0.0552**
	(2.43)	(2.42)	(2.44)	(2.43)	(2.43)
Flows	0.0317*	0.0318**	0.0319**	0.0318*	0.032**
	(1.95)	(1.97)	(1.97)	(1.96)	(1.98)
Return	0.0288**	0.0286**	0.0291**	0.0290**	0.0287**
	(2.21)	(2.20)	(2.23)	(2.22)	(2.21)
TER	0.0319*	0.0320*	0.0319*	0.0319*	0.0320*
	(1.88)	(1.89)	(1.89)	(1.88)	(1.88)
Age	0.0031**	0.0033**	0.0030**	0.0033*	0.0032*
	(2.01)	(2.02)	(2.01)	(2.03)	(2.03)
Index		-0.0632**			
		(-2.12)			
Retail			0.0396**		
			(2.41)		
Institutional				0.0074*	
				(1.68)	
Retirement					-0.0112
					-(1.53)
R^2	0.1093	0.1095	0.1096	0.1094	0.1094
adj. R^2	0.1078	0.1078	0.1080	0.1081	0.1079

Notes: The table shows the results for the Fama-MacBeth regressions conducted on the twelve-month rolling performance gap as dependent variable. Performance gap was calculated with money-weighted returns according to MIRR. Performance gap, standard deviation, return and TER are annualized and in percent. TNA and flows are given in Mio

\$. Standard errors are corrected according to Newey-West (1987) with a lag order of four, and t-statistics are reported in parenthesis below the respective estimation coefficient. The superscripts $***$, $**$ and $*$ denote the rejection of the null hypothesis regarding parameter insignificance at 1%, 5% and 10% significance levels, respectively.

Once again, older funds and those with high returns are more prone to poor timing-performance. Thus, good performing and older funds seem to attract investors with poor timing-performance. A higher total expense ratio is also positively correlated with the performance gap. Therefore, the double penalization as mentioned earlier, is also present when calculating a performance gap based on MIRR instead of IRR. As in Table 21, fund size has no significant influence on the performance gap of the average investor, based on MIRR calculations. Hence, the same interpretations as before can be applied. Rising net cash flows lead to timing underperformance in all of the models. Thus, funds with high inflows seem to attract investors with poor timing and magnitudes of these flows. Differences in significance and magnitude of coefficients of the IRR and MIRR models might also be due to the elimination of unrealistic assumptions like the reinvestment assumption for IRR, which biases returns and therefore have an impact on the momentum or return-chasing effect.

Our analysis so far includes results and the implications of money-weighted return calculations based on IRR and a dynamic version of MIRR, which is prone to subjectivity as mentioned above. Future research might incorporate other and possibly even more realistic approaches to measuring the actual return an investor receives. Besides the modified and plain internal rate of return, there are several other ways of calculating a performance gap. Future research on the performance gap might also include the implementation of different panel regression methods like hybrid or mixed effects models, as well as Hausman-Taylor instrumental variable models. Our analysis has focused on the timing ability of the average investor in US mutual equity funds. Due to their recent popularity, examining the timing performance for passive vehicles like exchange-traded funds might yield interesting results, especially since we find that investors in index funds show better timing-performance. Nonetheless, their buy-and-hold nature seems to leave little room for active investment decisions which would probably result in a better overall timing-performance. Additionally, we were able to build subsamples of institutional and retail share classes based on fund characteristics, and use them as a proxy for the respective net cash flow. Having access to the actual net cash flows for each of these groups as in Clare and Motson (2010), might yield more detailed implications. However, estimating fund flows is common praxis (see Sirri and Tufano, 1998).

4.6 Concluding Remarks

The present study contributes to the body of finance literature by extending the analysis of overall differences of buy-and-hold and actual investor fund return, with a study of different measures of money-weighted returns, as well as a more detailed approach to illustrating when timing out- and underperformance takes place. A simple buy-and-hold approach ignores the fact that many investors trade on a regular basis, and that the money-weighted return the average investor receives might be different to the performance of the underlying fund itself. Through rolling window calculations, we show when timing performance is superior or poor. Providing these detailed insights reveals that timing underperformance is not a permanent issue for our sample, but definitely present on average. Overall, we find an annualized performance gap of 0.20 % according to IRR and 0.52%, based on MIRR calculations for the sample of all funds. Since this annual performance gap is an average value, some investors have experienced better or worse performance throughout their personal investment history. Nevertheless, we find some periods of timing-outperformance enhancing returns of the average investor. Considering the upswing right before the Great Financial Crisis, the average investor was able to outperform a simple buy-and-hold return. These phases of good timing-performance are rather rare. Especially during crisis periods when volatility is high, timing underperformance is prevalent and reaches peaks of up to 4.20%. Thus, following a simple buy-and-hold or passive strategy yields better overall returns for the average investor, at least during phases of crisis. These findings are not only significant for the full sample of funds, but are also found to be robust across all sub-samples.

Dividing the sample into index, non-index as well as retail and institutional funds shows that investors in index funds tend to have a lower timing underperformance compared to the average investor in non-index funds, which is in line with the theory on passive investment vehicles. The performance gap according to IRR calculations is higher for retail funds compared to institutional funds, which suggests poor timing performance due to a lack of sophistication of the respective investor. Considering the MIRR-calculation approach, institutional funds have a higher on average performance gap than retail funds, which suggests that institutional investors are also prone to irrational behavior, as discussed in the introduction.

Additionally, we demonstrate that two different calculation methods of money-weighted returns, e.g. internal rate of return and modified internal rate of return, lead to partially different implications concerning the performance gap. As the literature review has shown, calculating the performance gap on the basis of modified internal rate of return follows a more realistic approach compared to IRR. Thus, these money-weighted returns state the actual returns the average investor receives. We also show that overall timing underperformance is even higher when considering the performance gap calculation via MIRR, irrespective of the length of the rolling window calculations. However, we introduce subjectivity into the calculation of money-weighted returns with the MIRR approach. Thus, we refrain from explicitly preferring one calculation method to the other.

We conduct regression analyses to find potential drivers of the performance gap and to gain further insights into the timing-performance of different investor clienteles following various investment strategies. Variables like the standard deviation of returns, total expense ratio or the age of a fund are typical drivers of the performance gap. We were able to show that a positive correlation of timing-underperformance and phases with economic downturns exist, irrespective of the actual calculation method of money-weighted returns. Additionally, the dummy variable indicating financial turmoil is highly significant throughout every regression. All of the findings above confirm our Hypothesis 1 of timing-performance being especially poor when standard deviation is high, thus during times of crises.

The models show that investments in index funds yield a lower performance gap, whilst the timing performance of an average investor in retail funds tends to be poor. The same effect is apparent for institutional funds in models with a performance gap based on MIRR money-weighted returns. We were not able to find institutional investors with overall better timing-performance, irrespective of the calculation method. In concordance with our Hypothesis 2, this shows that different investor clienteles are prone to irrational behavior, leading to poor timing-performance. Overall, our results suggest that the average investor misses the optimal point of reinvestment after the peak of the crises, and stays invested for too long right before the start of an economic downturn. Thus, simple buy-and-hold strategies outperform market-timing strategies most of the time. Especially during times of crisis, the average investor should not react at all and rather follow such a buy-and-hold approach. As the literature review has shown, all investors, if professional or not, are subject to biases and typical irrational behavior, leading to unnecessary losses in their actual return. Some suffer more from the impact of wrong timing, others have a lower performance gap dependent on the investor clientele to which they belong. Additionally, investors should consider changing their investment behavior towards passive strategies involving lower fees and requiring less action. Hence, they can avoid losing returns due to their own irrationality and falsely timed investment decisions, as the sub-samples for index and retirement share classes demonstrate.

4.7 Bibliography

Ammann, M., Bauer, C., Fischer, S., & Muller, P. (2018). The Impact of the Morningstar Sustainability Rating on Mutual Fund Flows. European Financial Management, 25(3), pp. 520-553.

Andrew, C., & Nick, M. (2010). Do UK Retail Investors Buy at the Top and Sell at the Bottom? Working Paper.

Baltagi, B. H. (2005). Econometric Analysis of Panel Data. Chichester: John Wiley & Sons.

Barber, B., Odean, T., & Zheng, L. (2005). Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows. The Journal of Business, 78(6), pp. 2095-2120.

Bayleat, B. R., Cagle, J., & Glasgo, P. (2013). Teaching MIRR to Improve Comprehension of Investment Performance Evaluation Techniques. Journal of Economics and Finance Education, 12(1), pp. 39-50.

Black, F. (1986). Noise. The Journal of Finance, 41(3), pp. 528-543.

Bollen, N. P., & Busse, J. A. (2002). On the Timing Ability of Mutual Fund Managers. The Journal of Finance, 56(3), pp. 1075-1094.

Braverman, O., Kandel, S., & Wohl, A. (2005). The (Bad?) Timing of Mutual Fund Investors. CEPR Discussion Papers, 5243.

Cashman, G. D., Nardari, F., Deli, D. N., & Villupuram, S. V. (2014). Investor Behavior in the Mutual Fund Industry: Evidence from Gross Flows. Journal of Economics and Finance, 38(4), pp. 541-567.

Cheremushkin, S. (2012, January). There is No Hidden Reinvestment Assumption in Discounting Formula and IRR: Logical and Mathematical Arguments. SSRN.

Chieh-Tse Hou, T. (2012). Return Persistence and Investment Timing Decision in Taiwanese Domestic Equity Mutual Funds. Managerial Finance, 38(9), pp. 873–891.

Cuthbertson, K., Nitzsche, D., & O'Sullivan, N. (2010). The Market Timing Ability of UK Mutual Funds. Journal of Business Finance & Accounting, 37(1-2), pp. 270-289.

Del Guercio, D., & Tkac, P. (2002). The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds. Journal of Financial and Quantitative Analysis, 37(4), pp. 523-557.

Dellva, W. L., DeMaskey, A. L., & Smith, C. A. (2001). Selectivity and Market Timing Performance of Fidelity Sector Mutual Funds. The Financial Review, 36(1), pp. 39-54.

Dichev, I. D. (2007). What Are Stock Investors’ Actual Historical Returns? Evidence from Dollar-Weighted Returns. American Economic Review, 97(1), pp. 386-401.

Dichev, I. D., & Yu, G. (2011). Higher Risk, Lower Returns: What Hedge Fund Investors Really Earn. Journal of Financial Economics, 100(2), pp. 248-263.

Dichtl, H., & Drobetz, W. (2011). Portfolio Insurance and Prospect Theory Investors: Popularity and Optimal Design of Capital Protected Financial Products. Journal of Banking & Finance, 35(7), pp. 1683-1697.

Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. The Journal of Political Economy, 81(3), pp. 607-636.

Fisher, K. L., & Statman, M. (2002). Blowing Bubbles. Journal of Psychology and Financial Markets, 1, pp. 53-65.

Frazzini, A., & Owen, L. (2008). Dumb Money: Mutual fund flows and the Cross-Section of Stock Returns. Journal of Financial Economics, 88(2), pp. 299-322.

Freiburg, M., & Grichnik, D. (2013). Institutional Reinvestments in Private Equity Funds as a Double-Edged Sword: The Role of the Status Quo Bias. Journal of Behavioral Finance, 14(2), pp. 134-148.

Friesen, G. C., & Sapp, T. R. (2007). Mutual fund flows and Investor Returns: An Empirical Examination of Fund Investor Timing Ability. Journal of Banking and Finance, 31(9), pp. 2796–2816.

Greene, W. H. (2012). Econometric Analysis. New York: Pearson.

Greenwood, R., & Shleifer, A. (2014). Expectations of Returns and Expected Returns. Review of Financial Studies, 27(3), pp. 714-746.

Griffin, D., & Tversky, A. (1992). The Weighing of Evidence and the Determinants of Confidence. Cognitive Psychology, 24(3), pp. 411-435.

Grinblatt, M., & Han, B. (2005). Prospect Theory, Mental Accounting, and Momentum. Journal of Financial Economics, 78(2), pp. 311-339.

Grinblatt, M., & Keloharju, M. (2000). The Investment Behavior and Performance of Various Investor Types: A Study of Finland's Unique Data Set. Journal of Financial Economics, 55(1), pp. 43-67.

Gruber, M. J. (1996). Another Puzzle: The Growth in Actively Managed Mutual Funds. The Journal of Finance, 51(3), pp. 783-810.

Hirshleifer, D., & Shumway, T. (2003). Good Day Sunshine: Stock Returns and the Weather. The Journal of Finance, 58(3), pp. 1009-1032.

Hsu, J. C., Myers, B. W., & Withby, R. J. (2016). Timing Poorly: A Guide to Generating Poor Returns While Investing in Successful Strategies. Journal of Portfolio Management, 42(2), pp. 90-98.

Hurley, T. M., Rao, X., & Pardey, P. G. (n.d.). Re-Examining the Reported Rates of Return to Food and Agricultural Research and Development. American Journal of Agricultural Economics, 96(5), pp. 492–504.

Ingersoll, J. E., Spiegel, M. E., Goetzmann, W. N., & Welch, I. (2007). Portfolio Performance Manipulation and Manipulation-Proof Performance Measures. The Review of Financial Studies, 20(5), pp. 1503-1546.

Ippolito, R. A. (1992). Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry. Journal of Law and Economics, 35(1), pp. 45-70.

Jain, P. C., & Wu, J. S. (2000). Truth in Mutual Fund Advertising: Evidence on Future Performance and Fund Flows. The Journal of Finance, 55(2), pp. 937-958.

James, C., & Karceski, J. (2006). Investor monitoring and differences in mutual fund performance. The Journal of Banking and Finance, 30(10), pp. 2787 - 2808.

Jegadeesh, N., & Titman, S. (2011). Momentum. Annual Review of Financial Economics, 3, pp. 493-509.

Johnston, K., Forbes, S., & Hatem, J. J. (2002). Reinvestment Rate Assumptions in Capital Budgeting: A Note. Journal of Economics and Finance Education, 1(2), pp. 22-23.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), pp. 263-291.

Keef, S., & Rosuh, M. (2001). Discounted Cash Flow Methods and the Fallacious Reinvestment Assumption: A Review of Recent Texts. Accounting Education, 10, pp. 105-116.

Keim, D. B., & Madhavan, A. (1995). Anatomy of the Trading Process Empirical Evidence on the Behavior of Institutional Traders. Journal of Financial Economics, 37(3), pp. 371-398.

Keswani, A., & Stolin, D. (2008). Which Money Is Smart? Mutual Fund Buys and Sells of Individual and Institutional Investors. The Journal of Finance, 63(1), pp. 85-118.

Kierulff, H. (2008). MIRR: A better measure. Business Horizons, 51, pp. 321-329.

Luo, J.-S., & Li, C.-A. (2008). Futures Market Sentiment and Institutional Investor Behavior in the Spot Market: The Emerging Market in Taiwan. Emerging Markets Finance and Trade, 44(2), pp. 70-86.

Muñoz, F. (2016). Cash Flow Timing Skills of Socially Responsible Mutual Fund Investors. International Review of Financial Analysis, 48, pp. 110-124.

Navone, M., & Pagani, M. (2015). Brothers from Different Mothers: How Distribution Fees Change Investment Behavior. Journal of Banking and Finance, 51(C), pp. 12-25.

Nesbitt, S. L. (1995). Buy High, Sell Low: Timing Errors in Mutual Fund Allocations. The Journal of Portfolio Management, 22(1), pp. 57-60.

Newey, W., West, & Kenneth. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), pp. 703-708.

Odean, T. (1998). Are Investors Reluctant to Realize Their Losses? The Journal of Finance, 53(5), pp. 1775-1798.

Petersen, M. A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. The Review of Financial Studies, 22(1), pp. 435–480.

Phalippou, L. (2008). The Hazards of Using IRR to Measure Performance: The Case of Private Equity. Journal of Performance Measurement, 12.

Rich, S. P., & Rose, J. T. (2014). Re-examining an old question: Does the IRR Method Implicitly Assume a Reinvestment Rate? Journal of Financial Education, 40(1-2), pp. 152-166.

Ross, S. A., Westerfield, R. W., & Jordan, B. D. (2013). Fundamentals of Corporate Finance. New York: McGraw-Hill Irwin.

Sapp, T., & Tiwari, A. (2004). Does Stock Return Momentum Explain the "Smart Money" Effect? The Journal of Finance, 59(6), pp. 2605-2622.

Shiller, R. J. (2016). Irrational Exuberance. New Jersey: Princeton University Press.

Sirri, E. R., & Tufano, P. (1998). Costly Search and Mutual Fund Flows. The Journal of Finance(53), pp. 1589-1622.

Walker, J. S., Check, H. F., & Randall, K. L. (2011). Does the Internal Rate of Return Calculation Require a Reinvestment Rate Assumption? There is still no Consensus. Pennsylvania Economic Review, 18(1), pp. 37-59.

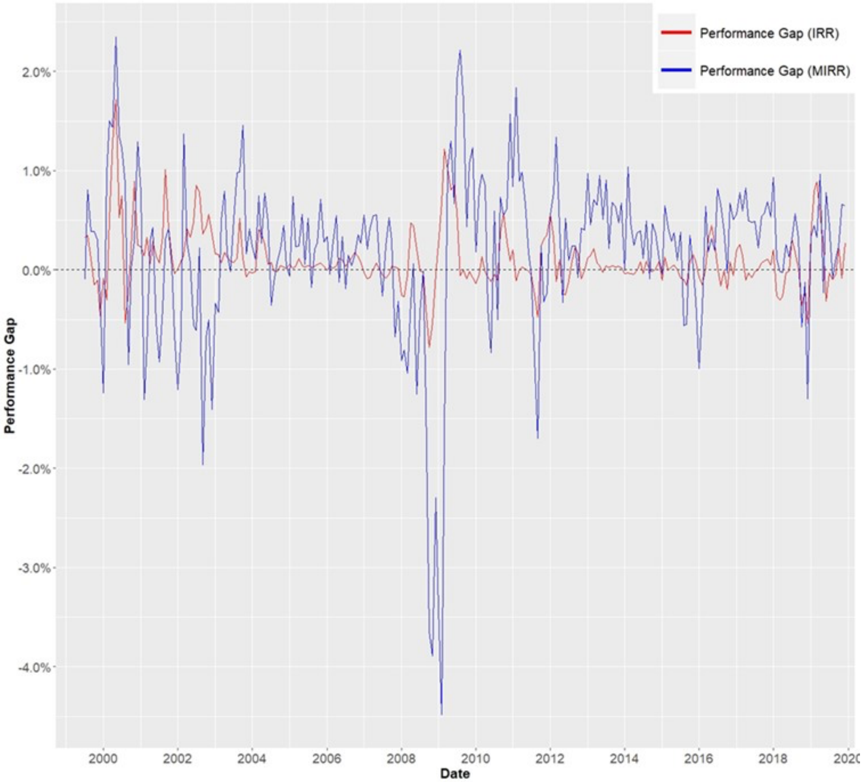
Wooldrige, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. Cambridge: The MIT Press.

Zheng, L. (1999). Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability. The Journal of Finance, 54(3), pp. 901-933.

Zweig, J. (2002, April). Funds that Really Make Money for their Investors. Money Magazine, pp. 124-134.

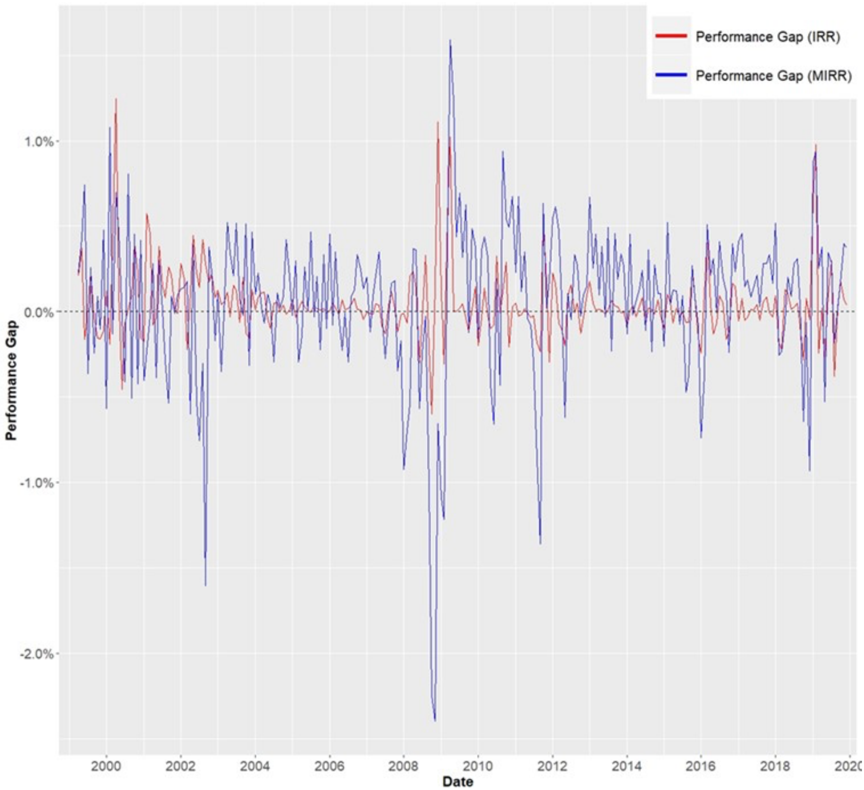
4.8 Appendix

Figure 10: Rolling six-months performance gap based on IRR or MIRR



Notes: The performance gaps are annualized and stated in percent. Bear in mind that due to the six-months rolling window, a performance gap for the time shown in the graph actually consists of TNA, fund flows and returns (money-weighted and time-weighted) for the previous six months. A positive performance gap shows that the average investor return was less than the buy & hold return (Performance Gap = TWR – MWR).

Figure 11: Rolling three-months performance gap based on IRR or MIRR



Notes: The performance gaps are annualized and stated in percent. Bear in mind that due to the three-months rolling window, a performance gap for the time shown in the graph actually consists of TNA, fund flows and returns (money-weighted and time-weighted) for the previous three months. A positive performance gap shows that the average investor return was less than the buy & hold return (Performance Gap = TWR – MWR).

5 Conclusion

5.1 Conclusion

This thesis sheds new light on the dynamic behaviour of risk in financial markets as well as on the outcomes of the (irrational) behaviour of investors trading in these markets. It analyses different cases along the investment experience of any individual, helping retail and institutional investors to have a better understanding of risk and how to cope with it in various situations. Understanding how individuals handle choice under uncertainty and when they invest their money, is not only of great interest for academic research but also economic policy makers since most financial decisions are based on the trade-off between risk and return. This chapter summarizes the insights gained, based on the empirical research presented in the cumulative dissertation. Furthermore, it discusses the broader implications of these findings and provides avenues for future research. Thus, chapter 1 outlines the motivation as well as the (economic) relevance of the outlined topics concerning academia and praxis. Chapters 2 to 4 cover the articles of this cumulative dissertation showing results incorporating research on different asset classes, types of investors as well as underlying models and methodological approaches.

Chapter 2 shows that it is necessary to be conscious of potential risk factors and one's exposure to these risks, in particular the risk of inflation, respectively deflation. This topic is of interest for institutional as well as retail investors. Although, hedging against inflation or deflation might be a rather sophisticated issue conducted mainly by institutional investors. Nevertheless, retail investors might find it interesting to know if their investments, especially direct real estate investments provide hedging capabilities. The empirical findings of Chapter 2 demonstrate that real estate prices and rents are strongly linked to consumer prices. Overall, results for the two countries in the sample are different no matter if the inflationary or deflationary market phase is considered. For the period mainly driven by inflation, real estate is able to provide a hedge against expected and unexpected inflation in some cases. The overall deflationary phase shows that in several instances, real estate is able to provide value stability in real terms and therefore a protection against deflation. However, an effective hedge is not given for each sub-sector at every time.

The research conducted in Chapter 3, back tests the one-day-ahead risk metric forecasting accuracy of the AR-GARCH-EVT-Copula methodology compared to the variance-covariance and the historical simulation methodology for portfolios that contain securitized real estate. The empirical results of the back-test as well as the optical inspection of results demonstrate a general reduction in model hits for the AR-GARCH-EVT-Copula approach in comparison to both benchmarks across various international financial markets. Nevertheless, it is also shown that the findings are subject to the chosen risk metric, since the test results indicate more substantial accuracy enhancements for the $CVaR_\alpha$ than for the VaR_α of the portfolios. The improvements are also subject to the portfolio constituents. The implications of this article are mainly of importance for risk managers in financial institutions as well as financial supervision authorities because the risk exposure of portfolios usually is obliged to stay within predetermined limits. Additionally, precise estimates of risk are important for performance evaluation. In particular, typical approaches like variance-covariance and historical simulation do not account for all of the (empirical) problems concerning the underlying data and therefore do not provide a model for forecasting price risks that is sufficient enough. Thus, risk managers should constantly monitor the risk of their portfolios and adjust holdings if necessary. The AR-GARCH-EVT-Copula model proposed in this article provides a sound approach for these daily monitoring, forecasting and risk assessment purposes.

Chapter 4 presents a paper that aims to illustrate the impact of investor irrationality concerning the returns of mutual funds an average investor actually receives. Calculating the difference of time-weighted and money-weighted returns using rolling-windows, demonstrates when timing-performance is good or poor. According to these calculations it becomes apparent, that periods of economic turmoil like the dot-com bubble or the GFC are phases of timing-underperformance. Nevertheless, it is also shown that calculations are prone to the length of the respective rolling window. Hence, the analysis provides more detail compared to already existent studies, which state a single figure as performance gap, which is based on simple averages over the full available period. Additionally, Fama-MacBeth regressions show that variables like the standard deviation of returns, total expense ratio or the age of a fund are typical drivers of the performance gap. Timing-performance concerning an investment into retail funds seems to be rather poor whereas investments into more passive vehicles like index and retirement funds show a superior timing-performance. Following the results of this article, the conclusion can be drawn that particularly retail investors should reflect their investment behavior and rather follow a simple buy-and-hold approach. Especially during times of crisis when there are plenty of convenient investment opportunities, this might seem unusual. However, the study shows that timing-performance of the average investor is poor in these periods characterized by high volatility of underlying prices.

However, the research conducted in this cumulative dissertation is subject to various research limitations. Considering the individual articles of the thesis, the evidence of the first paper is subject to data limitations. There are only two countries that exhibited long-term phases of deflation whilst delivering high quality data that can be used for research. Additionally, there are minor issues like Hong Kong not having long enough data history considering a government bond index or Japan only providing data on a half-yearly basis. Regarding the methodology, conducting cointegration tests in order to assess long-run hedging capabilities (e.g. Ganesan and Chiang, 1998; Adrangi et al., 2004) and therefore to provide further insights. In combination with Granger causality tests (e.g. Glascock et al., 2008; Obereiner and Kurzrock, 2012) to shed light on short-run dynamics, the empirical and economic results of Chapter 2 could be enhanced further. Additionally, the implications drawn in this chapter are short of linkages between the reason for particular sub-sector being hedge against inflation or deflation and the actual underlying economics and industry of the respective country.

The approach of forecasting price risk via AR-GARCH-EVT-Copula approach is limited by several assumptions. First, bivariate portfolios and therefore so-called pair-copulas are used, although there are methodologies to implement a portfolio consisting of multiple asset classes like the DCC-GARCH-Copula or Dynamic-Mixture-Copula approach including a dynamic, time-varying structure (e.g. Engle, 2002; Capiello et al., 2006; Patton, 2006; Heinen and Valdesogo, 2009; Fermanian and Wegkamp, 2012). Therefore, the latest models allow for multiple assets and enable circumventing the disadvantage of describing the whole dependence through a single parameter. Second, the study equally weights the portfolios of securitized real estate and stocks or bonds. Deriving a time-varying optimal portfolio weight is another option for future research. Third, index-level data is used, which may not describe the reality as is. Finally, utilizing the AR-GARCH-EVT Copula approach of Chapter 3 in the context of portfolio optimization represents yet another subject of future research (Chakkalakal et al., 2018).

Considering the measurement of the performance gap, there are various possibilities how money-weighted returns can be calculated. Approaches like the average internal rate of return (e.g. Magni, 2013) or a mixture of all existent methodologies might proof useful in deriving a more realistic performance gap. Apart from the actual underlying calculation method, finding the drivers of the timing-performance is important. Other than Fama-MacBeth regression, there are several possibilities of setting up a model and identifying the determinants. Fixed or random effects models are another way of analyzing the mutual fund data. It is widely recognized that fixed-effects models have an advantage over random-effects models when analyzing panel data because they control for all cluster-invariant characteristics, measured or unmeasured (Allison, 2009; Halaby, 2004; Wooldridge, 2013). However, such models are facing many drawbacks like random-effects estimators having to fulfill the assumption of all regressors being uncorrelated with the time-constant error term in order to yield unbiased estimates. In order to circumvent these disadvantages, Allison (2009) as well as other authors proposed to estimate within effects in random-effects models. Thus, resulting in a so-called hybrid model, based on a correlated random-effects (CRE) model using a GLS approach, including a between and within-cluster component. All models estimated with this approach would be generalized linear mixed models (GLMMs). These models also go by many other names, e.g. multilevel models, hierarchical models. According to Schunck (2013), hybrid models are a useful extension to the standard random-effects and fixed-effects approaches. Additionally, a hybrid model with variable slopes or even an instrumental-variables approach after Hausman and Taylor (1981) might shed further light on the discussion of statistically significant determinants and therefore actual drivers of timing-performance.

5.2 Bibliography

Allison, P. D. (2009). Fixed Effects Regression Models. Pennsylvania: SAGE Publications.

Andragi, B., Arjun, C., & Kambiz, R. (2004). REIT Investments and Hedging Against Inflation. Journal of Real Estate Portfolio Management, 10(2), pp. 97-112.

Capiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. Journal of Financial Econometrics, 4(4), pp. 537-572.

Chakkalakal, L., Hommel, U., & Li, W. (2018). Transport Infrastructure Equities in Mixed-Asset Portfolios Estimating Risk with a GARCH-Copula CVaR Model. Journal of Property Research, 35(2), pp. 117-138.

Engle, R. (2002). Dynamic Conditional Correlation. Journal of Business & Economic Statistics, 20(3), pp. 339-350.

Fermanian, J.-D., & Wegkamp, M. H. (2012). Time-Dependent Copulas. Journal of Multivariate Analysis, 110, pp. 19-29.

Ganesan, S., & Chiang, Y. (1998). The Inflation-Hedging Characteristics of Real and Financial Assets in Hong Kong. Journal of Real Estate Portfolio Management, 4(1), pp. 55-67.

Glascock, J., Feng, L., Fan, L., & Bao, H. X. (2008). Inflation Hedging Characteristics of Real Estate Assets in Hong Kong. Statistics & Decision(301), pp. 1-25.

Halaby, C. N. (2004). Panel Models in Sociological Research: Theory into Practice. Annual Review of Sociology, 30, pp. 507-544.

Hausman, J. A., & Taylor, W. E. (1981). Panel Data and Unobservable Individual Effects. Econometrica, 49(6), pp. 1377-1398.

Heinen, A., & Valdesogo, A. (2009). Modelling International Financial Returns with a Multivariate Regime-Switching Copula. Journal of Financial Econometrics, 7(4), pp. 437-480.

Magni, C. A. (2013). The Internal-Rate-of-Return Approach and the AIRR Paradigm: A Refutation and a Corroboration. The Engineering Economist, 58(2), pp. 73-111.

Obereiner, D., & Kurzrock, B.-M. (2012). Inflation-Hedging Properties of Indirect Real Estate Investments in Germany. Journal of Property Investment & Finance, 30(3), pp. 218-240.

Patton, A. J. (2006). Modelling Asymmetric Exchange Rate Dependence. International Economic Review, 47(2), pp. 527-556.

Schunck, R. (2013). Within and Between Estimates in Random-Effects Models: Advantages and Drawbacks of Correlated Random Effects and Hybrid Models. Stata Journal, 13(1), pp. 65-76.

Wooldridge, J. (2013). Introductory Econometrics: A Modern Approach. Mason: South-Western College Publishing.