Evaluation of Investment Strategies

Igor Vilcek a, *

a Tatra Asset Management, Bratislava, Slovakia

ABSTRACT

The paper is focused on an ex-post investment performance analysis. Firstly, we create a suite of return, risk, return to risk and benchmark related measures with immediate real life applications in mind. We define and describe these measures and provide real data examples where appropriate. Secondly, we take a look at common mistakes performed by practitioners. For each measure, we assess potential misuses and potential misinterpretations of conclusions resulting from them. We also include the visual analysis. Moreover, we analyze advantages and disadvantages of specific measures and show how to use them appropriately. We furthermore stress that it is necessary to evaluate measures in conjunction with each other. We complement the analysis by brief discussion of non-quantifiable, more qualitative, criteria. Lastly, in the practical section, we apply all performance measures and discussed principles in evaluation of the ten years of historical data of the Slovak equity pension funds in the funded pillar.

Journal of Insurance and Financial Management

ARTICLE INFO

Article History
Submitted 05 Mar 2017
Accepted 18 Mar 2017
Available online 07 Jul 2017

JEL Classification
C00
C10
C50
G00
G11

Keywords
Performance Evaluation
Investment Performance Measurement
Portfolio Performance
Fund Performance Analysis
Risk-Return Analysis
Funded Pillar of the Slovak Pension System

*Corresponding Author:
vilcek.igor@gmail.com

Author(s) retain copyright of the submitted paper (Please view the Copyright Notice of JIFM).
1. Introduction

The number of investment products available for investors nowadays is reaching overwhelming levels and is further growing. It is still more complicated and complex for an investor to assess which investment product is suitable for his investment objective. For a product an investor needs to consider almost endless number of aspects such as the future return potential, an alignment of the product against his investment objective, investment horizon, risks associated with it, costs, threats of a potential misleading marketing as well as historical performance of the same or similar products. The last is perhaps the most easily accessible and quantifiable and represents the foundation for this paper.

Many practitioners and even some investment professionals do not evaluate the performance of investment strategies correctly. It is a very common practice to judge an investment only by means of a price chart of an equity curve of an investment. This, stand-alone, is not sufficient for assessment of investment’s past risk and return characteristics. Moreover, even carrying out a so called visual inspection of the chart is often done incorrectly by analysts. Scaling, frequency of the displayed returns or the length of the charts is often misused and incorrectly interpreted, whether consciously or unconsciously.

Not uncommon is putting too much emphasis on the return side of the strategy. Practitioners very often base their investment decisions only on the past returns and do not put enough focus on all of the risks associated with an investment. It is often forgotten that if a specific risk event such as default is triggered, it may substantially or even completely reduce returns on an investment. For that reason it is of uttermost importance to focus on the return in conjunction with the risk associated and never on the return alone.

Even in case a potential investor conducts a return to risk analysis, it may be insufficient, calculated incorrectly or lacking other important return to risk measures necessary to be taken into account. Many practitioners focus on just one or two return to risk measures. They sometimes use the annualization incorrectly or compare investments between each other in spite of each of them being scaled differently. All of these introduces misuse of a performance evaluation.

Last but not least, it is not enough to assess only quantifiable statistics of an investment themselves. Besides them, one needs to take into account effects difficult to measure such as liquidity, credit quality, currency risk or details regarding investment strategy. It is worth mentioning that these may often only be obtainable from detailed reports or in the worst case
only by means of contacting the investment manager.

This paper attempts to address most of the above mentioned issues. We focus on an ex-post investment performance analysis. This analysis is purely based on the historical price data and does not cover predictions of a future performance, fundamental analysis or analysis of single trades. We thus do not focus on an ex-ante evaluation of a potential strategies but rather on an analysis of existing strategies with long enough track record. Nonetheless, we are aware of the fact that an ex-post analysis of the historical data is not fully sufficient to completely assess a suitability of an investment. Ex-ante analysis, however, is beyond the scope of this paper.

As we are not aware of the relevant literature which describes the ex-post performance measures, provides real data examples, discusses their pros and cons, the common misuses of the measures and which applies the measures again to the real data, we aim to accomplish this.

We will focus on an evaluation of the strategies as such, rather than compared to some benchmark. The reason is twofold. Firstly, most of the investors are usually concerned about the performance of a strategy as such and not about outperformance against some abstract benchmark. Secondly, a choice of the benchmark is often disputable and not always unambiguous. Hence, the majority of the paper will be devoted to the return, risk and return to risk measures which do not involve comparison against benchmark. Nonetheless, because it is still important for many analysts to also assess the performance against some pre-specified benchmark, we do include a section devoted to benchmark related measures. Aim of this paper is not to cover all of the existing measures. The aim is to describe as many different measures as possible which do complement each other and may add value when introduced into the decision making process.

Purpose of this text is as following. Firstly we aim to create a suite of return, risk, return to risk and benchmark related measures suitable for the immediate real life application. We define and describe the measures and attempt to provide real data examples where appropriate. Next, we aim to discuss the common mistakes made by practitioners, the common misuse of the measures and a potential misinterpretation of the resulting values of the measures. We also devote a section to the visual analysis. Moreover, we aim to discuss the advantages and disadvantages of the specific measures and describe the cases in which it is appropriate to use a specific measure. Furthermore, we want to stress that it is necessary to evaluate the measures in conjunction with each other rather then to arrive at conclusions based just on a single measure. We complement the analysis by brief discussion of the criteria which are difficult to
be measured. Last but not least, we apply all of the defined measures and discussed principles to the ten years of historical data of the Slovak equity pension funds in the funded pillar (defined contribution pension scheme).

The paper is organized as follows. The first chapter introduces. The second chapter summarizes the relevant literature. The third chapter focuses on the entire ex-post evaluation in terms of defining and describing the measures. First section describes the return measures and addresses the visual inspection, second section the risk measures, third section the return to risk measures, fourth section the other quantifiable measures, fifth section the benchmark related measures and sixth section the criteria difficult to measure. In the fourth chapter we apply the measures defined and described in the third chapter to the real data. In the first section we briefly introduce the Slovak funded pension pillar, in the second section we describe our data and summarize the measures and in the third section we carry out the visual as well as the quantitative (and very briefly also qualitative) evaluation. The final chapter concludes.

2. Literature Review

Literature analyzing an investment performance started emerging as soon as in the 1960s, see for example (Sharpe, 1966), (Treynor, 1965) or (Jensen, 1968). The majority of the literature since then has been focused on a single specific phenomenon, a new performance measure or an evaluation of a specific investment segment, see for example (Treynor, 1973), (Sortino, 1991) or (Keating, 2002). Only few of the papers attempted to provide a more general framework or a review applicable to the large, diverse universe of assets, for example (Fama, 1992). In our paper we focus on an easy to implement general framework for evaluation of an investment performance.

Closest to our paper in terms of content is the work of Veronique le Sourd in (Le Sourd, 2007). The paper is devoted to the performance measurement for traditional investments. It provides a comprehensive review of performance measures with many of them involving factor models and regression analysis. It also describes models that take a step away from modern portfolio theory and allow a consideration of cases beyond mean-variance theory. An author starts with different methods used in the return calculation, continues with absolute and relative risk-adjusted performance measures and also with the review of a new research on the Sharpe ratio. The second part of the paper is devoted to the measures of risk, downside risk and to the higher moments analysis. It also contains some more advanced models of a performance measurement using conditional beta and methods that are not dependent on the market model. It concludes
with the factor models review. We appreciate the comprehensiveness of the review, the mathematical exactness as well as the inclusion of many of the relevant references to the original papers. On the other hand, the paper does not provide any practical applications or suggestions for implementation and some of the measures from the paper are really hard and exhausting to implement. Moreover, many of the measures listed are very similar to each other. We also miss the discussion of the advantages and disadvantages of the measures and appropriateness of their utilization in specific cases. Based on the aforementioned facts about the paper, we decided to focus on simple to implement performance measures, diverse enough to complement each other, and provide many real data examples and suggestions for implementation.

One of the classic theoretical books from the field of the portfolio analysis which is definitely a must-read for any professional interested in the subject is the work of Elton, Gruber, Brown and Goetzmann in (Elton, 2014). The book devotes a section also to the evaluation of the investment process, mutual funds and portfolio performance, but is not primarily focused on this topic.

The very comprehensive book on the topic of quantitative finance and portfolio analysis, which also contains many of the VBA code ready for the practical implementation, is the work of Wilmott in (Wilmott, 2006). It may be of a high value for any reader who is interested in more details in specific topics, such as modeling of the volatility, mentioned also in our paper.

The more practical paper devoted to the performance evaluation has been written by Eling in (Eling, 2008). His work is focused on the comparison of the Sharpe ratio measure with other performance measures in terms of ranking the mutual and hedge funds’ performance. Eling has analyzed a data set of 38,954 funds investing in various asset classes. He has found that the ranking order of the funds utilizing the Sharpe ratio is virtually identical to the ranking order generated by other performance measures used in his paper. Eling thus argues that the Sharpe ratio is an appropriate performance measure even in cases when returns of investments are not normally distributed, which is the case for hedge funds, mutual funds and many other asset classes analyzed in his paper.
3. Ex-Post Evaluation

In the following chapter we will describe the most important ex-post performance evaluation measures. As a reader may already find many of these in an existing literature, what we consider even more important, and one of the main contributions of this paper, we present the evaluation measures along with the real world examples of their advantages and disadvantages. We discuss what they are mainly suited for and, on the other hand, what we cannot conclude purely by examining them.

Ex-post evaluation of an investment strategy can be interpreted as an evaluation in case that we already know the outcome of the strategy. In other words, we already have the real performance data of the fund employing the strategy, or at least real out of sample results of a trading account. Ex-post evaluation of a strategy is often insightful for many different types of professionals. Definitely, it is frequently being carried out by a company managing the strategy, to have a thorough understanding of its real world, or out of sample, performance. Next, it is naturally of uttermost importance for existing or prospective investors in the strategy. Last but not least, many third party evaluators, such as the regulator, researchers or journals discussing investment opportunities carry out the analysis often as well.

At first sight this evaluation process seems to be quite straightforward - we have the data, for which we need to calculate some known descriptive statistics, evaluate them and finally decide if we are satisfied with the observed outcome or not. As easy as it may seem, there are however many pitfalls that an evaluator has to face. Sometimes even investment professionals with years of experience omit important decision factors, neglect statistics needed to properly assess the particular strategy or oversimplify their analysis. We think it is crucial to consider as many performance statistics as possible and, at least, start with an ample universe of potential evaluation measures. Thereafter we may choose the most suitable subset for an investment that is to be evaluated.

Before proceeding to the definitions of the specific evaluation measures, it is useful to categorize the measures into basic groups.

Firstly, it is important to be aware that the measures described below could be calculated from two types of fund’s (or strategy’s) data. Each type has a distinct function in evaluation of the strategy. The first is a computation of the measures from the realized returns of the whole

---

1 We however do not aim to list every single existing evaluation criterion. Since the number of investment performance statistics can easily count into hundreds, we focus on the most important statistics that we think are inevitable to consider at least in some specific cases. For more comprehensive and also different list of performance measures we refer a reader for example to (Le Sourd, 2007).
strategy (daily, weekly or different). The second is a computation of the statistics from the closed trades which the strategy has made. Each of them produces different results and each of them is differently accessible. On the one hand it is almost impossible to get the data of individual trades for a person not involved in the management of the strategy. On the other hand, it is very easy to obtain the daily equity curve data of a publicly listed fund, and hence its strategy returns, for any external researcher. In this chapter we will focus on the measures computed from the first type of data, since the majority of these measures can be easily adapted also to the second type of data.

Secondly, a strategy can be evaluated as such, i.e. on a stand-alone basis, or with regards to a chosen benchmark. The first type of assessment is more suitable for absolute return strategies, which usually do not have a pre-specified market benchmark. The second type of evaluation applies better to a so called „beta” strategies or „market” strategies, i.e. investments which are in some way tied to a specific market benchmark. Since most of the investors are usually concerned about the performance of a strategy as such, and not about outperformance against some abstract benchmark, we will focus our analysis on the first type of measures, i.e. absolute return. For the sake of integrity we add a separate category discussing measures related to a benchmark.

Last but not least, evaluation measures can describe only the return part of an investment, or only the risk attributes of an investment, or both of them. Division based on these differences will be our chosen categorization. Basically, evaluation measures can be divided into 5 distinct categories - return measures, risk measures, return to risk measures, other quantifiable measures and criteria difficult to measure. As mentioned above, most of them can be related to either calendar returns or closed trades (e.g. return distribution of the daily returns or return distribution of the trades). In this paper we will not be analyzing the trade specific data, which is very hard to obtain, and rather focus on historical returns. To address the issue of a comparison of an investment against the benchmark we also add the last category which we call the benchmark related measures.
3.1. Return Measures

Calendar Days Versus Trading Days

In every calculation we have to firstly decide whether we will use a calendar day convention or a trading day convention. This can be illustrated on an example of a sample starting on 31.12.2014 and ending on 31.12.2015. In terms of the calendar day convention we have the sample of length $n = 366$ calendar days with $t_0 = 31.12.2014$, $T = 31.12.2015$ and hence, if we assign a serial number to each day, $T - t_0 = 365$. In terms of the trading day convention we have a sample of length $d = 253$ trading days with $t_0 = 1$ (first trading day), $T = 253$ (253-th trading day), $T - t_0 = 252$ and $d - 1 = 252$ return observations. We will always state what convention we are using.

Simple Return

If $P_0$ and $P_T$, where $T > t_0$ are the prices of an evaluated asset at times $t_0$ and $T$ respectively, then let us define the return on an investment from (either calendar or trading) period $t_0$ to period $T$ as follows

$$r_{t_0}^T = \frac{P_T}{P_0} - 1. \quad (1)$$

This is the simple arithmetic return for the whole period, or in financial jargon also flat return, without annualization. Throughout the paper we will utilize this simple arithmetic measure of return instead of log-returns, because it exactly represents an investor’s return on their investment, without any approximation. Log-returns of course do have important properties when used for modeling involving normal distribution, but throughout this section we will not be modeling the aforementioned.

We will omit the superscript when dealing with just 1-tick\footnote{The smallest tick or, in other words, the highest frequency used in this paper will be 1 day, that is daily frequency. However, all of the measures and calculations are applicable in the same way to any other frequency, higher or lower.} returns instead of whole period returns (such as a 1-day return instead of a 1-year return) and denote as $r_i$ only, where

---

\footnote{We omit the prefix „historical“ from any of the following measures for the sake of readability.}

\footnote{Measured by the number of trading days on the New York stock exchange. The average number of trading days per year on the American stock exchanges is 252.}
Note that, for example, $r_i = r_{t_0+i}^{t_{i+1}}$, $i = 1, \ldots, T-t_0$.

3.1.1. Visual Inspection

Many professionals base their investment decisions on a visual inspection of the charts of investments. This is usually just a part of the decision making process. However, whether one admits it or not, it is still heavily used, despite all of the perils of doing so. These include but are not limited to looking just at the historical (and not the future potential) price evolution and many biases related to imperfections of the human judgement. We do not argue against visual inspection, we, however, stress the importance of the correct visual inspection.

Logarithmic Charts

A special treatment should be applied to the very long (years or tenths of years) charts or charts in which the securities displayed experience an exponential growth. One of the ways how to deal with such charts is to partition the analysis to more shorter horizons and examine each of them separately. Another way, which enables an examination of the whole chart at a time, is to use the (natural) logarithm of the time series of the prices instead of the original time series of the prices.

---

5 It is not clear whether a visual inspection should be categorized into a return, risk, return to risk or criteria difficult to measure category. Although it most likely belongs into the last mentioned category, we include it into the return measures category. We do so because we think it is important to address the main issues about a visual inspection in the beginning rather than in the end of the chapter.
Why is this important and why may this cause problems? Consider an example in the figure (1). In the figure is displayed an American equity market price index S&P 500 (without dividends) from 1928 to 1996. Consider now an analyst trying to quickly assess the risk of this potential investment just by examining the chart. The analyst decides to do this by finding the largest peak to through losses (maximum drawdown, see section (3.2.2)) of an index in the chart. The largest loss he is able to observe is the crash of the index in 1987, when the index lost $-33.2\%$ from its peak to its through.

The problem with the non-logarithmic chart of such a long history is that an exponential growth which an index experienced makes it almost unable to correctly analyze the periods prior to the growth. And because most of the growth occurred during the 1980s and the 1990s, the price evolution during 1928-1980 is almost impossible to observe. The figure (2) displays the natural logarithm of the prices of an index during the same period.

---

* S&P 500 is the most widely used American stock market index consisting of 500 largest (measured by market capitalization) American stocks.
Figure 2


Data source: Bloomberg

The figure (2) is much more revealing when considering price evolution during earlier part of the history displayed. We can now much more easily see the huge losses which an index experienced during the Great Depression in the earlier 1930s. The maximum peak to through loss (maximum drawdown) reached $-86.1\%$ in 1932.

**Frequency Of The Charts**

For the purposes of visual inspection, we prefer as high frequency as possible. This is also the reason why we use the daily frequency of charts and returns in most of the examples throughout the paper. As already mentioned, our preferred frequency is the daily frequency, contrary to some of the literature on this topic. Many hedge funds, mutual funds or financial advisors report their performance on a weekly, monthly or some even on a quarterly basis. Although the return realized on an investment is the same regardless of the frequency of reporting chosen, the judgement of the risk becomes highly misleading with lower frequencies. Especially drawdowns (largest historical peak to through losses of an investment, see section (3.2.2)) but also many other risk characteristics may be highly underestimated when measured over low frequencies. Actually, in practice many funds rely on a monthly pricing of their net asset values,

---

7 For example (Le Sourd, 2007) advocates for using lower frequencies, such as monthly, because of the smaller noise. They argue, firstly, that lack of synchronization of different investments over a single day may cause problems. Secondly they argue that the common assumptions about asset returns (normal distribution, independent, identically distributed returns) are more plausible for returns of lower frequencies. Despite all of these, we prefer higher frequencies such as daily to be able to assess risk of an investment on a daily, not only a monthly basis.
because that way they may be better able to smooth the performance, hiding the intra-month price swings.

An example of how the risk of an investment may be underestimated when using lower frequencies can be illustrated on the American equity market index S&P 500 with net (after tax) reinvested dividends. Let us calculate the historical maximum drawdown for the index (that is the maximum historical peak to through loss, see section (3.2.2)) during the period starting on 4.1.1999 and ending on 29.2.2016. We will calculate the drawdown using daily, monthly, quarterly and yearly price data of the aforementioned index. The maximum drawdown when using the daily data is \(-55.7\%\), when using monthly data \(-51.4\%\), when using quarterly data \(-46.4\%\) and when using yearly data \(-38.4\%\). The reason why the drawdown gets larger (more negative) with higher frequency is pretty straightforward. Lower frequencies smooth the data and therefore the sample then contains fewer price deviations. Consider also this simple numerical illustration. Assume an index starts the month with the value of 1, in the middle of the month falls to the value of 0.9 and ends the month by the jump back to the value 1. Its largest monthly peak to through loss would be 0 (from 1 to 1) but its largest daily peak to through loss would be \(-10\%\). This illustrates the importance of using high enough frequencies to assess the risk of an investment.

**Scaling**

Although it should be quite obvious for everyone involved in an investment analysis, it is still important to emphasize that different investments should be compared fairly, on the same scale, with the same frequency and with the same (correct) adjustments applied. This means that, firstly, the frequency of returns (and thus pricing) should be the same for the investments under inspection.\(^8\)

Secondly, when charted in the same figure for the purposes of comparison, it is essential to scale the base (first) value of all charted investments to the same value.

If \(\{P_t\}_{t=0}^T\) is the original time series of prices of an investment and \(\{\overline{P}_t\}_{t=0}^T\) is the new scaled time series with the new base value set to \(\overline{P}_0\) (most commonly \(\overline{P}_0 = 1\)) then we obtain the new scaled time series from the original one as

\(^8\) Common mistake in this area is to compare the volatility of the monthly returns with the volatility of the daily returns. Even after annualization of the both aforementioned returns, they produce different results. The volatility of the monthly returns will be almost always quite different from the volatility of the daily returns. And this applies to almost any performance measure chosen.
\[
\bar{P}_t = \bar{P}_0 \cdot (1 + r'_0), \quad t_0 < t \leq T, \tag{2}
\]

where \( r'_0 \) is the flat return from (1).

In practice, we often witness a comparison of two different investments or a comparison of the current price evolution of the security with some piece of its history, where each price is displayed on a different axis of the same figure. This highly distorts the comparison, because we cannot infer if the (percentage) magnitude of the moves in price of investments is similar or not.

Consider an example from the blog The Mathematical Investor (The Mathematical Investor, 2014). As the blog points out, in February 2014 an article appeared on a web page which is focused on investing (Market watch, 2014). The article described a „scary” parallel between the price evolution of the Dow Jones Industrial Average (American stock market price index consisting of 30 companies) from July 2012 to February 2014 and the price evolution of the same index but from March 1928 to October 1929 (see figure (3)). The article suggested, that
if the history repeats itself, the index will experience a huge loss. Even if we accept the hypothesis that the history can repeat itself, when analyzing figure (3) we would make a false conclusion because the two time series are not scaled correctly.

In the figure (4) we do scale the two time series to a base value of 1 according to (2). We can now see that the magnitude of the price evolution of the two time series was entirely different and hardly comparable.

The Dow Jones Industrial Average price index (scaled to the base value of 1) during the two distinct periods, from 29.2.1928 to 27.11.1929 and from 2.7.2012 to 31.3.2014. Data source: Bloomberg

**Data Adjustments**

Many investments require proper data adjustments in order for an investor to be able to analyze their factual risk and return characteristics. These data adjustments are related to the character of an investment. For example if an investment pays dividends, this has to be taken into account in an analysis (either by reinvesting the dividends or by accumulating them separately). Another example can be a continually held futures contract (which usually expires after several months), which has to be rolled at each expiry date and thus all of the roll costs and roll prices have to be accounted for. Some other data adjustments include, but are not limited to, stock splits, stock mergers and acquisitions, exercising of a conversion right for a convertible bond or a specific country changing its currency. We leave a more thorough explanation of the
aforementioned adjustments for a further research. Most of them are actually related to an ex-ante performance evaluation. Hence, we rather describe in detail only the first, most commonly used (and also often neglected) adjustment when evaluating performance of investment funds, the treatment of dividends.

![Figure 5](image)

**Figure 5**
iShares iBoxx USD High Yield Corporate Bond ETF from 2.1.2009 to 27.4.2016 with three different data adjustments - price of the ETF only, price of the ETF with reinvested dividends (taxed at 30%) and price of the ETF with reinvested dividends (without tax). Data source: *Bloomberg*

Correct treatment of dividends is important for any asset which pays a dividend, whether regularly or not. If a fund (or stock or other security) pays a dividend, its price naturally falls at the moment of a payment, because the net asset value of an asset is reduced by the dividend amount. Consequently, if an investor analyzes only the price chart (without dividends) of a dividend paying asset, a return resulting from the dividend stream will be omitted. The dividend may be reinvested into the security or accumulated separately. Either method is chosen, this has to be incorporated into an analysis. Another important aspect to consider are taxes. A standard retail investor is taxed on a dividend he receives. This may or may not be included in

---

9 The day on which an asset begins to be traded without a dividend is called an *ex-date*. The date on which the dividend is actually transferred (in terms of funds) to a shareholder’s account is called a *payable date*. In this paper we will refer to a reinvestment of the dividend as to a reinvestment made on the close of an *ex-date*. This, however, may be done also differently, for example on a payable date, or on an open of an *ex-date*.
the analysis, depending on the purpose.

We illustrate the difference between an analysis of a dividend paying asset performed only by means of the price chart and the price including reinvested dividends taxed at different rates in the figure (5). The figure depicting the exchange traded fund focused on American high yield bonds, which pays a regular dividend, shows how huge a difference can be when considering different dividend treatments.

3.1.2. Cumulative Annual Growth Of Return

Let \( n \) be the number of calendar days in the period starting at (the close of the day) \( t_0 \) and ending at (the close of the day) \( T \), that is excluding \( t_0 \) and including \( T \) in the count. Let \( r^T_{t_0} \) be the simple arithmetic return from (1). We define a calendar day cumulative annual growth of return as

\[
CARC^T_{t_0} = (1 + r^T_{t_0})^{\frac{365}{n}} - 1,
\]

or when using trading days (or other periods such as weeks or months), with \( d \) being the number of trading days (or other periods) in the period starting at (the close of the day) \( t_0 \) and ending at (the close of the day) \( T \) and with \( p \) being the average number of trading days (or other periods) per year in the sample (usually \( p = 252 \) for daily frequency), then we define a trading day cumulative annual growth of return as

\[
CART^T_{t_0} = (1 + r^T_{t_0})^{\frac{p}{d}} - 1.
\]

It is worth underscoring that the calendar day and the trading day cumulative annual growth of return are not the same.\(^{10}\) Nevertheless, these are our preferred return measures, since they automatically adjust for the length of the measured horizon.

---

\(^{10}\) Consider for example how they treat weekends and holidays. Let us for example calculate an annualized return from the close of the last day of the month \( m \) to the close of the fourth day of the month \( m + 1 \), that is return over four calendar days. Assume a month starts with a working day (1), then a weekend follows (2,3) and continues with a working day (4). According to the calendar day convention \( 4/365 \) of the year just passed but according to the trading day convention only \( 2/p \) (for example \( 2/252 \)) of the year passed.
3.1.3. Rolling Return

Simple return from (1) and cumulative annual growth of return from (3) are just single point measures. That means they describe the return over the whole period and do not characterize any development within the period. In practice, however, it is very important to analyze also moving (or rolling) measures. These are simply the standard single point measures (such as $r$, CARC, or measures which will follow) measured over some predefined horizon (e.g. 1 year) and rolling in time after every tick (1 day in this paper).

With $(t_0 + n) \leq t \leq T$ we can define a simple rolling $n$-day (either trading or calendar) return as

$$Rr^T(t) = r^T_{t-n}. \quad (5)$$

![Figure 6](chart.png)

Rolling 1-year ($n = 365$) calendar return of the index S&P 500 with net (after tax) reinvested dividends. Data source: Bloomberg

Chart (6) depicts an example of a rolling 1-year ($n = 365$) calendar return. Rolling measures are especially useful when graphically charted, whereas single point measures can be charted as well as listed in a table.
3.1.4. Calendar Returns

Calendar returns are pretty straightforward to calculate and are undoubtedly the most often used measure by individual investors to judge their investments. They are anchored to a specific year or a month. In case of a month (year), we obtain them by substituting the last day of a month (year) \( k - 1 \) for \( t_0 \) in (1) and the last day of a month (year) \( k \) for \( T \), where \( k \) stands for a month (year) under the review. It is important to use exactly the closing prices of the month (year) to obtain the correct calendar month (year) return without any shifting. Calendar returns may obviously be calculated for any desired frequency such as a week, a quarter, a decade etc.

3.1.5. Distribution Of Returns

A distribution of returns of a particular investment is one of the essential statistics that need to be considered when evaluating an investment. Of uttermost importance are especially the characteristics of a left tail of the distribution, such as the worst month, worst day, \( 5^{th} \) (or \( x^{th} \)) percentile of daily (monthly) returns, etc. We recommend to analyze the distribution of more frequencies of returns at once (daily, monthly, yearly) and depict it also graphically.

![Empirical distribution (histogram) of the monthly returns of the index S&P 500 with net (after tax) reinvested dividends. The \(-10\%\) bin represents returns from the interval \((-\infty, -0.1]\) and for example the \(2\%\) bin represents returns from the interval \((0.01,0.02]\). Similarly for other bins. Data source: Bloomberg](image.png)
The figure (7) displays the empirical distribution (histogram) of the monthly returns of the net total return American equity index S&P 500.

3.2. Risk Measures

Investors tend to focus most of their time and effort on generating returns. However, it is very often the case that their wrong expectations about risk will not only erase returns that an investment has generated, but also force them to abandon their investments too soon and with a loss. The thorough understanding and analysis of the concept of risk is therefore crucial. We devote this section to the most widely used analytically quantifiable measures of risk. The risk criteria difficult to measure will be described in section (3.6).

Beginning with this section, we will also describe advantages and disadvantages of the measures discussed, providing empirical charts where appropriate.

3.2.1. Volatility

The most widely used measure of the risk is without any doubt volatility. Historical volatility (or standard deviation)\(^{11}\) can be verbally described as variation of returns of an investment around their mean return. More precisely, sample historical volatility is defined as follows:

\[
\sigma_f = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (r_i - \bar{r})^2},
\]

where \(n\) is the length of the return sample\(^{12}\), \(r_i\) denotes return \(r_{t+i}\) from (1), \(f\) denotes the frequency at which we calculate the volatility and \(\bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i\).

Following calculation may be performed at any desired frequency, that is, \(i\) may refer to day, week or month. One then obtains daily, weekly or monthly volatility. For the purposes of comparing volatility with return or other measures, it is important to standardize all of the measures used at a given (same) frequency. We prefer (and will use throughout the paper) an annualization\(^{13}\). We have already mentioned that we will use the daily frequency as a tick, that is the smallest time unit. Hence, when \(f\) denotes the daily frequency, \(i\) represents days and

\(^{11}\) Modeling of the volatility is a very comprehensive topic. We may refer a reader interested in more details for example to (Wilmott, 2006). The more complex methods of calculating volatility such as EWMA (exponentially weighted moving average model) or GARCH (generalized autoregressive conditional heteroskedastic model) are very useful for the forecasting or modeling purposes. For the purposes of evaluation of the historical performance, however, historical volatility as one of the means of risk comparison among investments should be sufficient. We will therefore utilize historical volatility throughout the paper.

\(^{12}\) Note that here we are using trading days, not calendar days, because the returns are realized only on the trading days.

\(^{13}\) In practice, we often observe incorrect comparisons of return against volatility (even on websites devoted to investment modeling), when return is annualized and volatility is not. Sometimes even two volatilities calculated at different frequencies are being compared to each other.
our annualized volatility is calculated as follows:

$$\sigma = \sqrt{p \cdot \sigma^f},$$  \hspace{1cm} (7)

where $p$ stands for the number of trading days (or other periods corresponding to the frequency $f$) per year. We emphasize once again that we are using the number of trading days and not the number of calendar days per year.

We may (and will later in the practical section) calculate also the $(n – trading day)$ rolling annualized historical volatility:

$$Rv^r(k) = \sqrt{\frac{p}{n-1} \sum_{i=k-n}^{k-1} (r_i - \bar{r}_n(k))^2},$$  \hspace{1cm} (8)

where $k$ is the trading day of calculation, $p$ number of trading days per year, $d$ is the number of trading days in the sample, $(n+1) \leq k \leq d$ and $\bar{r}_n(k) = \frac{1}{n} \sum_{i=k-n}^{k-1} (r_i)$.

Although volatility is the most widely adopted risk measure for an investment, it has its flaws and does not describe some of the major risks of an investment. Consider the following real life example why volatility may not be sufficient to assess the risk of an investment.

An investor afraid of a potential financial crisis decided to make an investment on 1.1.2008 into the American high yield bonds. He calculated daily historical volatility of the American high yield bond index from the last 5 years, which he annualized and correctly arrived at a final figure of 3.58% a.p.a. To be sure whether his decision was good, he also calculated the same volatility but from American treasury bonds (with the maturity of 7 to 10 years), which are considered as „safe heaven” investments. He again correctly arrived at value 5.49% p.a.. He compared the volatility of his high yield investment idea with that of the safe treasuries and because the volatility of high yield bonds was even lower than the volatility of treasuries, he made his investment.
However, exactly a year later this investor more than just regretted his investment. He observed the decline of the investment from its peak to its through of more than 30%. See the figure (8).

What has the investor done wrong? The list of the mistakes could be quite comprehensive, hence we name just a few. For example, he did not analyze what stands behind his investment (character of the investment), he did not look far enough into history and he did not calculate a very important measure, maximum drawdown (see section (3.2.2)).

**Semi-Deviation**

One of the disadvantages of the standard deviation is that it penalizes also the high deviation above the mean, that is, extremely positive returns. This may influence the results substantially when the distribution of returns is asymmetric and does not resemble the normal distribution, which is often the case in an investment reality. The desired property of semi-deviation is that it considers volatility only of the below-mean returns (annotation same as in (6)):

\[
\sigma_s^f = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (r_i - \bar{r})^2}.
\]  

(9)
Downside Deviation

Although semi-deviation does not penalize returns above the mean as opposed to standard historical volatility, it may still penalize positive returns (when the mean return is high), or not penalize negative returns (when the mean return is low). The property of penalizing only negative returns (or only the returns below a specific threshold) is the essence of another measure, downside deviation. We use the same annotation as in (6) and denote as \( MAR \) a pre-specified minimum acceptable return threshold:

\[
\sigma'^ D = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (r_i - MAR)^2}.
\] (10)

Practical usefulness of the downside deviation can be demonstrated with the following example. In the period starting on 11.12.2013 and ending on 11.12.2014 the Russian ruble depreciated strongly against euro, see the figure (9). An investment in euro against ruble was therefore a very profitable investment. Now, let us consider the volatility of this EUR/RUB currency pair. Annualized daily historical volatility for the period measured stands at \( 16.63\% \ p.a. \). However, annualized daily downside deviation (with \( MAR = 0 \) ) is only \( 9.99\% \ p.a. \). This difference is caused by the fact, that majority of the volatility of the currency pair was caused by positive performance. This is already adjusted for by the downside deviation, hence its value is lower.

![Figure 9](bloomberg_currency_pair_eur_rub.png)

Currency pair euro against Russian ruble (spot rate). Data source: Bloomberg
3.2.2. Maximum Drawdown

One of the most important risk measures (if not the most important one) is the maximum drawdown. Simply said, it is the worst case historical loss of an asset. The worst case loss means a loss from the top to the bottom. In other words it is the loss realized by an investor who bought an investment at the worst possible time and sold again at the worst possible time.

If \( P_t \) is a price of an asset at time \( t \) (either trading or calendar day), then we can define the maximum price of an asset up to the time \( t \) as

\[
M_t = \max_{u \in [t_0, t]} P_u,
\]

the drawdown at time \( t \) as

\[
DD_t = \frac{P_t}{M_t} - 1,
\]

(which creates an entire time series of the drawdown \( \{DD_t\}_{t=t_0}^T \) and consequently the maximum drawdown during the entire window under inspection, up to the time \( n \) as

\[
MDD_n = \min_{t \in [t_0, n]} DD_t.
\]

Why should the drawdown be so important in a risk management? The volatility gives us a clue about the distribution of returns or about the probability of the move of an investment in an either direction. However, it does not describe at all how much exactly an actual investor could have lost on his investment in the history. And precisely this is described by the maximum drawdown.

We may (and will later in practical section) calculate also the empirical distribution of the drawdowns. The distribution of the drawdowns may be calculated in two different ways. Firstly, it may be calculated from the entire time series sample \( TS \) of the drawdown as:

\[
Q^{TS}(\alpha) = \{ x : P(TDT_0^T \leq x) = \alpha \},
\]

where \( Q \) is the empirical quantile function and \( t_0, T \) denote the beginning and the end of the evaluation period respectively. Percentile calculated from the entire time series sample of the drawdown may, at least optically, understate (or even overstate) the drawdown figure. Consider for example (of an understatement) the median of the time series of the drawdown
for a fictive asset from the figure (10). This asset spends most of the time near 0% drawdown but once in a few weeks generates a loss.

![Fictive asset](image)

**Figure 10**

The time series of a price of a fictive asset and the time series of the drawdown of an asset.

The median of the time series of the drawdown for this asset is 0% (because majority of the time the drawdown of an asset is 0). The problem is that we know that the asset experiences the drawdown quite regularly and hence the median calculated in the way above is misleading.

The second way how to calculate the distribution of the drawdowns is to create a list of all of the local maximum drawdowns that an asset has experienced in the period under investigation. A new local drawdown begins whenever $P_t < M_t$ and subsequently ends when $P_t = M_t$. We may calculate the list of local maximum drawdowns as follows.

Firstly, take an entire time series of the drawdown $\{DD_t\}_{t=0}^T$ and append 0 as a last $(T+1)$ value. Secondly, divide this sample in such non-overlapping, non-empty subsamples starting on some $(t_a + 1)$ and ending on some $(t_b - 1)$ that $DD_{t_a} = DD_{t_b} = 0$ and $DD_{t_c} < 0$ for all $t_a < t_c < t_b$, while naturally $t_a \geq t_0$ and $t_b \leq (T+1)$. All of the non-zero values of $\{DD_t\}$ are now a part of some subsample, which we may denote by $S_i$. If we now for each $S_i$ take a minimum over all $\{DD_t\}$ which are part of $S_i$, we create a list of all of the local maximum drawdowns. That is,
\[
LMDD_i = \min_{DD_t \leq S_i} DD_t, \forall i.
\]  

(15)

To make this obscure definition of a practically very simple calculation more clear, we provide an example. Consider a following

\[
\{DD_t\}_i^{10} = \{0\%,-0.5\%,0\%,0\%,-1\%,-2\%,-0.5\%,-1\%,0\%,-0.7\%\}.
\]

The resulting list of local maximum drawdowns is \((-0.5\%,-2\%,-0.7\%)\). We obtained it by taking the minimum of each of the subsamples

\[
S_1 = \{-0.5\%\}, S_2 = \{-1\%,-2\%,-0.5\%,-1\%\}, S_3 = \{-0.7\%\}.
\]

Henceforth, we will be using the terms list of local maximum drawdowns and list of drawdowns interchangeably.

Let's denote the list of local maximum drawdowns \(\{LMDD_1, LMDD_2, \ldots, LMDD_n\}\) created in such a way during the period beginning with \(t_0\) and ending with \(T\) as \(DDL_{t_0}^T\). Now we may calculate the empirical distribution (and quantile) from this list as:

\[
Q^L(\alpha) = \{x : P(DDL_{t_0}^T \leq x) = \alpha\}.
\]

(16)

To give a specific example, the list of local maximum drawdowns for a fictive asset from the figure (10) is \((-2\%,-6\%,-4\%)\) and hence the median drawdown from the list is \(Q^L(0.5) = -4\%\).

We will now discuss one of the shortcomings of the maximum drawdown figure and how we may overcome it.

The figure (11) illustrates why the maximum drawdown solely is not sufficient for assessing the risk of an investment. Consider the two indices in the chart. Their maximum drawdown is almost the same (−54.5% for MSCI Europe and −55.71% for S&P 500). But the main difference is how long did the drawdown last. The time length of the drawdown of MSCI Europe was much longer than that of S&P 500, which recovered back to its price maximum approximately 9 months earlier. This characteristic of the drawdown is described by the next measure.

### 3.2.3. Time To Recovery

Time to recovery is the time (in calendar days) measuring how long did an investment take to recover from its drawdown back to the new maximum. In other words, it is the time length of the drawdown. It can be measured for every drawdown and the distribution may be calculated of them. The most widely used form is the longest (maximum) time to recovery over the whole time sample under inspection. If we use the same annotation as in (11), we define the time to recovery at time \( t \) as

\[
TTR_t = t - \arg\max_{u \in [0, t]} P_u,
\]

(17)
and the maximum time to recovery as

\[ \text{MTTR}_\alpha = \max_{z \in [0, \alpha]} TTR_z. \] (18)

### 3.2.4. Value At Risk

Value at risk is the estimated maximal loss on an asset (or portfolio) on a given time horizon with a given level of confidence. If we denote by \( \alpha \) the confidence level, by \( L \) the unknown loss of a portfolio and by \( l \) the actual specific value of the loss, which we aim to solve for, then the general definition of value at risk is as follows:

\[ \text{VaR}_\alpha = \inf \{ l \in \mathbb{R} : P[L > l] \leq 1 - \alpha \} \] (19)

Many methods of calculating value at risk do exist in practice. The simplest is the non-parametric historical simulation which is a \((1 - \alpha)\times\%\) th quantile of historical asset (or portfolio) returns on a given time frame. Parametric methods make specific assumptions about the distribution of the returns (normal, student or many other) and then calculate value at risk as expected mean return minus multiple of standard deviation (corresponding to a specific distribution). Other methods model asset returns via GARCH (or another) model and calculate value at risk among this framework. All of the aforementioned deal with the expected loss within a certain level of confidence. On the other hand, conditional value at risk analyzes the loss once it happens, that is, models the tail risk. We will not go into more detail and instead refer a reader to (Jorion, 2006).

For the purposes of this paper we will define value at risk based on a parametric method assuming normal distribution and calculate it as follows:

\[ \text{VaR}_\alpha = \bar{r} + z_\alpha \cdot \sigma, \] (20)

where \( \bar{r} \) is a mean return (calculated from the single tick frequency - daily in this paper) in a period under investigation, \( z_\alpha \) is the \( \alpha \)-quantile of the standardized normal distribution and

---

14 There are entire books focused on the topic. We do not aim to cover all existing variations and the details of the calculation, but we recognize the value and importance of this statistic, thus we do list it here in a simple form.
\( \sigma \) is the volatility (not annualized) of single tick returns from (6).

The main disadvantage of value at risk is that (with an exception of the historical simulation) it is model dependant. More advanced models need to be calibrated and often contain many parameters which can introduce over-fitting. Moreover, value at risk typically utilizes only some recent data history and does not take into account whole history of an asset. The outcome is just a single number which may and may not be helpful in a specific cases. Value at risk is therefore a perfect tool for banks to model their capital requirements, but less practical tool for investment analysts.

### 3.3. Return To Risk Measures

One of the most challenging and crucial roles of an investment analyst or a portfolio manager is an assessment of a particular investment’s return against its risk. There are many types of investors. One group focuses on finding the assets with the best return to risk profile. Another group does have specific return objective and aims to find investments capable of providing such a return, with the lowest risk possible. Also vice versa, many investors have a given risk target which should not be exceeded and try to maximize their return subject to this risk hurdle. All of the above mentioned need to very thoroughly analyze and use return to risk measures. In general, the higher return to risk, the better. However, special caution has to be applied in case of the negative return to risk values, because many of the common measures have an inverted (or highly altered) logic in such a case.\(^{15}\)

#### 3.3.1. Sharpe Ratio

Likely the most famous return to risk measure is that of William F. Sharpe introduced in 1966 in (Sharpe, 1966). The Sharpe ratio is simply the excess return of an asset (return minus risk free return) divided by asset’s volatility. An ex-ante (forward looking) Sharpe ratio is defined as follows:

\[
SHR = \frac{E[r - r_f]}{\sigma},
\]

\(^{15}\) We will not be adjusting each measure (and thus providing a special framework) for the case of negative values. Negative return to risk measures are caused almost always by the negative numerator, that is, negative excess (or classic) return of an investment. In that case special logic applies. Take an example of the Sharpe ratio from section (3.3.1). It compares excess return against volatility. When it is positive, the higher Sharpe ratio means better risk adjusted performance (higher returns with lower volatility). When it is negative, the question what is now better arises. A higher Sharpe ratio, that is less negative return with high volatility or a lower Sharpe ratio, that is, more negative return with lower volatility? The answer is not straightforward and we will not focus on answering it.
where \( r \) is a single tick (daily in this paper) return from section (3.1), \( r_f \) a single tick risk free rate and \( \sigma \) volatility from (6), calculated from single tick returns. The choice of a risk free rate is a topic on its own. In practice the return of short term (3 months or shorter) treasury bills is often used. However now, in 2016, having witnessed short term bills of numerous countries crossing zero and being in a negative territory, the 0 value is increasingly being used as \( r_f \).

Since this chapter deals with an ex-post evaluation we will define the Sharpe ratio as

\[
SHR = \frac{CART_{t_0}^T - r_f}{\sigma_{t_0}^T},
\]

where \( CART_{t_0}^T \) is not the expected return but annualized historical return from (4). \( \sigma_{t_0}^T \) is the annualized historical volatility (here we denote by subscript \( t_0 \) and superscript \( T \) the start and end of the horizon from which the volatility is calculated) from (7) and \( r_f \) is the annualized risk free rate corresponding to the period under investigation. Everything is calculated utilizing trading days.

We may calculate also the \((n - \text{trading day})\) rolling annualized Sharpe ratio:

\[
RSHR^n(t) = \frac{CART_{t-n}^t - r_f(t)}{Rv^n(t)},
\]
The main disadvantage of the Sharpe ratio lies in its reliance only on volatility as a risk measure. Consider again the example from section (3.2.1) and figure (8). Now an investor also calculates the 5 year annualized Sharpe ratio of high yield bonds and gets a value of $3.42$ which is almost triple when compared to $1.21$ of treasury bonds. This may again lead to a wrong conclusion (as already shown in the section mentioned above) that high yield bonds are undoubtedly the better investment. It is important that an investor considers also many other aspects such as drawdown, time to recovery, character of an investment, credit risk and many other, rather than solely the Sharpe ratio.

3.3.2. Sortino Ratio

The Sortino ratio (introduced by Sortino and van der Meer in (Sortino, 1991)) is very similar in nature to the Sharpe ratio. The difference between the two is the usage of a predefined hurdle rate, $minimum\ acceptable\ return\ (MAR)$, instead of a risk free rate. This hurdle rate is also incorporated into the altered standard deviation calculation, so that only returns below the hurdle rate are used in the calculation.
where \( \bar{r}_{t_0}^{T} \) is the average (single period, that is daily, not annualized) return from \( t_0 \) to \( T \), \( MAR \) is the aforementioned minimum acceptable return threshold (of the same frequency) and \( \sigma_D \) is the (single period, that is daily, not annualized) downside deviation from (10). Throughout the paper we will use the annualized Sortino ratio (to stay in line with all other measures), which we obtain as follows:

\[
SRR' = \frac{CART_{t_0}^{T} - MAR_a}{\sqrt{p \cdot \sigma_D}},
\]

where \( CART_{t_0}^{T} \) is from (4), \( p \) is the number of trading days per year, \( MAR_a \) is the annualized minimum acceptable return and \( \sigma_D \) is the (single period, that is daily, not annualized) downside deviation from (10). The main advantages and disadvantages of using the ratio correspond with that for the Sharpe ratio, with the difference of a better penalization of only below-hurdle returns for the Sortino ratio.

![Figure 13](image)

Shanghai Composite price index and Russell 2000 price index from 11.6.2012 to 3.4.2015, normalized to the value of 1 as of 11.6.2012. Data source: Bloomberg
One of the important deficiencies of the Sharpe ratio and the Sortino ratio is that they do not (fully) take into consideration if the (positive) performance of an investment was achieved consistently throughout the whole period under investigation or just during a short time period. Consider the following example of the two equity market indices - Chinese Shanghai Composite and American small capitalization Russell 2000. During the period depicted in the figure (13) both appreciated roughly the same in their value achieving almost identical cumulative annual growth of return. If we choose the same $MAR$ equal to 0 for both of them, then the only material difference between the two indices (relevant for the calculation of the Sortino ratio) is their different downside deviation. Volatility of Shanghai Composite was slightly higher than for the Russell 2000 in the period under investigation, which resulted into the Sortino ratios of 1.61 and 1.93 respectively. Now although the Sortino ratio for the Russell 2000 is 20% higher than for the Shanghai Composite, the ratio ignores the path through which these indices achieved their terminal value. By visual inspection, the path of the Russell 2000 was much more linear (and hence more stable) than the path of the Shanghai Composite. This is much better reflected in the value of another measure - the DVR ratio.

### 3.3.3 DVR Ratio

The DVR ratio (named most likely after its author David Varadi - we may refer a reader to his blog (Varadi, 2016)) connects the Sharpe ratio with $R^2$ from the linear regression of the price of an investment against time:

$$DVR = SHR \cdot R^2,$$

where $SHR$ is the Sharpe ratio from (22) and $R^2$ is the coefficient of determination from the following linear regression ($P_t$ being the price of an asset at the discrete time tick $t$ and $d$ number of trading days):

$$P_t = \alpha + \beta \cdot t, \quad t = 1, \ldots, d.$$  \hfill (27)

Now consider the recent example from the section (3.3.2). The $R^2$ for the regression of the Shanghai Composite index against time is 0.32 and for the Russell index 0.88. That results into values of $DVR$ of 0.34 and 1.16 respectively. The difference of 239% between the $DVR$ ratios is now much higher than the difference between the Sortino ratios\textsuperscript{16} and better reflects the difference in their price paths.

\textsuperscript{16} The difference in DVR ratios is also much higher than the difference between the Sharpe ratios which stands at 22% for the Sharpe ratios of 1.09 and 1.33 for the Shanghai Composite and the Russell 2000 respectively.
3.3.4 Calmar Ratio

The Calmar ratio simply compares the return of an investment against its maximum drawdown:

\[
CR = \frac{CART_t^T - T_{c0}}{MDD},
\]

where \( CART_t^T \) is from (4) and \( MDD \) (from the same corresponding period) is from (13).

![Figure 14](image)


Now consider an example in the figure (14) as an illustration why the Calmar ratio (and the maximum drawdown as well) is very sensitive to the specific period during which we measure it and hence it may not be sufficient to use it alone as a measure to compare investments. During the period under investigation the S&P 500 returned 15.91% p.a. while suffering the maximum drawdown of –27.31% and the MSCI Emerging Markets returned 17.48% p.a.
with the maximum drawdown of $-30.06\%$. Their Calmar ratios are almost identical (0.58 both), yet, when examining the chart above, an investor would be tempted to prefer S&P 500 because of its smoother price path. One of the ways how we can mathematically describe smoothness for an investment is by looking at the frequency, or percentile of drawdowns.

The main disadvantage of using the Calmar ratio is that it does not take into account a frequency and length of drawdowns (because $MDD$ represents just a single number from the entire sample). This can be hardly achieved by examining just one number (and not the whole historical evolution of the drawdowns), but can be partially accomplished if we replace the maximum drawdown in the denominator of (28) by some percentile (greater than 0) of the drawdown.

### 3.3.5 CAR To Xth Percentile Of Drawdown

Useful modification of the Calmar ratio is achieved by replacing the maximum drawdown in the denominator of (28) by a percentile (quantile) of the drawdown. Levels such as 1%, 5%, or 50% are often used in practice. This way we are able to lessen the sensitivity of the Calmar ratio (calculated from the event (or single point) sensitive maximum drawdown number) by using information from the distribution of drawdowns. We may calculate the ratio in two different ways, differing in whether we use the quantile of the entire time series of the drawdown ($TS$) in the denominator of the ratio or the quantile of the list ($L$) of the drawdowns (see also section (3.2.2)).

\[
CPD^{TS}(x) = \frac{\text{CART}^{T}_{t \rightarrow t_{0}}}{Q^{TS}(x)}, \quad (29)
\]

\[
CPD^{L}(x) = \frac{\text{CART}^{T}_{t \rightarrow t_{0}}}{Q^{L}(x)}, \quad (30)
\]

where $\text{CART}^{T}_{t \rightarrow t_{0}}$ is the trading day cumulative annual growth of return from period $t_{0}$ to period $T$ from (4) and $Q^{TS}(x)$ and $Q^{L}(x)$ are the $x$-th quantiles of the time series of the drawdown and of the list of local maximum drawdowns from (14) and (16) respectively.

If we now calculate $CPD^{TS}(5\%)$ for the recent example from the section (3.3.4), instead of very close values of the Calmar ratio (0.58 for both investments) we now get much more distinct values of 1.19 for S&P 500 and 0.8 for MSCI Emerging Markets. This clearly favors American equities because of their less frequent losses (drawdowns).
3.3.6 Omega Ratio

The main shortcoming of the measures described in the paper so far is that they do not fully take into account an entire distribution of returns (the exact shape of the distribution). For example two assets may have identical Sharpe ratios, but their return distributions may be entirely different. The Omega ratio introduced in (Keating, 2002) incorporates also the higher moments than just the first two - mean and variance of the portfolio returns distribution. The Omega ratio uses a threshold (minimum acceptable return) and calculates the ratio of the upside above the threshold against the downside below the threshold. The calculation is as follows:

\[
\Omega(MAR) = \frac{\int_{a}^{b} (1 - F(x))dx}{\int_{a}^{MAR} F(x)dx},
\]  

(31)

where \( F(x) \) is the empirical cumulative distribution function of returns of an asset, \( MAR \) is the minimum acceptable return threshold, \( a \) is the minimum return and \( b \) is the maximum return. An illustration of how the Omega ratio is calculated is depicted in the figure (15). More on the analysis of the Omega ratio can be found for example in (Frey, 2009) or (Winton Capital Management, 2003).

![Figure 15](image)

Empirical cumulative distribution function of the daily returns of S&P 500 Net Total Return index from 31.12.2014 to 31.12.2015, the minimum acceptable return threshold chosen as 0% and an illustration of the Omega ratio (ratio of the surface of the region marked by „U” against the surface of the region „D”). Data source: Bloomberg
3.3.7 Return To Maximal Loss

This measure compares the average return of an asset (on a given time horizon and on a given frequency) against the worst return. It is useful to calculate it on different frequencies, that is, as a daily return (d), weekly return (w), monthly return (m) or yearly return (y):

\[ \text{RML}^f = -\frac{\bar{r}}{\min_{i \in [t_0, T]} (r_i)} \]  

(32)

where \( f \) denotes the desired frequency (d, w, m, y) of the returns \( r \).

3.3.8 Ulcer Performance Index

The Ulcer performance index belongs to the group of measures which compare a return of an investment against its risk. It is different from the other similar measures because it uses a unique measure of risk - the sum of squares of the time series of the drawdown of an asset (the Ulcer index). This is distinct from the Calmar ratio or from the CPD because the two aforementioned use just a single point (maximum drawdown or percentile of the drawdown) in the calculation. On the contrary, the Ulcer performance index incorporates the entire drawdown evolution into the calculation. It was introduced by P. Martin in (Martin, 1989).

\[ UPI = \frac{\text{CART}^T_{t_0} - r_f}{UI} \]  

(33)

where \( \text{CART}^T_{t_0} \) is the trading day cumulative annual growth of return from period \( t_0 \) to period \( T \) from (4), \( r_f \) is the annualized risk free interest rate and \( UI \) is the above mentioned Ulcer index defined as:

\[ UI = \sqrt{\frac{\sum_{i \in I} (DD_i)^2}{d}} \]  

(34)

where \( DD_i \) is the time series of the drawdown from (12) and \( d \) is the number of trading days in the period under inspection.
3.4 Other Quantifiable Measures

Some of the measures may be hardly classified into one of the above mentioned sections. They are neither just the return or risk measures, nor measures comparing return against the risk. Such measures include higher moments of an investment returns and also some more complex ratios. All of these are however still important for the evaluation of investment strategies.

3.4.1 Statistical Significance

The majority of research papers involved in an analysis of potential investment strategies test these strategies for statistical significance. In case of research papers which deal with an evaluation of existing investments (or funds), tests for a statistical significance are used less often. Investors in such funds are much more interested in the already realized return and risk characteristics of an investment rather than in their statistical significance.

Suppose that returns of an investment \( r_i = \frac{P_i}{P_{i-1}} - 1, i = 1, \ldots, n \) (where \( n \) is the number of return observations in the sample and hence \( t_n = T \)) are independent, identically distributed and follow the normal distribution \( N(\mu, \sigma^2) \), then

\[
\frac{\bar{X} - \mu}{S} \sqrt{n} : t_{n-1},
\]

(35)

where \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} r_i \), \( S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (r_i - \bar{X})^2} \), \( \mu \) is the unknown true mean return of the whole population and \( t_{n-1} \) is the Student’s t-distribution with \( (n-1) \) degrees of freedom. One may then formulate a hypothesis of the form

\[ H_0 : \mu \leq MAR \quad \text{vs.} \quad H_1 : \mu > MAR, \]

where \( MAR \) is our minimum acceptable return (often set to 0, which is the special case). The test statistic for such hypothesis based on (35) is

\[
TS = \frac{\bar{X} - MAR}{S} \sqrt{n}.
\]

(36)
We reject the null hypothesis \( H_0 \) if \( TS > t_{n-1,\alpha} \), where \( \alpha \) is our pre-specified level of significance and \( t_{n-1,\alpha} \) is the \( \alpha \)-percent critical value ((\( 1-\alpha \)) percent quantile) of the t-distribution with \( n-1 \) degrees of freedom. The interesting property of the test statistic is that if \( MAR = r_f \), where \( r_f \) is the risk free rate used in the calculation of the daily Sharpe ratio\(^1\), then \( TS = \sqrt{n} \cdot SHR \) and we are actually testing if the scaled Sharpe ratio exceeds the critical value of the t-distribution. See more for example in (Pav, 2016).

Although many of the above mentioned assumptions do not hold in reality (returns are definitely not independent and also the rejection of \( H_0 \) does not imply that \( H_1 \) is true), many practitioners use the test to assess if the mean return of the fund (or strategy) is statistically greater than 0 (or another \( MAR \) threshold)\(^2\).

**Factor Models**

More often than just testing for a significance of the return being greater than zero, practitioners do run a series of regressions of the form

\[
    r_i = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots + \beta_k \cdot X_k + \varepsilon_i, \quad i = 1, \ldots, n, \tag{37}
\]

where \( \varepsilon_i \) is the error term and \( X_i \) are some common known factors such as the passive market return or famous Fama and French factors ((Fama, 1992)) value and size or many other. Then the main analysis involves testing of the statistical significance of the \( \alpha \) coefficient, or in other words, if the excess return of a fund (or strategy) cannot be explained by just the common already existing market factors. For the very nice review of existing factor models see for example (Le Sourd, 2007).

The problem with p-values is that their statistical significance does not imply economical significance. That is the reason why practitioners very scarcely base their investment decisions on the p-values. Hence, p-values do often serve as a necessary but not sufficient condition for accepting a potential strategy. The strategy validation through statistical significance is however much more often used for an ex-ante evaluation of strategies, rather than for an

\(^1\) In (22) we calculated the Sharpe ratio from annualized return and annualized standard deviation. The Sharpe ratio may be also calculated from daily returns (not annualized) and daily standard deviation (not annualized). We denote this as the daily Sharpe ratio.

\(^2\) In practice we often witness the race for the highest t-statistic when creating investment strategies, which is quite tricky, because statistically, higher t-statistic does not imply better strategy. The only implication which we can make is the rejection of the null hypothesis.
assessment of already existing live strategies (ex-post). We will not focus on evaluating strategies by means of a statistical significance in this paper, because of the already mentioned reasons regarding uncertain economical significance.

3.4.2 Skewness

Useful measure from the standard statistical analysis which helps to assess the tail risk (or reward) of an investment is skewness. Roughly speaking, positive skewness implies heavier right tail of the distribution (that is an investment more prone to the extreme positive returns) and negative skewness vice versa, the heavier left tail of the distribution (that is an investment more prone to the extreme negative returns). Skewness of a symmetric distribution equals zero. The calculation for the sample skewness goes as follows:

$$SK = \frac{1}{n} \sum_{i=1}^{n} (r_i - \bar{r})^3 \left( \frac{1}{n-1} \sum_{i=1}^{n} (r_i - \bar{r})^2 \right)^{-3/2},$$

where $r_i$ are the simple one tick asset returns, $n$ is the number of return observations in the sample and $\bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i$.

Skewness is just a single value which has its advantages and disadvantages as well. The main advantage is an easy comparison across different assets. The disadvantage is that it does not tell us anything about the structure of the tails of the asset returns’ distribution. One asset may have much higher skewness than the other asset and still the left tail risk (measured by other measures such as e.g. maximum drawdown) of the first asset may be higher than that of the second asset.

Consider an example of the 10 year sample of monthly returns of the two indices, American equity market index S&P 500 and American iBoxx high yield bond index from the figure (16). Skewness of the S&P 500 monthly returns is $-0.76$ and skewness of the iBoxx high yield bond monthly returns is $-1.28$. Both of them are negative, which suggests higher probability of extreme negative, rather than positive returns. Skewness of the high yield bond index is almost 70% lower than skewness of the equity market index. One may conclude that the left tail risk of the high yield bonds should be higher than that of equities. However, when we look at the maximum drawdown (see (13)) of both indices, we get the value of $-55.7\%$ for the S&P 500 and $-32.9\%$ for the iBoxx high yield. In reality, the left tail risk (when using
maximum drawdown as a measure) was actually almost 70% higher (and not lower) for the equity market index when compared to the high yield bond index.

Figure 16

Monthly returns of the S&P 500 Net Total Return index and iBoxx USD Liquid High Yield Index from 1.1.2006 to 31.12.2015. The \(-10\%\) bin represents returns from the interval \((-\infty,-0.1]\) and for example the \(2\%\) bin represents returns from the interval \((0.01,0.02]\). Similarly for other bins. Data source: Bloomberg

3.4.3 Kurtosis

Kurtosis is the second measure from the standard statistical analysis which helps to describe the tail risk of an asset. Again roughly speaking, kurtosis measures how heavy are the tails of the distribution of an asset returns (or in other words the „peakedness” of an asset returns). The higher the kurtosis the heavier the tails of the distribution. Kurtosis of a standard normal distribution equals 3. Higher (lower) kurtosis of returns than 3 thus implies that an asset is more (less) prone to extreme movements of an unknown direction (either positive or negative). The calculation of the sample kurtosis is as follows:

\[
KU = \frac{1}{n-1} \sum_{i=1}^{n} (r_{i} - \bar{r})^4 - \left(\frac{1}{n-1} \sum_{i=1}^{n} (r_{i} - \bar{r})^2\right)^2,
\]

where \(r_i\) are the simple one tick asset returns, \(n\) is the number of return observations in the sample and \(\bar{r} = \frac{1}{n} \sum_{i=1}^{n} r_i\).
Again, kurtosis does tell us how heavy the tails of the distribution of returns are (when compared to other distributions), but does not tell us other information about the structure of the tails. When we again analyze the example from the section (3.4.2) and from the figure (16) and calculate the sample kurtosis for both indices, we get the value of 1.78 for the S&P 500 equity index and 8.85 for the iBoxx high yield bond index. Because skewness is more negative for high yield bonds and kurtosis is again higher for high yield bonds we now may be tempted to conclude that high yield bonds are more prone to extreme movements than equity markets and that a probability of an extreme negative movement is higher than that of an extreme positive movement for high yield bonds. However, we already know that the maximum drawdown of the equity market index was much higher than that of the high yield bonds. A lesson to learn from this analysis is that it is crucial to take into consideration all of the information we have about the distribution and about the asset, not just the single (or a small subset) measure.

3.4.4 Fractal Efficiency

Many investors do care a lot about smoothness and stableness of their investments rather than about specific return or risk targets. The reasons behind preferring smooth path of investments are numerous but almost all of them have one in common - greater confidence in such an investment and thus lower probability of inappropriate transactions (such as buying high and selling low) performed by investors. The smoothness or stableness can be described by many of the already mentioned risk measures and risk to return measures, but there are also measures especially developed for measuring smoothness. First of them is the fractal efficiency.

The fractal efficiency is very simple to calculate and, roughly said, measures the length of the price path of an investment which is compared against a hypothetical investment with the same return but linear price path. Formula for the fractal efficiency is as follows:

\[
FE = \frac{P_T - P_{t_0}}{\sum_{i=1}^{d-1} |P_i - P_{i-1}|},
\]

where \( P_i \) is the price of an asset at time \( i \) and \( d \) is the number of trading days from \( t_0 \) to \( T \), hence \( t_{d-1} = T \). For longer time horizons, where the price path tends to create a geometric (exponential) growth path, the logarithmic fractal efficiency ratio should be used:
\[ LFE = \frac{\ln P_t - \ln P_{t_0}}{\sum_{i=1}^{t-1} |\ln P_i - \ln P_{i-1}|} \quad (41) \]

### 3.4.5 K-Ratio

The K-ratio introduced by L. Kestner in (Kestner, Futures-1996) and (Kestner, Stocks and Commodities-1996) is another from the series of measures which aim to quantify the linearity (or stableness) of an investment. The K-ratio is slightly more sophisticated than the fractal efficiency and should be able to better capture the degree of linearity of an asset. The K-ratio is essentially the scaled t-statistic of the beta coefficient from the regression of a (logarithm of the) price of an asset regressed against time. This cumbersome verbal description is much easier understood in a formula:

\[ KR = \frac{\hat{\beta}}{SE(\hat{\beta})} \cdot \frac{\sqrt{p}}{d} \quad (42) \]

where \( d \) is the number of observations in an entire sample, \( p \) is the number of observations per calendar year (that is frequency)\(^{19} \), \( SE \) is the standard error and \( \hat{\beta} \) is an estimate of the beta coefficient from the following regression:

\[ \ln P_t = \alpha + \beta \cdot t + \varepsilon_t, \quad t = 1, \ldots, d. \quad (43) \]

If it was not for the scaling factor \( \sqrt{p/d} \) then the K-ratio would be essentially just the t-statistic of the beta coefficient. In order to compare return streams of differing lengths and varying periodicity an aforementioned adjustment has to be made.

Now consider again an example from the section (3.3.2) and from the figure (13). In the chart the price paths of the American small cap equity index Russell 2000 and the Chinese equity index Shanghai Composite were depicted. In the period under inspection, both of the indices reached a very similar terminal value (and hence almost identical cumulative annual growth of return). The main difference is between the path through which they arrived to that value. We have already shown in the section (3.3.2) that volatility and downside deviation of the Shanghai Composite was higher, which resulted into the slightly lower Sortino and Sharpe ratios for Shanghai Composite. We have also shown in the section (3.3.3) that when using the DVR ratio instead of the Sharpe or the Sortino ratio solely, the difference between the Shanghai

---

\(^{19}\) If the frequency is monthly, then \( p = 12 \), if it is weekly then \( p = 52 \) and if it is daily then approximately \( p = 252 \).
Composite and the Russell 2000 widens in favor of the Russell 2000 because of its much more linear price path. We will now calculate the fractal efficiency and the K-ratio for both aforementioned indices to support this view. Surprisingly, the fractal efficiency ratio for the Shanghai Composite index is higher at 0.114 than that of the Russell 2000 index at 0.096. This is the result of the fact, that the price path of the Russell 2000 was slightly longer, although by visual inspection more linear. The K-ratio reflects this linearity much better with Russell 2000 scoring a value of 1.63 against 0.41 for the Shanghai Composite. The difference of 296% is even bigger than the difference between the DVR ratios.

3.5 Benchmark Related Measures

As opposed to the above used categorization, this section contains the mix of return, risk and return to risk measures. All with regards to a specified benchmark. Benchmark is a predefined standard, criterion or a gauge against which a real fund or an investment is measured and compared to. It is usually an index which is easily replicable or which is similar in style to the investment under consideration. There is an entire area of literature devoted to the right choice of a benchmark, see for example (Christopherson, 2012), (Philips, 2012) or (Savani, 2013). We do not aim to delve deeper into the topic of benchmarking, thus as an example we will illustrate all of the below listed measures on a pretty straightforward pairs of an investment and their benchmark.

The investments under inspection will be the two most well known and most widely used American exchange traded funds investing into the emerging market equities, the iShares MSCI Emerging Markets ETF and the Vanguard FTSE Emerging Markets ETF, both including net (after tax) reinvested dividends and including costs (that is after fees). The benchmark used is the MSCI Daily Total Return Net Emerging Markets index (index with net reinvested dividends). The risk free rate is the 3-month US LIBOR (London - Interbank Offered Rate, calculated by the Intercontinental Exchange). Source for all the data is Bloomberg. The period under evaluation is 18.3.2005 - 29.1.2016 and the frequency of the time series is weekly.

20 The aim of this paper is primarily to provide a framework for evaluating investments on a stand-alone basis and strategies not tied to a specific benchmark. Moreover, the choice of the benchmark is often subjective, which introduces more bias into an evaluation. Thus, we decided to not create an exhaustive list of all benchmark related measures and many factor models. These are used primarily for an evaluation against benchmark factors. We instead describe just the basic and most common measures, upon which the more complex ones are based.

21 The official benchmark for the iShares ETF is the aforementioned MSCI index. The official benchmark for the Vanguard ETF, however, is different (FTSE Emerging Markets All Cap China A Net Tax (US RIC) Transition Index). This index unfortunately does not have a sufficiently long history. Therefore, in our analysis we will be comparing both ETFs to the same MSCI index benchmark, which is not the official benchmark of the Vanguard ETF. Since the purpose of this section is not to select the winning ETF but rather to illustrate the measures, we do not consider this to be an obstacle.

22 In this section we decided to lower the frequency by one level to weekly. The reason of this change is closely related to the footnote (7). More specifically, when calculating linear regressions of returns against some factors or when calculating a standard deviation of the differences in returns of the two assets, noise and outliers can distort the results more heavily. It is so because we are comparing two securities day by day (instead of comparing only a single point measure such as the Sharpe ratio calculated from each time series of two assets...
These two ETFs and the index are depicted in the figure (17).

![Figure 17](image)

**Figure 17**


### 3.5.1 Jensen’s Alpha

The Jensen’s alpha (see (Jensen, 1968)) is defined as the differential between the return of an asset in excess of the risk free rate and the return explained by the market model. In other words it is an estimate of the intercept from the regression of the excess asset returns on the excess benchmark returns. More specifically, it is an estimate $\hat{\alpha}$ from the following regression:

$$ (r_i - r_f) = \alpha + \beta \cdot (r_b - r_f) + \varepsilon_i, \quad i = 1, \ldots, n, $$

where $r_i$ is the single period return of an asset under investigation, $r_f$ is the risk free return (of the same frequency), $r_b$ is the return of a benchmark (or in other words market return), $\varepsilon_i$ is the error term and $n$ is the number of return observations in the entire sample.

---

separately, as in the previous sections). The differences between daily returns may also arise because of the different pricing hours of the markets included in such a diverse index as MSCI Emerging Markets.
The Jensen’s alpha of the Vanguard ETF in the period under investigation is \(2.46 \times 10^{-3}\%\) daily (or \(0.13\%\) p.a. annualized), with the p-value of 0.97 which is statistically insignificant on all relevant levels of significance. Jensen’s alpha of the iShares ETF in the period under investigation is \(-1.9 \times 10^{-3}\%\) daily (or \(-0.1\%\) p.a. annualized), with the p-value of 0.98 which is again statistically insignificant on all relevant levels of significance. This means that both ETFs (even after fees) track the benchmark pretty closely.

**Beta**

Closely related to the Jensen’s alpha is the beta coefficient. Both of them are calculated from the same capital asset pricing model from (44). See more on the model for example in (Elton, 2014). In contrary to the Jensen’s alpha which measures the ability of a manager or an asset to deliver additional returns when compared to the benchmark, the beta coefficient measures the sensitivity of an investment to the market movements. It is calculated as an estimate \(\hat{\beta}\) of the beta coefficient from the regression (44). Roughly speaking, beta equal to 1 means that should the benchmark move \(x\%\) in some direction, the investment will follow with the same sensitivity of the \(x\%\) in the same direction. Beta greater (lower) than 1 means higher (lower) move in the same direction than the benchmark. Negative beta means a move in the opposite direction than the benchmark.

The beta of the Vanguard ETF is 1.02 (with the p-value of 0) and the beta of the iShares ETF is 1.06 (with the p-value of 0). Both of the betas are statistically significant at any level of significance. This again supports the argument that both ETFs do track the benchmark pretty closely (even after fees).

### 3.5.2 Black-Treynor Ratio

The Jensen’s alpha measures just the additional value of an investment (in terms of a return only) against the benchmark and does not describe risk of this additional value at all. To improve on this, Black and Treynor in (Treynor, 1973) developed a simple combination of the Jensen’s alpha and the beta coefficient to arrive at the return to risk Black-Treynor ratio:

\[
BTR = \frac{\hat{\alpha}}{\hat{\beta}},
\]  
(45)
where $\hat{\alpha}$ is the estimate of an alpha coefficient from the regression (44) and $\hat{\beta}$ is the estimate of a beta coefficient from the same regression. In this way an investment is penalized for being too sensitive (or even leveraged) to the movements of the benchmark and, on the other hand, rewarded for delivering alpha with a low degree of the sensitivity to the benchmark.

An annualized Black-Treynor ratio of the Vanguard ETF is $0.13\%$ and for the iShares ETF $-0.09\%$.

### 3.5.3 Tracking Error

Tracking error, similar to the correlation, measures the degree of a co-movement between two assets but in a slightly different way. It is the standard deviation of the difference between returns of an investment and returns of the benchmark:

$$TRE^f = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} ((r_i - r_i^b) - \bar{r}_d)^2},$$

where $n$ is the length of the return sample, $r_i$ denotes the single period return of an investment, $r_i^b$ return of the benchmark, $f$ denotes the frequency at which we calculate the volatility and $\bar{r}_d = \frac{1}{n} \sum_{i=1}^{n} (r_i - r_i^b)$. We may annualize it in the same way as in (7), that is

$$TRE = \sqrt{p \cdot TRE^f},$$

where $p$ stands for the number of trading weeks (or other periods corresponding to the frequency $f$) per year. Our already stated original frequency is weekly, which we annualize by $p = 52$.

The annualized tracking error of the Vanguard ETF is $9.7\% $ p.a. and the annualized tracking error of the iShares ETF is $10.6\%$ p.a.. These tracking error values are quite high if we take into account the fact that both of the ETFs are supposed to track the benchmark as closely as possible. One of the explanations is technical - because of the weekly data (which is still a quite high frequency for measuring tracking error) the same issue with the problematic pricing of different countries as in the footnote (22) may arise. We calculated also the annualized tracking error for the monthly data which resulted into the values of $4.7\%$ p.a. for the Vanguard ETF.
and 5.7% p.a. for the iShares ETF. The second explanation is more fundamental. Because the emerging market companies are still not developed enough, and the same applies for their stock markets, it is much harder to track their indices than it is for their developed counterparts.

### 3.5.4 Information Ratio

While the tracking error gives us an information about the risk that an investment will not track the benchmark, it does not contain any information about the returns of an investment compared to its benchmark. Both of these are incorporated into the information ratio:

\[
IR = \frac{CART^a - CART^b}{TRE},
\]

(48)

where \(TRE\) is the tracking error of an investment \(a\) against its benchmark \(b\) from (47) and \(CART^a\) and \(CART^b\) are the trading day cumulative annual growth of returns from (4) for an investment \(a\) and its benchmark \(b\) respectively (both of them for the entire horizon \([t_0, T]\) under investigation).

Information ratios for both of the ETFs are negative (after costs), specifically, -0.04 for the Vanguard ETF (-0.37% p.a. excess return with 9.7% p.a. tracking error) and -0.06 for the iShares ETF (-0.66% p.a. excess return with 10.6% p.a. tracking error).

### 3.5.5 Treynor Ratio

The Treynor ratio (see also (Treynor, 1965)) compares return of an investment in excess of the risk free rate against the sensitivity of an investment to its benchmark, that is against its beta:

\[
IR = \frac{CART^a - CART^f}{\hat{\beta}},
\]

(49)

where \(\hat{\beta}\) is an estimate of a beta coefficient from the regression (44) and \(CART^a\) and \(CART^f\) are the trading day cumulative annual growth of returns from (4) for an investment \(a\) and the risk free investment \(f\) respectively (both of them for the entire horizon \([t_0, T]\) under investigation).

The Treynor ratio of the Vanguard ETF is 2.8% and for the iShares ETF 2.5%.
3.6 Criteria Difficult To Measure

An investor may perform all of the analysis above regarding his/her potential investment, be satisfied with it and yet the decision to invest may still be wrong. Except from the quantitative analysis of an investment it is also essential to perform the qualitative (or fundamental) analysis. It is not the main focus of this paper to provide a comprehensive framework for the fundamental investment analysis, but we do include a short list of the most important areas that need to be considered. It is very difficult to quantify the below mentioned qualities, hence we provide just short descriptions of the topic.

3.6.1 Reputation Of The Manager

Assume an investor faces a decision to invest in the two otherwise identical investments (or products) but from the two different managers. First is a renowned bank with a hundred year long tradition and second is a new manager which has only a few years of a track record. Who would he choose? He would surely decide to invest with the first.

Length of the track record of the manager, his history of dealing with specific market crises, assets under management, sustainability and continuity of his management, size of his investment team, risk management processes and many other have to be considered thoroughly, because they can influence a potential investment.

3.6.2 Credit Risk

Closely related to the reputation of the manager is credit risk. Credit risk is the risk that a guarantee, undertaker or warrantor of an investment (such as a government is with government bonds, a company with corporate bonds, an institution with certain certificates and derivatives) will be unable to fulfill his obligations (due to for example default). Government bonds are the traditional example - should the government default on its debt, the bondholders will experience haircut or complete default.

Even many investment products other than bonds, such as exchange traded notes, certificates or numerous other derivatives have an issuer who legally serves as a warrantor for an investment. This bears an additional risk, because an underlying investment of such a certificate may be performing well, but if the institution which issued it goes bankrupt, usually so does also the value of the certificate (regardless of the value of the underlying security).23

---

23 A quite recent example for this situation was the bankruptcy of the Lehman Brothers in 2008. The Lehman Brothers defaulted not only on its own corporate bonds (and other debt) but it also defaulted as an issuer of many derivatives and certificates. All of the investments in for example certificates tied to the safe government bonds, but with some alterations (such as different currency, leverage) embedded in the
We may classify in the same category as credit risk also other risks, such as political risk or country risk. An institution which serves as a warrant for an investment may have good financial conditions on its own, but if a local government decides to nationalize the institution, pose some special levy on it or otherwise affect its operation, this may all increase the credit risk of an institution. Not to mention some special geopolitical risks such as a war. Therefore, country of domicile and country of operation of the institution have to be considered as well.

3.6.3 Liquidity Risk

Liquidity risk is the risk that an entry or exit from an investment will not be made at such a time and at such a price that an investor states in his objective, due to the investment being not liquid enough (in terms of the trading volume) to achieve it. Consider for example the small capitalization stock Synta Pharmaceuticals. In January 2016 on average around 1.5 million shares of the stock per day were traded on the American exchanges. The average price of the stock was around 0.3$. This results into the average daily volume of around 0.45 million $ per day. Now consider an institutional investor who wants to invest 1 million $ into the stock. If he gave the order to the broker to invest that amount immediately into the stock, its price would skyrocket immediately because there would not be enough sellers of the stock at its current price. The same would apply for the sale of the stock.

Hence, liquidity of an investment has to be thoroughly considered at the time of the entry and also at the potential future time of the sale. Not taking liquidity into account can introduce unexpected costs and destroy an investment objective at all.

3.6.4 Pricing Risk

By pricing risk we mean all risks associated with not executing an order (of buying or selling) an investment at the price desired. This is closely related to the liquidity risk, because a low liquid investment will usually have also a higher pricing risk and vice versa.

First of the common pricing risks is the limitation imposed by many funds and hedge funds to make (or redeem) the investments only at certain times (end of the week, end of the month, only quarterly or other periodical schemes). This introduces huge delay risk, because an investor is unable to redeem or invest at any time desired but has to wait until a provider of an investment enables him an access to do so.

Second common risk is associated with bid-ask spreads of an investment. Many low liquid

certificate and issued by the Lehman Brothers lost their entire value after the default.
securities may have wide bid-ask spreads, which means that without a comprehensive execution algorithm or a lengthy manual execution, price at which a trade would be made may be highly unfavorable (that is higher for purchases and lower for sales) for an investor. When examining historical data, the most widely used form of the data is the mid price (that is an average of the bid and the ask price), which is the price not directly available for an investor at the time of the trade. The real trade price can in reality diverge many basis points from the mid price, depending on the size of the order and an actual order book (that is the structure of all of the currently placed orders).

3.6.5 Market Risk

No matter how well an investment has been chosen and analyzed, no matter how high and established reputation has a manager and no matter how liquid and well priced an asset is, if it is an investment into the market which undergoes serious problems, its value will plunge together with the market. It is therefore necessary to become familiar with all of the risks associated with a specific market segment to which an investment belongs. For example no matter how well managed are the derivatives tied to the evolution of the volatility of American stocks (derivatives tied to the VIX index\(^{24}\)), they bear a huge intra-day as well as overnight risk of strong and sudden moves of even tenths of percentage points.

Gap Risk

Gap risk is the risk of a sudden jump in the price of an investment by such an extent, that an investor is unable to trade a security at the desired price. The gap risk is most pronounced between the two trading sessions, that is, between the close of the exchange on a day \(T\) and the opening of the exchange on a day \(T+1\). The reason why, is that in this situation it happens almost always, just to a various degree. Gap risk, however, also exists intra-day, usually because of the happening events having huge impact on the price of a security.

Gap risk is a huge risk management issue and especially for the investments, where a manager makes a promise (either hard, that is legal, or soft, that is just during the „usual market behaviour“) that an investment will not lose more than a pre-specified percentage (or absolute) amount.

\(^{24}\) http://www.cboe.com/micro/vix/
Consider the very recent and striking example of the Swiss franc, depicted in the figure (18). The Swiss national bank on the 15.1.2015 decided to abandon their floor value of the minimum 1.2 euro per franc, which it had maintained for the preceding several months. This resulted in the move of almost \(-30\%\) drop of the value of euro when measured in terms of Swiss francs. Franc started the day with the price of 1.2 euro per franc, after the announcement of the abandonment of the floor, it plunged within minutes to the low of 0.85 and closed the day at 0.975. Consider now a Swiss investor who invested in the euro and set a stop-loss to close his investments, should the euro against franc fall below 1.1 hurdle. Unless this investor had some special trading skills combined with luck, he could not be able to fulfill his stop-loss goal. If he sold his euro position at the end of the day, he would have lost 11.4\% more than his defined stop-loss level. And this poses a gap risk.

3.6.6 Details Regarding Investment Strategy

Details regarding an investment strategy are again very challenging to assess. They can evolve and change over time and there is also an endless list of them, differing from one investment manager to another. Nevertheless, it is crucial to have as deep knowledge as possible about a structure of an investment and about an investment process carried out by a manager of an investment. Below is just a short list of some common risks associated with the style of the management of the portfolio.
Concentration Risk

Concentration risk is the risk of not diversifying a fund or a portfolio enough. It is often the case that managers achieve spectacular short term performance by picking (with a questionable degree of luck and skill) just a few securities and, hence, run just a concentrated portfolio. The risk associated with such a portfolio is that the same securities which may have performed well may perform also terribly and highly under-perform some well known broad based benchmark such as the S&P 500 index.

Timing Risk

Timing risk is the risk that a manager will make huge and sudden shifts in the composition of a portfolio at wrong times. The more the manager is active versus the benchmark, the higher the timing risk. Passive investments such as ETFs have almost no timing risk when compared to their benchmark. However, for all of the investments (passive or active) there is a timing risk of an investor (not the manager), the risk of entering into or exiting from an investment by investors at an inconvenient time.

Leverage Risk

If a manager tries to harvest some consistent market premium which is, however, small in magnitude, he very often introduces a leverage into his investment process. The leverage enables him to borrow capital and invest more than just 100% into the strategy. That way the manager is able to multiply returns achieved (minus costs of borrowing). The risk of such a multiplication is that it leverages not only gains but, naturally, also losses. If an anomaly or a premium starts to disappear and behave differently than the manager has been used to, the risk of huge losses magnified by the leverage arises.

Currency Risk

It is essential to understand what a currency exposure of an investment is. The simplest case without the currency risk is an investment consisting of assets denominated only in the local currency (for example euro assets for a euro investor). This is, however, very often not the case. In practice, for example many euro denominated mutual and pension funds investing into the US dollar denominated assets and not hedging their dollar exposure strongly benefited from the appreciation of the US dollar in the 2015. On the contrary, the exact opposite was experienced by the US denominated funds investing in the euro denominated assets and not hedging their currency exposure. They experienced substantial losses only because of the currency movement and not the underlying security movement.
Costs

Every investment product has some costs associated with an investment. The most common types of costs start with entry and exit fees which are one-off, that is subtracted from the value of a client’s investment in the beginning and in the end. Another basic type of a fee is the management fee, which is usually charged continuously, that is subtracted from the value of an investment on a daily basis. Next typical type of the fee is the performance fee which is usually subtracted from the value of an investment on each day an investment exceeds a predefined hurdle, such as a new maximum over the past 3 years.

It is crucial to know in advance all of the fees, their character and how and when they will be deducted, because they may have a material impact on an investment (see for example (Faber, 2015)). It is also very important to distinguish between the gross (without fees) and net (after fees) performance of an investment. Many performance reports do contain just the gross performance which is substantially lower after the application of fees.

Fees may also come in many other, more hidden forms. For example many certificates tied to the evolution of equity market indices tie their value to the price indices. Since price indices do not contain dividends, but the real investment into the securities which are in the index does yield dividends, this way an issuer of the certificate is essentially keeping the dividends for himself.

4. The Practical Ex-Post Evaluation

We will demonstrate, analyze and compare all of the measures from the chapter (3) for the six equity funds managed by six different pension companies of the Slovak funded pillar (defined contribution) pension system. We have chosen the Slovak funded pillar (DC) pension system funds for several reasons. Firstly, an author has already been involved in the DC pension research, see for example (Melichercík, Journal of Economics - 2015), (Melichercík, Acta Mathematica - 2015) or (Vilcek, 2013). Secondly, the funds in the funded pillar employ a quite similar strategy in managing their assets, therefore they constitute a good peer group. They also now have a track record of more than ten years of managing pension assets. Furthermore, the number of six (funds) is very well suited for the purposes of this analysis. It is enough to uncover differences and hence demonstrate measures from this paper and on the other hand it is not too extensive to exhaust a reader. Last but not least, an author’s current occupation is not

---

25 By „quite similar“ we mean a long term strategy with a mandate of creating a pension income for a pensioner in the long run. Of course we are aware of the fact that the strategies may differ substantially in carrying out this mandate.
related to the Slovak funded pillar at all and therefore the analysis should not be biased because of any potential conflict of interests.

The reason why we have chosen the equity fund category is pretty straightforward. The guaranteed bond fund, and the non-guaranteed equity fund are now the only two funds which pension companies are obliged to manage. Hence, our choice has been narrowed to the two options, of which we have chosen for the purposes of our analysis the equity funds because of their higher diversity and less strict legal limits.

4.1 The Funded Pillar Of The Slovak Pension System

4.1.1 Brief History

Until 2005, Slovakia employed only a compulsory Pay As You Go (PAYG) pension system. In simplicity, the system was, and still is, based on the pension payments to the current pensioner’s being paid out of the wages of the currently working population. That, naturally, has raised questions about sustainability of such a system, should the population age rapidly (which happens to be the case). Hence, the pension reform from 2005 introduced two additional pillars - the mandatory (for new savers) defined contribution scheme (funded pillar) and the voluntary defined contribution scheme (third pillar). The PAYG system (first pillar) remained managed by the government, whereas the other two pillars have begun to be managed by the private pension companies via pension funds.

During 2005-2008 each pension company in the funded pillar managed three funds - conservative, balanced and growth fund with relatively free investment mandate and not so strict legal limits. After the crisis of 2008, the Slovak government ratified an amendment to the pension law, which made it compulsory for the pension companies to evaluate performance of all of their funds on a rolling 6-month basis. In case of any loss on any 6-month horizon, the pension company was obliged to compensate such a loss to the pensioners out of their own capital. This resulted into highly conservative investment strategies of all of the funds, where the majority of portfolios consisted of term deposits and short term bills.

Another amendment of the pension law from 2011 canceled the guaranty of not making a loss on any 6-month horizon in all of the funds except the conservative (bond) fund, where the guaranty was changed from the 6-month horizon to the 5-year horizon. The pension companies were also obliged to start managing four funds with the new names - the bond fund, the mixed fund, the equity fund and the index fund. This amendment had not resulted into any material changes in investment strategies of the pension companies, because another amendment of the
law was supposed to be ratified a year later.

4.1.2 Current Setup Of The Pillar

Based on the amendment of the pension law from August 2012, each pension company is now obliged to manage at least two funds - the guaranteed bond fund and the non-guaranteed equity fund, while any additional funds are voluntary. In the guaranteed bond fund a pension company is now obliged by law to compensate any loss made on any rolling 10-calendar year period to pensioners from its own capital. In other funds, no guaranty has been imposed by the law. This finally resulted into huge changes of the investment strategies employed in all of the funds. Majority of the funds have increased their risk with an aim of increasing a potential long term return.

4.2 Description Of The Measures And Data

4.2.1 Fund Data

We obtained the daily price data for the six equity funds managed by six pension companies (in the alphabetical order of the pension company):

- Vital, non-guaranteed equity fund of Aegon,
- Progres, non-guaranteed equity fund of Allianz,
- Akciovy, non-guaranteed equity fund of Axa,
- Profit, non-guaranteed equity fund of ING,
- Prosperita, non-guaranteed equity fund of DSS Postovej banky (our abbreviation - Postova) and
- Profit, non-guaranteed equity fund of VUB Generali (our abbreviation - VUB).

The data was obtained from the web pages of the pension companies (in case they provided it)\textsuperscript{26} or from the web page www.manazeruspor.sk in case of Aegon and Allianz. The data was normalized to the base value of 1 according to (2) as of 2.1.2006 and ends on 21.3.2016. The dates of the data were aligned to one of the shortest time series, the ING’s fund, to avoid repeating values.\textsuperscript{27} The price data of the funds prior to 1.1.2009 (abandonment of the Slovak


\textsuperscript{27} There happen to be some dates in the sample, on which certain pension companies do provide the price data and some do not. In case there is some missing price data on a certain date (aligned to the aforementioned ING’s time series) for some fund, the last available price data point
koruna and introduction of the euro) have been already adjusted in our data source according to the official conversion rate. The data for the risk free rate represented by the 3-month Euribor were obtained from Bloomberg.

4.2.2 Performance Measures

We calculated the performance measures for each fund for the four different examination periods:

1. the whole sample 2.1.2006-21.3.2016,
2. the first part of the sample, 2.1.2006-30.6.2009, during which the pension companies were not obliged to guarantee a profit; this includes the financial crisis,
3. the second part of the sample, 30.6.2009-28.3.2013, during which the pension companies were obliged to guarantee a profit on a rolling 6-calendar month basis and hence invested very conservatively
4. and finally the third part of the sample, 28.3.2013-21.3.2016 during which the pension companies started to again invest in line with their long term capital appreciation mandate, because of the missing capital protection requirements.

For the sake of a better readability in the tables below we use the names of the measures instead of their official abbreviations defined in chapter (3). This is the description of all of the analyzed measures:

- **Calendar year returns** - these are the simple yearly returns from the section (3.1.4) for a year \(i\) starting on the close of the last day of the year \((i-1)\) and ending on the close of the last day of the year \(i\); an exception is the year 2006, where the calculation starts on the close of 2.1.2006 and the year 2016 where the calculation ends on the close of 21.3.2016
- **Flat return** - simple return for the whole period under examination from (1)
- **Annualized return** - cumulative annual growth of return based on trading days from (4)
- **Best/worst day/month/year** - based on the flat return for the calendar days/months/years; in case the month (year) in the sample under examination has not ended yet (for example in 2016 we have only the data up to and including 21.3.2016), the month to date (year to date) figure is used (flat, not annualized)

is used instead. That is, there may be some repeating values for some funds. This is also the reason why we align to the shortest instead of to the longest time series.
• **Xth perc of 252-day return** - the percentile of the rolling 252-trading day return from (5)

• **Historical volatility** - the annualized historical volatility from (7)

• **Xth perc of 252-day volatility** - the percentile of the rolling 252-trading day annualized historical volatility from (8)

• **Semi deviation** - the annualized historical semi deviation from (9)

• **Downside deviation** - the annualized historical downside deviation from (10) with \( MAR = 0 \)

• **Max drawdown** - the maximum drawdown from (13)

• **Xth perc of TS drawdown** - the percentile of the time series of the drawdown from (14)

• **Xth perc of list of drawdowns** - the percentile of the list of local maximum drawdowns from (16); also the drawdown on the last day in the sample, that is not yet ended drawdown, is included into the calculation

• **Time to recovery** - maximum time to recovery in days from (18)

• **Value at risk** - daily 95% Value at Risk from (20)

• **Sharpe ratio** - the annualized Sharpe ratio from (22) with the risk free rate equal to the average of the 3-month Euribor over the specific period under inspection

• **Xth perc of 252-day Sharpe** - the percentile of the rolling 252-trading day annualized Sharpe ratio from (23)

• **Sortino ratio** - the annualized Sortino ratio from (25) with \( MAR = 0 \)

• **DVR ratio** - the DVR ratio from (26)

• **Calmar ratio** - the Calmar ratio from (28)

• **CAR to 5th perc of TS drawdown** - the cumulative annualized trading day return to 5th percentile of the time series of the drawdown from (29)

• **CAR to 5th perc of list of dd** - the cumulative annualized trading day return to 5th percentile of the list of local maximum drawdowns from (30)

• **Omega (daily/monthly returns)** - the Omega ratio from (31) calculated from the daily/monthly returns with \( MAR = 0 \)

• **Return to max loss (daily/monthly)** - return to maximal loss calculated from the daily/monthly returns from (32)

• **Ulcer performance index** - Ulcer performance index from (33) with the risk free rate equal to the average of the 3-month Euribor over the period under inspection

• **t-statistic** - the one-sided Student’s t-test test statistic for the hypothesis
\[ H_0 : \mu \leq MAR \text{ vs. } H_1 : \mu > MAR, \] where \( \mu \) is the daily return over the period under inspection and \( MAR = 0 \) from (36)

- \textit{p-value of t-statistic} - p-value for the one-sided t-test of the above mentioned hypothesis
- \textit{Skewness (daily/monthly)} - the skewness of the daily/monthly returns from (38)
- \textit{Kurtosis (daily/monthly)} - the kurtosis of the daily/monthly returns from (39)
- \textit{Fractal efficiency} - the fractal efficiency from (40)
- \textit{K-ratio} - the K-ratio from (42)

There is not any specific benchmark for the equity pension funds in the funded pillar. It is also impossible to benchmark the whole 10-year period with a uniform benchmark, because of the significant changes in the investment policies induced by the changes in the pension law. Therefore, we decided to create our own artificial benchmark. The benchmark is the daily average of returns of all of the funds under inspection. We are aware of the fact that when we create a benchmark this way, it is impossible for all of the funds to beat the benchmark. We are also aware that this may distort the regression analysis, because we will be regressing a fund’s returns on the benchmark which already contains, among other, also the fund’s returns. In spite of this we will use this benchmark because we believe it remains one of the most fair and simple ways of constructing the benchmark for the purposes of our analysis.

The benchmark related measures are calculated from the weekly, instead of the daily returns to reduce the noise. See also section (3.5) and footnote (22) for more thorough explanation.

- \textit{Jensen’s alpha} - the annualized Jensen’s alpha from (44) with the risk free rate equal to the 3-month Euribor during the period under inspection and the benchmark described in the paragraph above
- \textit{Beta} - the beta against the benchmark from section (3.5.1)
- \textit{Black-Treynor ratio} - the annualized Black-Treynor ratio from (45)
- \textit{Tracking error} - the annualized tracking error from (47)
- \textit{Information ratio} - the information ratio from (48)
- \textit{Treynor ratio} - the Treynor ratio from (49)
4.3 Evaluation

We will firstly perform a visual analysis of the Slovak pension funds’ performance. Next, we will continue by examining the return measures, risk measures, return to risk measures and benchmark related measures for the whole sample. An analysis of the first part of the sample containing the financial crisis in 2008 will follow. The next two sections will deal with the analysis of the second and third, most recent, part of the sample. The last section will briefly comment on criteria difficult to measure.

4.3.1 Visual Evaluation

Firstly, we perform a visual analysis of the Slovak pension funds’ performance which is carried out in the charts (19)-(28). The equity curves and the time series of the drawdown are charted for all funds.

Rolling return, volatility, Sharpe ratio, Value at Risk and the return distribution is charted only for the VUB equity fund, with the aim of not exhausting a reader. Nonetheless, the character of the figures (not reported here) containing the aforementioned measures depicted only for the VUB fund is quite similar for all of the funds in our analysis.

Table 1

Correlation of the daily returns among the Slovak funded pillar equity pension funds for the different calculation periods. Row and column names denote the pension companies which manage the funds (Ae=Aegon, Al=Allianz, Ax=Axa, In=Ing, Pb=Postova, Vu=VUB). Data source and normalization: see section (4.2.1).

<table>
<thead>
<tr>
<th></th>
<th>Ae</th>
<th>Al</th>
<th>Ax</th>
<th>In</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>[2.1.2006 - 21.3.2016]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ae</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>0.90</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ax</td>
<td>0.82</td>
<td>0.87</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In</td>
<td>0.83</td>
<td>0.80</td>
<td>0.85</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.87</td>
<td>0.84</td>
<td>0.91</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td>[2.1.2006 - 30.6.2009]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ae</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al</td>
<td>0.83</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ax</td>
<td>0.81</td>
<td>0.88</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In</td>
<td>0.71</td>
<td>0.75</td>
<td>0.76</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.82</td>
<td>0.93</td>
<td>0.94</td>
<td>0.88</td>
<td>0.76</td>
</tr>
</tbody>
</table>
Table 2

Correlation of the weekly returns among the Slovak funded pillar equity pension funds for the different calculation periods. Row and column names denote the pension companies which manage the funds (Ae=Aegon, Al=Allianz, Ax=Axa, In=Ing, Pb=Postova, Vu=VUB). Data source and normalization: see section (4.2.1).


<table>
<thead>
<tr>
<th></th>
<th>Ae</th>
<th>Al</th>
<th>Ax</th>
<th>In</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ax</td>
<td>0.24</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In</td>
<td>0.28</td>
<td>0.19</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.18</td>
<td>0.34</td>
<td>0.37</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Vu</td>
<td>0.29</td>
<td>0.17</td>
<td>0.26</td>
<td>0.20</td>
<td>0.63</td>
</tr>
</tbody>
</table>

[28.3.2013 - 21.3.2016]

<table>
<thead>
<tr>
<th></th>
<th>Ae</th>
<th>Al</th>
<th>Ax</th>
<th>In</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ax</td>
<td>0.93</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In</td>
<td>0.83</td>
<td>0.89</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.89</td>
<td>0.81</td>
<td>0.90</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Vu</td>
<td>0.92</td>
<td>0.83</td>
<td>0.94</td>
<td>0.80</td>
<td>0.92</td>
</tr>
</tbody>
</table>

[2.1.2006 - 21.3.2016]

<table>
<thead>
<tr>
<th></th>
<th>Ae</th>
<th>Al</th>
<th>Ax</th>
<th>In</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ax</td>
<td>0.92</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In</td>
<td>0.89</td>
<td>0.88</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.87</td>
<td>0.92</td>
<td>0.89</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td>Vu</td>
<td>0.89</td>
<td>0.94</td>
<td>0.95</td>
<td>0.90</td>
<td>0.92</td>
</tr>
</tbody>
</table>

[2.1.2006 - 30.6.2009]

<table>
<thead>
<tr>
<th></th>
<th>Ae</th>
<th>Al</th>
<th>Ax</th>
<th>In</th>
<th>Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ax</td>
<td>0.89</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In</td>
<td>0.89</td>
<td>0.94</td>
<td>0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pb</td>
<td>0.90</td>
<td>0.94</td>
<td>0.93</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Vu</td>
<td>0.90</td>
<td>0.97</td>
<td>0.97</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Figure (19) depicts (in two parts, for the sake of clarity) the daily equity curves of the six equity funds in the whole, more than 10 year long, sample. We can see that the evolution of the wealth in each fund was quite similar. Correlation among the funds’ returns is high. For example the average correlation of the daily returns among funds for the whole sample is 0.84, see table (1). We calculated also the average correlation of the weekly returns to mitigate any potential impact in the delay of the pricing of the funds. The weekly correlation was even higher, 0.9 for the whole sample, see table (2).

We may also observe the three entirely different regimes in the investment policies employed by the funds. The first part of the sample, during which the pension companies were not obliged to compensate loss to pensioners, but were still subject to strict legal limits. Next, the second part, when the pension companies invested very conservatively to avoid any loss on a 6-month rolling horizon. Finally, the third part, when the pension companies were allowed by policymakers to invest the most aggressively so far.

Upper part of the figure (20) depicts the daily equity curves of the six equity funds in the first part of the sample spanning approximately three and a half years. This includes the final months of the equity market boom, that is 2006 and first half of 2007, as well as the financial crisis of

---

28 Recall that illiquidity of some securities in the portfolios may cause differences in funds’ returns on a daily basis. Consider an example of a local covered bond which may be repriced only when the trade on the exchange is made. The main, liquid bond market may experience significant moves which are priced on a daily basis but the covered bond will be repriced only after the next trade. Hence, if one fund possess the bond which was not traded after the market movement and the other fund possess another bond which, however, was traded after the movement, the pricing of the funds may be different, although the economical exposure remains the same for both funds. This potential difference should, however, be smoothed when using lower frequency returns.
2008. We can see that all of the funds experienced material losses and substantially elevated volatility during the crisis.

Lower part of the figure (20) displays the equity curves of the funds in the second part of the sample spanning approximately three years and nine months. This period was strongly influenced by the amendment of the pension law which made it compulsory for the pension companies to guarantee a rolling 6-month nonnegative performance for pensioners. In case of not fulfilling this guaranty, the pension company was obliged to compensate a pensioner any loss from the company’s own capital. It resulted into the portfolios consisting only of very low duration bonds, money market instruments and term deposits. This translated into very smooth equity curves.

Figure (21) depicts (in two parts, for the sake of clarity) the equity curves of the funds in the third part of the sample spanning approximately three most recent years. We can observe that the pension companies started to again invest more aggressively with a substantial exposure to equity markets. This is the result of the amendment of the pension law which canceled the guaranty obligation for the equity funds.

Figure (22) displays the daily time series of the drawdown for each fund. The crisis of 2008 caused the funds to experience drawdowns from \(-9\%\) to \(-15\%\). It then took the funds more than 5 years (with an exception of VUB, in which case a little less than 5) to recover this drawdown and create the new maximum value of the fund. Another volatile period accompanied with high drawdowns happened very recently in 2016 and culminated in February. Several funds experienced their new maximum drawdown and neither fund has recovered this drawdown yet, as of 21.3.2016.

Figure (23) depicts the rolling 21-trading day, 63-trading day and 252-trading day simple return of the VUB equity fund. We may notice swings ranging from \(-10\%\) to more than \(20\%\) when 252-day return is considered, especially in the first and the third part of the sample. The second part of the sample is characterized by very stable returns.

Chart (24) helps to analyze the distribution of the daily returns of the VUB equity fund for the whole sample. It contains the histogram of the aforementioned returns. The most frequent bin with more than \(40\%\) of the observations belongs to the interval \((0\%,0.1\%]\). We may also observe high kurtosis and a slight, hardly recognizable skewness.

The same statements may be applied to the monthly distribution of the returns of the VUB equity fund, which are depicted in the figure (25).
Figure (26) displays the rolling 21-trading day, 63-trading day and 252-trading day annualized historical volatility of the VUB equity fund. We may notice three distinct periods corresponding with the pension law amendments as mentioned several times earlier. Volatility in the first period peaked during the financial crisis and the collapse of the Lehman Brothers. However, it was still low when compared to the long-term pension funds managed in the developed countries, where their volatility approaches equity volatility (usually more than 10% p.a.). In the second period, when the pension companies were obliged to compensate any 6-month losses, the volatility was abnormally low (ranging between 0.15% p.a. and 0.4% p.a.). This is the result of the low risk investments made by fund managers at that time.

In the third period, when the obligations to compensate losses ceased to exist, the volatility increased to its historical maximum. This is in line with a higher risk (equity) exposure of the fund.

Rolling daily Value at Risk with 252-day calculation window and $\alpha = 95\%$ confidence level is charted in the figure (27). It has a similar character as the figure (26), that is, moderate risk in the first period, very low risk in the second period and high risk in the third period.

Figure (28) depicts the rolling 21-trading day, 63-trading day and 252-trading day annualized Sharpe ratio of the VUB equity fund with the risk free rate of 3-month Euribor. We again observe the three distinct periods corresponding to the already mentioned amendments of the pension law. The first period was also substantially affected by the financial crisis. We may observe that the rolling 252-day Sharpe ratio plunged to its historical low in 2008. Contrary to it, in the next period (2009-2013) the 252-day Sharpe ratio firstly increased to approximately 5 and then climbed even higher to create the historical maximum of more than 10.

Such value of the Sharpe ratio is typically unrealistic and almost unattainable for common pension funds. These funds have high exposure to risky assets, in order to increase their long-term return potential. However, in case of the pension funds in Slovakia during the aforementioned period, the rationale behind such Sharpe ratio is pretty straightforward. Majority of the assets bought by the portfolio managers had a volatility close to 0, because of their short maturity and low risk. Consider also term deposits, which, by definition, have 0 volatility, but of course only in case that the institution where the deposit is made does not default. If it does not, and their return exceeds the return of the risk free rate, they have a property of an infinite Sharpe ratio. All of these instruments helped volatility of the fund approach 0, whereas their return spread over the risk free rate of the 3-month Euribor was still
high over that period. This resulted into very high Sharpe ratios.

The most recent period shows the most realistic picture of the behavior of the equity pension fund portfolios. That is, 1-year Sharpe ratios in between the range of approximately -3 and 3.

**Figure 19**

Daily equity curves of the six Slovak funded pillar equity pension funds. Examination period is the whole sample starting on 2.1.2006 and ending on 21.3.2016. Names denote the pension companies which manage the funds. Data source and normalization: see section (4.2.1).
Daily equity curves of the six Slovak funded pillar equity pension funds. Examination period in the upper part of the chart is the first part of the sample starting on 2.1.2006 and ending on 30.6.2009. Examination period in the lower part of the chart is the second part of the sample starting on 30.6.2009 and ending on 28.3.2013. Data source and normalization: see section (4.2.1).
Daily equity curves of the six Slovak funded pillar equity pension funds. Examination period is the third part of the sample starting on 28.3.2013 and ending on 21.3.2016. Data source and normalization: see section (4.2.1).
Figure 22

Time series of the drawdown for each of the six funds. Calculations performed according to the section (3.2). Data source and normalization: see section (4.2.1).
Rolling 21-trading day, 63-trading day and 252-trading day simple return of the VUB equity fund. Calculations performed according to the section (3.1). Data source and normalization: see section (4.2.1).

Empirical distribution (histogram) of the daily returns of the VUB equity fund. The $-1.5\%$ bin represents returns from the interval $(-\infty, -0.015]$ and for example the $1\%$ bin represents returns from the interval $(0.009, 0.01]$. Similarly for other bins. Data source and normalization: see section (4.2.1).
Empirical distribution (histogram) of the monthly returns of the VUB equity fund. The $-3\%$ bin represents returns from the interval $(-\infty,-0.03]$ and for example the $2\%$ bin represents returns from the interval $(0.015,0.02]$. Similarly for other bins. Data source and normalization: see section (4.2.1).

Rolling 21-trading day, 63-trading day and 252-trading day annualized historical volatility of the VUB equity fund. Calculations performed according to the section (3.1). Data source and normalization: see section (4.2.1).
Figure 27

Daily returns and rolling daily Value at Risk with $\alpha = 95\%$ and calculation window of 252 trading days for the VUB equity fund. Calculations performed according to the section (3.4). Data source and normalization: see section (4.2.1).

Figure 28

Rolling 21-trading day, 63-trading day and 252-trading day annualized Sharpe ratio with the risk free rate equal to the 3-month Euribor for the VUB equity fund. Calculations performed according to the section (4.1). Data source and normalization: see section (4.2.1).
4.3.2 The Whole Sample

Table 3

Calendar year returns of the pension funds calculated in the whole period starting on 2.1.2006 and ending on 21.3.2016. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th>Year</th>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Penzijna</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>4.59%</td>
<td>5.10%</td>
<td>4.86%</td>
<td>2.96%</td>
<td>5.04%</td>
<td>4.35%</td>
</tr>
<tr>
<td>2007</td>
<td>4.26%</td>
<td>3.55%</td>
<td>3.73%</td>
<td>3.26%</td>
<td>2.09%</td>
<td>4.16%</td>
</tr>
<tr>
<td>2008</td>
<td>-11.03%</td>
<td>-6.76%</td>
<td>-6.80%</td>
<td>-6.69%</td>
<td>-8.46%</td>
<td>-7.48%</td>
</tr>
<tr>
<td>2009</td>
<td>0.90%</td>
<td>0.94%</td>
<td>-0.13%</td>
<td>1.24%</td>
<td>1.42%</td>
<td>1.29%</td>
</tr>
<tr>
<td>2010</td>
<td>1.19%</td>
<td>0.78%</td>
<td>1.21%</td>
<td>1.53%</td>
<td>1.23%</td>
<td>1.86%</td>
</tr>
<tr>
<td>2011</td>
<td>1.13%</td>
<td>1.30%</td>
<td>1.48%</td>
<td>1.26%</td>
<td>1.27%</td>
<td>2.12%</td>
</tr>
<tr>
<td>2012</td>
<td>3.51%</td>
<td>2.94%</td>
<td>2.20%</td>
<td>3.44%</td>
<td>4.46%</td>
<td>3.76%</td>
</tr>
<tr>
<td>2013</td>
<td>2.11%</td>
<td>4.33%</td>
<td>3.83%</td>
<td>3.15%</td>
<td>3.69%</td>
<td>5.52%</td>
</tr>
<tr>
<td>2014</td>
<td>7.57%</td>
<td>10.85%</td>
<td>8.60%</td>
<td>9.69%</td>
<td>8.56%</td>
<td>10.32%</td>
</tr>
<tr>
<td>2015</td>
<td>4.00%</td>
<td>2.52%</td>
<td>3.24%</td>
<td>2.63%</td>
<td>0.55%</td>
<td>3.92%</td>
</tr>
<tr>
<td>2016</td>
<td>-2.15%</td>
<td>-2.08%</td>
<td>-0.56%</td>
<td>-0.40%</td>
<td>-8.55%</td>
<td>-2.28%</td>
</tr>
</tbody>
</table>

We begin an analysis of the whole sample by listing the yearly returns for all of the funds (with 2016 being the year to date figure) in the table (3). We may observe that the most successful year so far was achieved by all funds in 2014, while the worst year on average was recorded in 2008. The two years with the lowest dispersion in returns among funds were the years 2010 and 2011, when the pension companies were obliged to compensate any 6-calendar month losses out of their own capital. On the contrary, the most dispersed calendar year performance has been recorded in the (not yet ended) 2016, followed by 2008.

Table 4

Return and risk measures of the pension funds calculated for the whole period starting on 2.1.2006 and ending on 21.3.2016. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Postova</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat return</td>
<td>15.92%</td>
<td>24.97%</td>
<td>23.03%</td>
<td>23.51%</td>
<td>10.35%</td>
<td>29.96%</td>
</tr>
<tr>
<td>Annualized return</td>
<td>1.46%</td>
<td>2.21%</td>
<td>2.05%</td>
<td>2.09%</td>
<td>0.97%</td>
<td>2.60%</td>
</tr>
<tr>
<td>Best day</td>
<td>2.30%</td>
<td>3.12%</td>
<td>1.47%</td>
<td>1.66%</td>
<td>2.51%</td>
<td>2.30%</td>
</tr>
<tr>
<td>Worst day</td>
<td>-2.95%</td>
<td>-4.94%</td>
<td>-2.05%</td>
<td>-2.98%</td>
<td>-3.52%</td>
<td>-3.45%</td>
</tr>
<tr>
<td>Best month</td>
<td>4.81%</td>
<td>8.29%</td>
<td>3.43%</td>
<td>3.82%</td>
<td>6.12%</td>
<td>5.03%</td>
</tr>
<tr>
<td>Worst month</td>
<td>-5.13%</td>
<td>-7.88%</td>
<td>-2.99%</td>
<td>-4.44%</td>
<td>-6.06%</td>
<td>-5.29%</td>
</tr>
<tr>
<td>Best year</td>
<td>7.57%</td>
<td>10.85%</td>
<td>8.60%</td>
<td>9.69%</td>
<td>8.56%</td>
<td>10.32%</td>
</tr>
<tr>
<td>Worst year</td>
<td>-11.03%</td>
<td>-6.76%</td>
<td>-6.80%</td>
<td>-6.69%</td>
<td>-8.55%</td>
<td>-7.48%</td>
</tr>
<tr>
<td>Min 252-day return</td>
<td>-14.73%</td>
<td>-15.57%</td>
<td>-8.32%</td>
<td>-8.33%</td>
<td>-16.25%</td>
<td>-9.17%</td>
</tr>
<tr>
<td>5th perc of 252-day return</td>
<td>-8.70%</td>
<td>-5.92%</td>
<td>-5.43%</td>
<td>-5.47%</td>
<td>-7.70%</td>
<td>-6.19%</td>
</tr>
<tr>
<td>50th perc of 252-day return</td>
<td>1.76%</td>
<td>1.85%</td>
<td>1.81%</td>
<td>1.92%</td>
<td>2.00%</td>
<td>2.75%</td>
</tr>
</tbody>
</table>
Let us now comment on the return and risk characteristics of the equity pension funds during the whole examination period, summarized in the table (4).

The highest absolute return has been achieved by the VUB fund, while the lowest by Postova fund. The latter is, however, highly attributable to the recent poor performance in 2016. The best day and the worst day range from 1.47% to 3.12% in the first case and from −4.94% to −2.05% in the latter. The most extreme values (both the highest best day and the lowest worst day among funds) have been recorded by the Allianz fund, which is related to its very high equity exposure. The least extreme values (both the lowest best day and the highest worst day among funds) have been recorded by the Axa fund, which, on the contrary, is related to its lower risk exposure. The average of the best day returns across funds is lower than the average of the worst day returns (both in absolute values), suggesting the negatively skewed returns. The same conclusion regarding the most and the least extreme values attributed to Allianz and Axa respectively may be applied to the monthly returns. However, the average of the best month returns across funds is now almost identical to the average of the worst month returns (in absolute values), suggesting skewness near zero. With the best and the worst year, the pattern among funds is less consistent and the results are mixed. With the minimum and the maximum rolling 252-day return spread of the funds, the results are similar to the best/worst day, that is, the highest extremes are recorded by Allianz, while the lowest by Axa. The highest median rolling 252-day return belongs to VUB (corresponding to the highest absolute return), followed by Postova and other companies with a little dispersion among them (with an

| 95th perc of 252-day return | 9.74% | 14.62% | 9.07% | 10.82% | 10.83% | 12.87% |
| Max 252-day return | 16.01% | 25.81% | 15.27% | 19.76% | 19.22% | 21.97% |
| Historical volatility | 4.65% | 6.41% | 3.21% | 3.90% | 5.18% | 4.82% |
| Min 252-day volatility | 0.16% | 0.14% | 0.08% | 0.18% | 0.17% | 0.19% |
| 5th perc of 252-day volatility | 0.20% | 0.22% | 0.10% | 0.19% | 0.20% | 0.21% |
| 50th perc of 252-day volatility | 2.33% | 2.72% | 2.29% | 2.63% | 3.32% | 2.58% |
| 95th perc of 252-day volatility | 9.06% | 14.03% | 6.06% | 7.70% | 10.67% | 10.44% |
| Max 252-day volatility | 10.96% | 16.11% | 7.01% | 8.63% | 12.66% | 11.02% |
| Semi deviation | 3.43% | 4.71% | 2.37% | 2.89% | 3.95% | 3.56% |
| Downside deviation | 3.40% | 4.67% | 2.32% | 2.84% | 3.93% | 3.51% |
| Max drawdown | -15.14% | -22.87% | -9.07% | -12.80% | -20.73% | -14.45% |
| 50th perc of TS drawdown | -5.05% | -3.20% | -3.02% | -2.00% | -3.38% | -2.27% |
| 5th perc of TS drawdown | -11.60% | -9.04% | -7.67% | -7.74% | -10.02% | -8.62% |
| 50th perc of list of drawdowns | -0.23% | -0.13% | -0.15% | -0.18% | -0.12% | -0.12% |
| 5th perc of list of drawdowns | -2.86% | -5.36% | -2.23% | -2.28% | -5.17% | -2.56% |
| Time to recovery | 2434 | 2006 | 1981 | 1916 | 2111 | 1762 |
| Value at risk | -0.48% | -0.66% | -0.33% | -0.40% | -0.53% | -0.49% |
exception of VUB, which wins by far in this measure). Worth noting is the case of Postova whose absolute return is worst, however median of the 252-day return is the second best after VUB. This confirms the aforementioned hypothesis that the return of Postova has been worsened mainly by the past few months.

The lowest historical volatility, and also the lowest minimum, median and maximum of the rolling 252-day volatility has been achieved by Axa. This is in line with the paragraph above. On the other hand, the highest historical volatility and the highest maximum of the rolling 252-day volatility is attributed to Allianz, while the highest median 252-day volatility to Postova. This suggests, that the risk of the Allianz fund may have been increased only recently, but by a high mark, whereas the risk of Postova has been consistently higher than that of the other funds. The semi deviation and downside deviation do not reveal any surprising results and their pattern correspond to that of the historical volatility. Ranking among funds is the same whether we use historical volatility, semi deviation or downside deviation. The maximum drawdown ranking among funds also approximately corresponds to the ranking by volatility. The highest drawdown was recorded by Allianz, recently, and the lowest by Axa (see also the figure (22)). The 50th and the 5th percentile of the time series of the drawdowns now reveal a different structure. The worst fund in this measure is Aegon, which we may interpret as more of the time spent in higher drawdowns than for any other fund. The least negative value for 50th percentile and almost the least negative for the 5th percentile is, on the contrary, recorded by ING, which suggests their fund spent less time in drawdowns (or in lower drawdowns) than the other funds. Values for the percentile of the list of local maximum drawdowns are markedly lower than values for the percentile of the time series of the drawdowns. This relates a lot to the drawdown experienced by the funds in 2008, which took them a long time to recover. This drawdown, therefore, spans a large part of the time series of the drawdown and hence lifts the median up, whereas in the list of the drawdowns this drawdown is included only as a single value, not influencing the median markedly. The 5th percentile of the list of the drawdowns is more in line with the maximum drawdowns or volatility, when compared among funds. The maximum time to recovery has been suffered during the sole occasion of the crisis of 2008. We may observe that even for the best fund it took almost 5 years to recover all of the losses and create a new maximum. For the worst, it took more than 6 and a half years. Value at Risk is again very much in line with the maximum drawdown or the historical volatility, that is, lowest for Allianz fund and highest for Axa.
Table 5

Return to risk and benchmark related measures of the pension funds calculated for the whole period starting on 2.1.2006 and ending on 21.3.2016. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Postova</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharpe ratio</td>
<td>-0.03</td>
<td>0.10</td>
<td>0.14</td>
<td>0.13</td>
<td>-0.12</td>
<td>0.21</td>
</tr>
<tr>
<td>Min 252-day Sharpe</td>
<td>-3.31</td>
<td>-3.10</td>
<td>-3.06</td>
<td>-3.35</td>
<td>-3.58</td>
<td>-3.58</td>
</tr>
<tr>
<td>5th perc of 252-day Sharpe</td>
<td>-2.02</td>
<td>-2.22</td>
<td>-2.19</td>
<td>-2.24</td>
<td>-2.58</td>
<td>-2.39</td>
</tr>
<tr>
<td>50th perc of 252-day Sharpe</td>
<td>0.96</td>
<td>0.37</td>
<td>1.31</td>
<td>1.02</td>
<td>0.89</td>
<td>1.13</td>
</tr>
<tr>
<td>95th perc of 252-day Sharpe</td>
<td>3.65</td>
<td>3.57</td>
<td>3.49</td>
<td>4.93</td>
<td>5.72</td>
<td>9.58</td>
</tr>
<tr>
<td>Max 252-day Sharpe</td>
<td>4.52</td>
<td>4.31</td>
<td>4.65</td>
<td>6.26</td>
<td>8.78</td>
<td>13.46</td>
</tr>
<tr>
<td>Sortino ratio</td>
<td>0.43</td>
<td>0.47</td>
<td>0.88</td>
<td>0.74</td>
<td>0.25</td>
<td>0.74</td>
</tr>
<tr>
<td>DVR ratio</td>
<td>-0.01</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
<td>-0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>Calmar ratio</td>
<td>0.10</td>
<td>0.10</td>
<td>0.23</td>
<td>0.16</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td>CAR to 5th perc of TS dd</td>
<td>0.13</td>
<td>0.24</td>
<td>0.27</td>
<td>0.27</td>
<td>0.10</td>
<td>0.30</td>
</tr>
<tr>
<td>CAR to 5th perc of list of dd</td>
<td>0.51</td>
<td>0.41</td>
<td>0.92</td>
<td>0.92</td>
<td>0.19</td>
<td>1.01</td>
</tr>
<tr>
<td>Omega (daily returns)</td>
<td>1.09</td>
<td>1.10</td>
<td>1.16</td>
<td>1.13</td>
<td>1.06</td>
<td>1.15</td>
</tr>
<tr>
<td>Omega (monthly returns)</td>
<td>1.44</td>
<td>1.49</td>
<td>1.73</td>
<td>1.59</td>
<td>1.24</td>
<td>1.75</td>
</tr>
<tr>
<td>Return to max loss (daily)</td>
<td>0.0021</td>
<td>0.0019</td>
<td>0.0041</td>
<td>0.0029</td>
<td>0.0013</td>
<td>0.0031</td>
</tr>
<tr>
<td>Return to max loss (monthly)</td>
<td>0.025</td>
<td>0.025</td>
<td>0.058</td>
<td>0.040</td>
<td>0.015</td>
<td>0.042</td>
</tr>
<tr>
<td>Ulcer performance index</td>
<td>-0.02</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>-0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>t-statistic</td>
<td>1.07</td>
<td>1.19</td>
<td>2.07</td>
<td>1.76</td>
<td>0.68</td>
<td>1.78</td>
</tr>
<tr>
<td>p-value of t-statistic</td>
<td>0.143</td>
<td>0.117</td>
<td>0.019</td>
<td>0.039</td>
<td>0.249</td>
<td>0.038</td>
</tr>
<tr>
<td>Skewness (daily)</td>
<td>-0.61</td>
<td>-0.89</td>
<td>-0.67</td>
<td>-1.01</td>
<td>-1.31</td>
<td>-0.83</td>
</tr>
<tr>
<td>Skewness (monthly)</td>
<td>-1.08</td>
<td>-0.49</td>
<td>-0.66</td>
<td>-0.36</td>
<td>-1.07</td>
<td>-0.51</td>
</tr>
<tr>
<td>Kurtosis (daily)</td>
<td>15.47</td>
<td>20.75</td>
<td>11.46</td>
<td>15.34</td>
<td>21.91</td>
<td>15.38</td>
</tr>
<tr>
<td>Kurtosis (monthly)</td>
<td>6.80</td>
<td>7.88</td>
<td>3.28</td>
<td>4.12</td>
<td>6.97</td>
<td>5.22</td>
</tr>
<tr>
<td>Fractal efficiency</td>
<td>0.039</td>
<td>0.042</td>
<td>0.071</td>
<td>0.060</td>
<td>0.023</td>
<td>0.063</td>
</tr>
<tr>
<td>K-ratio</td>
<td>0.20</td>
<td>0.33</td>
<td>0.33</td>
<td>0.37</td>
<td>0.28</td>
<td>0.41</td>
</tr>
<tr>
<td>Jensen’s alpha</td>
<td>-0.41%</td>
<td>0.24%</td>
<td>0.19%</td>
<td>0.23%</td>
<td>-0.92%</td>
<td>0.67%</td>
</tr>
<tr>
<td>Beta</td>
<td>0.95</td>
<td>1.45</td>
<td>0.69</td>
<td>0.82</td>
<td>1.07</td>
<td>1.02</td>
</tr>
<tr>
<td>Black-Treynor ratio</td>
<td>-0.44%</td>
<td>0.17%</td>
<td>0.28%</td>
<td>0.28%</td>
<td>-0.86%</td>
<td>0.66%</td>
</tr>
<tr>
<td>Tracking error</td>
<td>1.63%</td>
<td>2.82%</td>
<td>1.72%</td>
<td>1.65%</td>
<td>1.61%</td>
<td>1.06%</td>
</tr>
<tr>
<td>Information ratio</td>
<td>-0.27</td>
<td>0.10</td>
<td>0.08</td>
<td>0.12</td>
<td>-0.58</td>
<td>0.64</td>
</tr>
<tr>
<td>Treynor ratio</td>
<td>-0.18%</td>
<td>0.38%</td>
<td>0.59%</td>
<td>0.57%</td>
<td>-0.62%</td>
<td>0.94%</td>
</tr>
</tbody>
</table>

We will now focus on the risk to return characteristics and benchmark related measures of the equity pension funds during the whole examination period, summarized in the table (5).

Not every fund was able to generate the return above the 3-month Euribor during the whole sample. All of the Sharpe ratios are not high, which is primarily the result of low absolute returns when compared to the risk free rate. Nevertheless, the highest Sharpe ratio has been recorded by VUB and the lowest by Postova. The situation is slightly different if we analyze
the distribution of the rolling 252-day Sharpe ratios. The median is highest for Axa and lowest for Allianz, suggesting higher dispersion of the Sharpe ratio for example for the VUB fund, which scored the best in the entire sample Sharpe ratio measure. Minimum rolling 252-day Sharpe ratio value is similar for all funds, while the maximum varies more significantly, with the clear winner of VUB. The same applies to the 5th and 95th percentile of the rolling 252-day Sharpe ratios respectively. The Sortino ratio with minimum acceptable return threshold set to zero is highest for the Axa fund, followed by ING and VUB, and lowest for Postova. The reason why this ranking differs from that of created by the Sharpe ratio is that the risk free rate used in the Sharpe ratio subtracted the substantial portion of the Axa fund’s return and a less substantial portion of the VUB fund’s return. This caused VUB to win on the Sharpe ratio but lose on the Sortino ratio.

The price paths of all of the funds were very similar, which we were already able to observe from the figure (19). This translates into the same ranking of the funds based on the DVR ratio when compared to that based on the Sharpe ratio ($R^2$ for the price paths of the funds were not very distinct from each other). The ratio of the annualized return to maximum drawdown, the Calmar ratio, is highest for Axa, followed by VUB and lowest for Postova. If we, however, analyze the ratio of the annualized return to the 5th percentile of either the time series of the drawdown or list of the drawdowns, the highest value is again achieved by VUB and lowest by Postova. The reason behind this ranking switch is exactly the one we explained in sections (3.3.4) and (3.3.5). More specifically, the maximum drawdown is just a single point measure and can be heavily influenced by one-off events and outliers, whereas the percentile of the drawdown should represent the more consistent and stable measure of the tail risk. The Omega ratio, whether calculated from the daily or the monthly returns is highest for the Axa and VUB funds, while lowest for the Postova. This is consistent with several aforementioned measures. The return to maximal loss ratio has the same disadvantage as the Calmar ratio, because its denominator consists of a single point measure, which can be highly sensitive to outliers. The ranking of the funds (winner Axa, loser Postova) reflects exactly this fact, because it is heavily dependent on the worst day/month value, and less dependent on the average daily/monthly return which has been more similar across funds. On the measure of the Ulcer performance index scores highest the VUB fund and Allianz, while lowest Postova, which can be mostly attributed to not outperforming the 3-month Euribor and thus earning a negative numerator in the measure.

When analyzing the statistical significance of the daily returns of the funds, we have come to
the conclusion that not every fund was able to record the return significantly greater than zero. Only the funds of Axa, ING and VUB were able to generate the statistically significant return on the 10% and 5% level of significance and no fund was able to achieve a statistically significant return on the 1% level of significance. The t-statistic was highest for the Axa fund. The analysis above suggests that the returns of the funds may be negatively skewed, as consistent with the standard equity market behavior. Indeed, the skewness of either daily or monthly returns for all of the funds is negative. The most (least) negative value was recorded by Aegon (ING) for monthly and Postova (Aegon) for daily skewness. The pattern among funds, however, is not so clear as with several other measures. Kurtosis for all of the funds is much higher than kurtosis of the standard normal distribution, both on the daily as well as the monthly frequency. The only fund which was able to approach kurtosis of the normal distribution was the fund managed by Axa with its monthly returns. On the fractal efficiency measure, the highest value was recorded by Axa and the lowest by Postova. In terms of the linearity of the returns as measured by the K-ratio, the highest value was achieved by the VUB fund and the lowest by the Aegon fund.

As we have already explained in the section (4.2.2), our choice of the benchmark is a synthetic fund created as an average of all of the funds under examination, rather than the specific index benchmark. This means that, by definition of our construction, some funds will always be under-performing on a specific measure (with negative, for example Jensen’s alphas) and some will always be outperforming (with positive, for example Information ratios). This will depend on their scoring against the average, that is against the benchmark. Thus, the overall results should be regarded more as the fund’s ranking against the average rather than as an answer to the question whether the fund is beating its standard benchmark.

The negative annualized Jensen’s alpha, which may be interpreted as an alpha versus the average fund, has been recorded by Postova and Aegon, while the highest Jensen’s alpha has been achieved by VUB. This is closely related to the highest absolute return of the VUB, but does not have to necessarily be so, when we consider the methodology explained in the section (3.5.1). Moreover, the aforementioned alpha coefficient was statistically significant at the 5% level only for the VUB fund (p-values not reported here, are available from the author upon request). The majority of the funds had the beta close to 1, which had been expected, given the construction of the benchmark. The slight exceptions are the fund of Allianz with the beta of almost 1.5 and the fund of Axa with, on the contrary, beta of less than 0.7. The beta in this case can be interpreted as the fund’s sensitivity to the average movement of all of its peer funds.
The results are thus consistent with the previously calculated risk measures, such as historical volatility, according to which Allianz manages the most risky fund and Axa the least risky. The beta coefficients were statistically significant on all standard levels of significance with p-values of the order $10^{-200}$ and less (not reported here and available from the author upon request). The ranking of the funds by the Black-Treynor ratio is similar to the ranking by the Jensen’s alpha with the exception of Axa (second place, thanks to the lower beta), ING (third place) and Allianz (fourth place). The annualized tracking error against the benchmark has been lowest for the VUB fund and highest for the Allianz fund. On the measure of the information ratio, the highest value has been achieved by VUB by a wide margin, followed by the close values of ING, Allianz and Axa. The lowest information ratio has been achieved by Postova. Finally, ranking of the funds by the Treynor ratio is identical to the ranking by the Black-Treynor ratio.

### 4.3.3 First Part Of The Sample

#### Table 6

Return and risk measures of the pension funds calculated in the first period starting on 2.1.2006 and ending on 30.6.2009. See section (4.2) for the details regarding calculation.
We will now briefly comment on the return and risk measures of the funds during the first examination period starting in 2006 and ending in the middle of 2009 listed in the table (6). To not exhaust a reader we will point out only some notable differences.

During the period hit by the financial crisis the returns for all funds were lower than for the whole sample. Half of the funds recorded negative return (see also the figure (20)). Statistics of the best/worst day/month/year are more consistent across funds when compared to the entire sample, suggesting pension managers invested more uniformly. The highest difference between best/worst month/year and also minimal/maximal 252-day return and the 5th/95th percentile of the 252-day return has been achieved by Aegon, suggesting the most risky portfolio. The contrary (that is, the smallest differences) has been recorded by ING.

The historical volatility across funds was much more uniform than during the entire sample and also much lower than the volatility of the typical world pension funds, suggesting conservative investment policies. The highest dispersion in the distribution of the volatility (when analyzed via percentiles of the rolling volatility listed in the table (6)) had again the fund managed by Aegon and the lowest the fund managed by ING. The ranking of the funds based on the semi deviation or the downside deviation is identical to that based on the historical volatility. The maximum drawdown figure ranged from $-15.14\%$ for Aegon to $-8.86\%$ for ING. The time to recovery has been similar for all funds, because they had all peaked at approximately the same time and none of the funds recovered from its drawdown by the end of this period (see also figure (22)).
Table 7

Return to risk and benchmark related measures of the pension funds calculated in the first period starting on 2.1.2006 and ending on 30.6.2009. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Postova</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharpe ratio</td>
<td>-1.01</td>
<td>-1.02</td>
<td>-1.09</td>
<td>-1.21</td>
<td>-1.14</td>
<td>-1.18</td>
</tr>
<tr>
<td>Min 252-day Sharpe</td>
<td>-3.31</td>
<td>-3.10</td>
<td>-3.05</td>
<td>-3.34</td>
<td>-3.58</td>
<td>-3.58</td>
</tr>
<tr>
<td>5th perc of 252-day Sharpe</td>
<td>-2.31</td>
<td>-2.58</td>
<td>-2.58</td>
<td>-2.74</td>
<td>-3.11</td>
<td>-3.00</td>
</tr>
<tr>
<td>50th perc of 252-day Sharpe</td>
<td>-1.27</td>
<td>-1.38</td>
<td>-1.18</td>
<td>-1.33</td>
<td>-1.21</td>
<td>-1.25</td>
</tr>
<tr>
<td>95th perc of 252-day Sharpe</td>
<td>1.78</td>
<td>2.13</td>
<td>1.74</td>
<td>0.56</td>
<td>2.15</td>
<td>1.68</td>
</tr>
<tr>
<td>Max 252-day Sharpe</td>
<td>2.36</td>
<td>2.84</td>
<td>2.35</td>
<td>1.22</td>
<td>2.98</td>
<td>2.37</td>
</tr>
<tr>
<td>Sortino ratio</td>
<td>-0.25</td>
<td>0.23</td>
<td>0.11</td>
<td>-0.07</td>
<td>-0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>DVR ratio</td>
<td>-0.08</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Calmar ratio</td>
<td>-0.06</td>
<td>0.06</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>CAR to 5th perc of TS dd</td>
<td>-0.07</td>
<td>0.07</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>CAR to 5th perc of list of dd</td>
<td>-0.37</td>
<td>0.32</td>
<td>0.15</td>
<td>-0.08</td>
<td>-0.23</td>
<td>0.13</td>
</tr>
<tr>
<td>Omega (daily returns)</td>
<td>0.97</td>
<td>1.04</td>
<td>1.02</td>
<td>0.99</td>
<td>0.99</td>
<td>1.02</td>
</tr>
<tr>
<td>Omega (monthly returns)</td>
<td>0.78</td>
<td>0.82</td>
<td>0.77</td>
<td>0.85</td>
<td>0.95</td>
<td>1.04</td>
</tr>
<tr>
<td>Return to max loss (daily)</td>
<td>-0.0017</td>
<td>0.0020</td>
<td>0.0011</td>
<td>-0.0004</td>
<td>-0.0004</td>
<td>0.0009</td>
</tr>
<tr>
<td>Return to max loss (monthly)</td>
<td>-0.012</td>
<td>0.015</td>
<td>0.010</td>
<td>-0.003</td>
<td>-0.005</td>
<td>0.009</td>
</tr>
<tr>
<td>Ulcer performance index</td>
<td>-0.78</td>
<td>-0.46</td>
<td>-0.44</td>
<td>-0.45</td>
<td>-0.41</td>
<td>-0.33</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-0.31</td>
<td>0.35</td>
<td>0.18</td>
<td>-0.06</td>
<td>-0.12</td>
<td>0.17</td>
</tr>
<tr>
<td>p-value of t-statistic</td>
<td>0.622</td>
<td>0.363</td>
<td>0.427</td>
<td>0.526</td>
<td>0.549</td>
<td>0.432</td>
</tr>
<tr>
<td>Skewness (daily)</td>
<td>-0.21</td>
<td>-0.52</td>
<td>-0.51</td>
<td>-0.62</td>
<td>-0.80</td>
<td>-0.66</td>
</tr>
<tr>
<td>Skewness (monthly)</td>
<td>-2.40</td>
<td>-1.80</td>
<td>-1.31</td>
<td>-1.76</td>
<td>-1.96</td>
<td>-1.55</td>
</tr>
<tr>
<td>Kurtosis (daily)</td>
<td>13.95</td>
<td>4.79</td>
<td>4.65</td>
<td>4.48</td>
<td>35.92</td>
<td>4.67</td>
</tr>
<tr>
<td>Kurtosis (monthly)</td>
<td>7.05</td>
<td>4.16</td>
<td>1.60</td>
<td>3.53</td>
<td>5.10</td>
<td>2.43</td>
</tr>
<tr>
<td>Fractal efficiency</td>
<td>-0.018</td>
<td>0.015</td>
<td>0.007</td>
<td>-0.004</td>
<td>-0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>K-ratio</td>
<td>-0.16</td>
<td>0.05</td>
<td>0.04</td>
<td>-0.07</td>
<td>-0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>Jensen’s alpha</td>
<td>0.70%</td>
<td>0.27%</td>
<td>0.00%</td>
<td>-0.43%</td>
<td>-0.49%</td>
<td>-0.04%</td>
</tr>
<tr>
<td>Beta</td>
<td>1.37</td>
<td>0.93</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>Black-Treynor ratio</td>
<td>0.51%</td>
<td>0.29%</td>
<td>0.00%</td>
<td>-0.46%</td>
<td>-0.53%</td>
<td>-0.05%</td>
</tr>
<tr>
<td>Tracking error</td>
<td>2.01%</td>
<td>0.68%</td>
<td>0.78%</td>
<td>0.92%</td>
<td>0.92%</td>
<td>0.73%</td>
</tr>
<tr>
<td>Information ratio</td>
<td>-0.35</td>
<td>0.76</td>
<td>0.36</td>
<td>-0.14</td>
<td>-0.23</td>
<td>0.30</td>
</tr>
<tr>
<td>Treynor ratio</td>
<td>-3.29%</td>
<td>-3.50%</td>
<td>-3.79%</td>
<td>-4.25%</td>
<td>-4.31%</td>
<td>-3.84%</td>
</tr>
</tbody>
</table>

The main remarks regarding return to risk and benchmark related measures (listed in the table (7)) in the first part of the sample are as follows.

The Sharpe ratio as well as the median Sharpe ratio has been negative for all of the funds, implying that no fund was able to beat the return of the 3-month Euribor during the examination period. The distributions of the Sharpe ratios are close to each other when compared across funds. The DVR ratios are near zero and close to each other, because of the near zero $R^2$ from
the regressions of the funds’ prices against time. The Calmar ratio has been highest for the Allianz fund, given its highest return and almost the highest (the least negative) maximum drawdown figure. The same applies for the ratio of the annualized return against 5th percentile of the drawdown (either time series or list of the drawdowns). The Omega ratio calculated from the daily returns shows very small differences across funds. When calculated from the monthly returns, VUB achieved the highest value. The Ulcer performance index has been negative for all funds because of the under-performance against the 3-month Euribor. Moreover (and not surprisingly), no fund was able to achieve the daily return significantly greater than zero. The values of the K-ratio and the fractal efficiency are very low, because of the low returns generated by the funds.

Jensen’s alpha has been surprisingly highest for Aegon (and lowest for Postova), however none of the alpha coefficients has been statistically significant (p-values not reported here, are available from the author upon request) at the 10% level (the highest p-value 0.31). That suggests that the alpha may have been influenced by outliers and/or the sample may be not large enough. The beta for all of the funds except Aegon has been nearly identical. For Aegon, the higher beta suggests higher risk taking by the managers. The tracking error has been highest for Aegon, consistent with the fact that it also ranked top for several other risk measures. The Treynor ratio is highly negative, because the return on a risk free rate largely exceeded the return generated by the funds.

To sum up, the performance measures among funds during the period marked by the financial crisis did not vary notably and they were the worst among our three chosen sub-parts of the whole sample.
4.3.4 Second Part Of The Sample

Table 8

Return and risk measures of the pension funds calculated in the second period starting on 30.6.2009 and ending on 28.3.2013. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Postova</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat return</td>
<td>6.88%</td>
<td>6.71%</td>
<td>6.07%</td>
<td>7.53%</td>
<td>8.08%</td>
</tr>
<tr>
<td>Annualized return</td>
<td>1.79%</td>
<td>1.75%</td>
<td>1.59%</td>
<td>1.96%</td>
<td>2.10%</td>
</tr>
<tr>
<td>Best day</td>
<td>0.52%</td>
<td>0.18%</td>
<td>0.16%</td>
<td>0.18%</td>
<td>0.17%</td>
</tr>
<tr>
<td>Worst day</td>
<td>-0.16%</td>
<td>-0.15%</td>
<td>-0.27%</td>
<td>-0.16%</td>
<td>-0.51%</td>
</tr>
<tr>
<td>Best month</td>
<td>0.65%</td>
<td>0.68%</td>
<td>0.55%</td>
<td>0.70%</td>
<td>0.74%</td>
</tr>
<tr>
<td>Worst month</td>
<td>-0.06%</td>
<td>-0.16%</td>
<td>-0.35%</td>
<td>-0.35%</td>
<td>-0.29%</td>
</tr>
<tr>
<td>Best year</td>
<td>3.51%</td>
<td>2.94%</td>
<td>2.20%</td>
<td>3.44%</td>
<td>4.46%</td>
</tr>
<tr>
<td>Worst year</td>
<td>0.08%</td>
<td>0.55%</td>
<td>0.33%</td>
<td>0.14%</td>
<td>0.36%</td>
</tr>
<tr>
<td>Min 252-day return</td>
<td>1.01%</td>
<td>0.58%</td>
<td>0.90%</td>
<td>1.12%</td>
<td>0.94%</td>
</tr>
<tr>
<td>5th perc of 252-day return</td>
<td>1.07%</td>
<td>0.62%</td>
<td>0.98%</td>
<td>1.31%</td>
<td>1.06%</td>
</tr>
<tr>
<td>50th perc of 252-day return</td>
<td>3.35%</td>
<td>3.00%</td>
<td>2.10%</td>
<td>3.28%</td>
<td>4.34%</td>
</tr>
<tr>
<td>95th perc of 252-day return</td>
<td>3.52%</td>
<td>3.20%</td>
<td>2.31%</td>
<td>3.57%</td>
<td>4.64%</td>
</tr>
<tr>
<td>Historical volatility</td>
<td>0.47%</td>
<td>0.46%</td>
<td>0.43%</td>
<td>0.39%</td>
<td>0.50%</td>
</tr>
<tr>
<td>Min 252-day volatility</td>
<td>0.16%</td>
<td>0.14%</td>
<td>0.08%</td>
<td>0.18%</td>
<td>0.17%</td>
</tr>
<tr>
<td>5th perc of 252-day volatility</td>
<td>0.17%</td>
<td>0.14%</td>
<td>0.10%</td>
<td>0.18%</td>
<td>0.17%</td>
</tr>
<tr>
<td>50th perc of 252-day volatility</td>
<td>0.31%</td>
<td>0.24%</td>
<td>0.12%</td>
<td>0.31%</td>
<td>0.23%</td>
</tr>
<tr>
<td>95th perc of 252-day volatility</td>
<td>0.73%</td>
<td>0.71%</td>
<td>0.56%</td>
<td>0.59%</td>
<td>0.57%</td>
</tr>
<tr>
<td>Max 252-day volatility</td>
<td>0.73%</td>
<td>0.79%</td>
<td>0.75%</td>
<td>0.60%</td>
<td>0.71%</td>
</tr>
<tr>
<td>Semi deviation</td>
<td>0.26%</td>
<td>0.31%</td>
<td>0.31%</td>
<td>0.28%</td>
<td>0.40%</td>
</tr>
<tr>
<td>Downside deviation</td>
<td>0.21%</td>
<td>0.27%</td>
<td>0.28%</td>
<td>0.23%</td>
<td>0.37%</td>
</tr>
<tr>
<td>Max drawdown</td>
<td>-0.16%</td>
<td>-0.45%</td>
<td>-0.49%</td>
<td>-0.40%</td>
<td>-0.64%</td>
</tr>
<tr>
<td>50th perc of TS drawdown</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>5th perc of TS drawdown</td>
<td>-0.06%</td>
<td>-0.13%</td>
<td>-0.09%</td>
<td>-0.17%</td>
<td>-0.21%</td>
</tr>
<tr>
<td>50th perc of list of drawdowns</td>
<td>-0.02%</td>
<td>-0.01%</td>
<td>-0.01%</td>
<td>-0.01%</td>
<td>-0.01%</td>
</tr>
<tr>
<td>5th perc of list of drawdowns</td>
<td>-0.09%</td>
<td>-0.17%</td>
<td>-0.12%</td>
<td>-0.09%</td>
<td>-0.24%</td>
</tr>
<tr>
<td>Time to recovery</td>
<td>22</td>
<td>48</td>
<td>64</td>
<td>73</td>
<td>94</td>
</tr>
<tr>
<td>Value at risk</td>
<td>-0.04%</td>
<td>-0.04%</td>
<td>-0.04%</td>
<td>-0.03%</td>
<td>-0.04%</td>
</tr>
</tbody>
</table>

We will now briefly comment on the return and risk measures of the funds during the second examination period starting in the middle of 2009 and ending in the first quarter of 2013 listed in the table (8). This period has been characterized by low risk portfolios of all funds, because of the legal obligation to compensate losses occurring on any 6-calendar month rolling period.

Portfolios of the pension funds consisted mainly of short maturity fixed income assets, such as treasury bills, covered bonds and term deposits. All of the funds achieved to deliver positive returns. This is primarily due to the declining interest rates, which, naturally, caused the raise
in the prices of fixed income securities. Moreover, no credit events, that is defaults or suspension of interest payments, occurred on the securities held by the funds. Annualized returns exceeded the risk free rate (0.86% p.a.) by as low as 0.73% p.a. (Axa) to as high as 1.59% p.a. (VUB). There has not been any consistent pattern in the best/worst day/month/year among funds that may have helped to assess the differences in their level of riskiness. The highest differences between the minimal/maximal 252-day return and the 5th/95th percentile of the 252-day return have been recorded by Postova while the contrary (that is, the smallest differences) has been recorded by Axa. Volatility and also the distribution of the volatility has been remarkably stable, very low in absolute terms and also similar across funds, which is the consequence of low risk investments. The maximum drawdown reached as high value as −0.16% for Aegon and −0.81% for VUB, which suffered the largest drawdown during this period among funds. The time to recovery also counted in just a few weeks.

Table 9

Return to risk and benchmark related measures of the pension funds calculated in the second period starting on 30.6.2009 and ending on 28.3.2013. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th></th>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Postova</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharpe ratio</td>
<td>1.99</td>
<td>1.91</td>
<td>1.70</td>
<td>2.80</td>
<td>2.47</td>
<td>3.40</td>
</tr>
<tr>
<td>Min 252-day Sharpe</td>
<td>-0.95</td>
<td>-2.45</td>
<td>0.28</td>
<td>-0.70</td>
<td>-0.87</td>
<td>1.36</td>
</tr>
<tr>
<td>5th perc of 252-day Sharpe</td>
<td>-0.41</td>
<td>-2.10</td>
<td>0.65</td>
<td>-0.11</td>
<td>-0.30</td>
<td>3.01</td>
</tr>
<tr>
<td>50th perc of 252-day Sharpe</td>
<td>1.41</td>
<td>0.30</td>
<td>2.63</td>
<td>2.51</td>
<td>2.20</td>
<td>4.98</td>
</tr>
<tr>
<td>95th perc of 252-day Sharpe</td>
<td>4.05</td>
<td>3.91</td>
<td>3.93</td>
<td>5.35</td>
<td>7.76</td>
<td>12.34</td>
</tr>
<tr>
<td>Max 252-day Sharpe</td>
<td>4.54</td>
<td>4.33</td>
<td>4.75</td>
<td>6.30</td>
<td>8.83</td>
<td>13.54</td>
</tr>
<tr>
<td>Sortino ratio</td>
<td>8.37</td>
<td>6.59</td>
<td>5.62</td>
<td>8.34</td>
<td>5.69</td>
<td>6.48</td>
</tr>
<tr>
<td>DVR ratio</td>
<td>1.87</td>
<td>1.74</td>
<td>1.66</td>
<td>2.70</td>
<td>2.24</td>
<td>3.32</td>
</tr>
<tr>
<td>Calmar ratio</td>
<td>11.29</td>
<td>3.89</td>
<td>3.24</td>
<td>4.85</td>
<td>3.27</td>
<td>3.04</td>
</tr>
<tr>
<td>CAR to 5th perc of TS dd</td>
<td>27.58</td>
<td>13.66</td>
<td>18.38</td>
<td>11.84</td>
<td>10.13</td>
<td>18.99</td>
</tr>
<tr>
<td>CAR to 5th perc of list of dd</td>
<td>19.78</td>
<td>10.02</td>
<td>13.54</td>
<td>21.25</td>
<td>8.69</td>
<td>22.13</td>
</tr>
<tr>
<td>Omega (daily returns)</td>
<td>2.57</td>
<td>2.23</td>
<td>2.61</td>
<td>2.83</td>
<td>2.87</td>
<td>3.53</td>
</tr>
<tr>
<td>Omega (monthly returns)</td>
<td>60.03</td>
<td>25.51</td>
<td>15.77</td>
<td>7.19</td>
<td>8.92</td>
<td>23.03</td>
</tr>
<tr>
<td>Return to max loss (daily)</td>
<td>0.0444</td>
<td>0.0472</td>
<td>0.0233</td>
<td>0.0469</td>
<td>0.0160</td>
<td>0.0144</td>
</tr>
<tr>
<td>Return to max loss (monthly)</td>
<td>2.514</td>
<td>0.892</td>
<td>0.377</td>
<td>0.465</td>
<td>0.599</td>
<td>0.551</td>
</tr>
<tr>
<td>Ulcer performance index</td>
<td>32.72</td>
<td>13.64</td>
<td>8.59</td>
<td>10.43</td>
<td>8.58</td>
<td>8.84</td>
</tr>
<tr>
<td>t-statistic</td>
<td>7.37</td>
<td>7.26</td>
<td>7.15</td>
<td>9.61</td>
<td>8.04</td>
<td>10.03</td>
</tr>
<tr>
<td>p-value of t-statistic</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Skewness (daily)</td>
<td>5.99</td>
<td>0.56</td>
<td>-1.09</td>
<td>-0.12</td>
<td>-5.25</td>
<td>-12.92</td>
</tr>
<tr>
<td>Skewness (monthly)</td>
<td>1.88</td>
<td>1.65</td>
<td>0.37</td>
<td>0.11</td>
<td>0.59</td>
<td>-1.44</td>
</tr>
<tr>
<td>Kurtosis (daily)</td>
<td>104.14</td>
<td>7.60</td>
<td>25.32</td>
<td>13.72</td>
<td>86.15</td>
<td>295.62</td>
</tr>
<tr>
<td>Kurtosis (monthly)</td>
<td>4.08</td>
<td>3.59</td>
<td>5.28</td>
<td>5.30</td>
<td>1.86</td>
<td>5.01</td>
</tr>
</tbody>
</table>
The return to risk and benchmark related measures for the second part of our sample are listed in the table (9).

The Sharpe ratios have been very high in the period under examination. The reason why this was so, was the very low volatility of money market instruments (on the denominator side) and the attractive return spread above the risk free rate (3-month Euribor) for the same instruments (on the numerator side). The highest Sharpe ratio and also the highest percentiles of the rolling 252-day Sharpe ratio have been by a wide margin achieved by VUB. The lowest dispersion in the distribution of the rolling 252-day Sharpe ratio as measured by the percentiles in the table (9) has been, however, achieved by Axa. Ranking of the funds based on the Sortino ratio is completely different than the one based on the Sharpe ratio. The reason is the downside deviation, which now differs more substantially from the historical volatility. The ranking of the funds based on the DVR ratio is the same as that based on the Sharpe ratio, because all of the funds had very high \( R^2 \) values (all above 0.9) in the regressions of the funds’ prices against time. The main winner in the Calmar ratio measure is the Aegon fund, thanks to its extraordinarily high (least negative) maximum drawdown figure. All funds in the measures of the ratios of annualized return against the drawdown percentiles, the Omega ratio, the return to maximal loss, the Ulcer performance index or \( t \)-statistic achieved very high values in the period under examination, which were not repeated before or after the period. This was also the only period when some funds even managed to achieve positive skewness of the returns. The fractal efficiency ratio reached exceptionally high values as well, indicating that the path of the funds’ equity curves resembled the straight linear curves the most from all of our sub-periods, see also the figure (20). The same applies to the K-ratio. The highest values for the last two measures were recorded by VUB and the lowest by Allianz.

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{Fractal efficiency} & 0.441 & 0.379 & 0.443 & 0.476 & 0.482 & 0.562 \\
\hline
\text{K-ratio} & 2.06 & 1.66 & 3.73 & 2.77 & 1.68 & 3.79 \\
\hline
\text{Jensen’s alpha} & -0.09\% & -0.47\% & -0.23\% & 0.12\% & -0.19\% & 0.86\% \\
\hline
\text{Beta} & 0.90 & 1.20 & 0.86 & 0.85 & 1.37 & 0.82 \\
\hline
\text{Black-Treynor ratio} & -0.10\% & -0.39\% & -0.27\% & 0.14\% & -0.14\% & 1.05\% \\
\hline
\text{Tracking error} & 0.36\% & 0.32\% & 0.25\% & 0.31\% & 0.39\% & 0.23\% \\
\hline
\text{Information ratio} & -0.55 & -0.78 & -1.55 & -0.16 & 0.59 & 2.84 \\
\hline
\text{Treynor ratio} & 1.02\% & 0.73\% & 0.85\% & 1.27\% & 0.99\% & 2.18\% \\
\hline
\end{array}
\]

29 Most of the money market assets held by the funds had a domicile in Slovakia or in the neighboring low risk countries, such as the Czech republic. The return spread of the aforementioned securities against the core Eurozone countries, such as Germany or France, which form the main base for Euribor calculation, was quite substantial at the time. This is because investors at the time still priced some country risk of the instruments issued by the Slovak government, which created the return premium over the core Eurozone countries.
Unequivocally highest, and statistically significant with the p-value of the order $10^{-10}$, Jensen’s alpha has been achieved by the VUB fund. The most negative, and also significant with the p-value of 0.01, Jensen’s alpha was recorded by Allianz. The tracking error has been very low and similar across funds.

To sum up, for all funds, most of the return to risk measures peaked during the second part of the sample. This is because of the very low risk employed by the fund managers.

### 4.3.5 Third Part Of The Sample

#### Table 10

Return and risk measures of the pension funds calculated in the third period starting on 28.3.2013 and ending on 21.3.2016. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th>Fund</th>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Postova</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat return</td>
<td>11.69%</td>
<td>14.97%</td>
<td>14.94%</td>
<td>15.50%</td>
<td>3.14%</td>
<td>17.76%</td>
</tr>
<tr>
<td>Annualized return</td>
<td>3.78%</td>
<td>4.79%</td>
<td>4.78%</td>
<td>4.95%</td>
<td>1.04%</td>
<td>5.64%</td>
</tr>
<tr>
<td>Best day</td>
<td>2.18%</td>
<td>3.12%</td>
<td>1.47%</td>
<td>1.66%</td>
<td>2.51%</td>
<td>2.30%</td>
</tr>
<tr>
<td>Worst day</td>
<td>-2.95%</td>
<td>-4.94%</td>
<td>-2.05%</td>
<td>-2.98%</td>
<td>-3.52%</td>
<td>-3.45%</td>
</tr>
<tr>
<td>Best month</td>
<td>4.81%</td>
<td>8.29%</td>
<td>3.43%</td>
<td>3.82%</td>
<td>6.12%</td>
<td>5.03%</td>
</tr>
<tr>
<td>Worst month</td>
<td>-4.20%</td>
<td>-7.88%</td>
<td>-2.99%</td>
<td>-4.44%</td>
<td>-6.06%</td>
<td>-5.29%</td>
</tr>
<tr>
<td>Best year</td>
<td>7.57%</td>
<td>10.85%</td>
<td>8.60%</td>
<td>9.69%</td>
<td>8.56%</td>
<td>10.32%</td>
</tr>
<tr>
<td>Worst year</td>
<td>-2.15%</td>
<td>-2.08%</td>
<td>-0.56%</td>
<td>-0.40%</td>
<td>-8.55%</td>
<td>-2.28%</td>
</tr>
<tr>
<td>Min 252-day return</td>
<td>-7.89%</td>
<td>-15.57%</td>
<td>-4.71%</td>
<td>-7.90%</td>
<td>-16.25%</td>
<td>-8.41%</td>
</tr>
<tr>
<td>5th perc of 252-day return</td>
<td>-4.59%</td>
<td>-9.58%</td>
<td>-2.41%</td>
<td>-5.22%</td>
<td>-15.05%</td>
<td>-6.44%</td>
</tr>
<tr>
<td>50th perc of 252-day return</td>
<td>6.59%</td>
<td>10.18%</td>
<td>6.90%</td>
<td>7.69%</td>
<td>7.54%</td>
<td>9.84%</td>
</tr>
<tr>
<td>95th perc of 252-day return</td>
<td>14.27%</td>
<td>21.62%</td>
<td>13.78%</td>
<td>17.77%</td>
<td>17.23%</td>
<td>19.37%</td>
</tr>
<tr>
<td>Max 252-day return</td>
<td>16.01%</td>
<td>25.81%</td>
<td>15.27%</td>
<td>19.76%</td>
<td>19.22%</td>
<td>21.97%</td>
</tr>
<tr>
<td>Historical volatility</td>
<td>7.12%</td>
<td>11.38%</td>
<td>4.87%</td>
<td>6.34%</td>
<td>8.81%</td>
<td>8.32%</td>
</tr>
<tr>
<td>Min 252-day volatility</td>
<td>2.15%</td>
<td>6.70%</td>
<td>2.55%</td>
<td>4.08%</td>
<td>4.26%</td>
<td>4.92%</td>
</tr>
<tr>
<td>5th perc of 252-day volatility</td>
<td>2.46%</td>
<td>6.84%</td>
<td>2.58%</td>
<td>4.36%</td>
<td>4.37%</td>
<td>4.94%</td>
</tr>
<tr>
<td>50th perc of 252-day volatility</td>
<td>5.15%</td>
<td>8.36%</td>
<td>3.70%</td>
<td>5.59%</td>
<td>6.81%</td>
<td>7.18%</td>
</tr>
<tr>
<td>95th perc of 252-day volatility</td>
<td>10.65%</td>
<td>15.72%</td>
<td>6.80%</td>
<td>8.48%</td>
<td>12.35%</td>
<td>10.85%</td>
</tr>
<tr>
<td>Max 252-day volatility</td>
<td>10.94%</td>
<td>16.08%</td>
<td>6.99%</td>
<td>8.62%</td>
<td>12.63%</td>
<td>10.99%</td>
</tr>
<tr>
<td>Semi deviation</td>
<td>5.26%</td>
<td>8.43%</td>
<td>3.61%</td>
<td>4.74%</td>
<td>6.72%</td>
<td>6.20%</td>
</tr>
<tr>
<td>Downside deviation</td>
<td>5.15%</td>
<td>8.27%</td>
<td>3.47%</td>
<td>4.60%</td>
<td>6.68%</td>
<td>6.03%</td>
</tr>
<tr>
<td>Max drawdown</td>
<td>-12.91%</td>
<td>-22.87%</td>
<td>-8.53%</td>
<td>-12.80%</td>
<td>-20.73%</td>
<td>-14.45%</td>
</tr>
<tr>
<td>50th perc of TS drawdown</td>
<td>-0.37%</td>
<td>-2.06%</td>
<td>-0.78%</td>
<td>-0.70%</td>
<td>-1.35%</td>
<td>-1.39%</td>
</tr>
<tr>
<td>5th perc of TS drawdown</td>
<td>-8.26%</td>
<td>-16.37%</td>
<td>-6.05%</td>
<td>-9.57%</td>
<td>-14.72%</td>
<td>-10.94%</td>
</tr>
<tr>
<td>50th perc of list of drawdowns</td>
<td>-0.21%</td>
<td>-0.40%</td>
<td>-0.27%</td>
<td>-0.23%</td>
<td>-0.31%</td>
<td>-0.34%</td>
</tr>
<tr>
<td>5th perc of list of drawdowns</td>
<td>-2.71%</td>
<td>-9.02%</td>
<td>-3.29%</td>
<td>-3.13%</td>
<td>-5.79%</td>
<td>-5.40%</td>
</tr>
<tr>
<td>Time to recovery</td>
<td>343</td>
<td>341</td>
<td>341</td>
<td>341</td>
<td>343</td>
<td>343</td>
</tr>
<tr>
<td>Value at risk</td>
<td>-0.73%</td>
<td>-1.16%</td>
<td>-0.49%</td>
<td>-0.64%</td>
<td>-0.91%</td>
<td>-0.84%</td>
</tr>
</tbody>
</table>
The third part of our sample is the latest (approximately) 3-year period starting in the second quarter of 2013 and ending on 21.3.2016. The new pension legislation made it possible for the funds to be entirely invested in equities, because of both, the loose legal limits as well as the missing obligation to deliver the positive performance. The return and risk measures over this period are listed in the table (10).

All of the funds except Postova delivered the highest returns among our three sub-samples. This is now linked primarily to the rise of the equity markets, and not negligibly also to the further fall of interest rates, which resulted into the appreciation of fixed income assets. The funds have been investing heavily into equity assets as well as into longer duration fixed income assets, which paid off. The fund with the clearly lowest return has been managed by Postova, however, this is mainly due to the very low returns in the last year of the sample, not because of the consistent under-performance over the whole period. The highest difference between the best/worst day/month among funds has been recorded by Allianz, suggesting aggressive allocation into equities. The opposite, that is the lowest difference, achieved the Axa fund. Exactly the same for both of the aforementioned funds holds also for the differences between the minimal/maximal 252-day return and the 5th/95th percentile of the 252-day return (highest for Allianz, lowest for Axa).

This is also consistent with the historical volatility, which is highest among all three sub-samples and even higher than during the financial crisis. This suggests that the significantly different investment policies are being employed by the pension funds when compared to the past. The funds managed by Allianz, ING, Postova and VUB also suffered their new maximum drawdown in this third part of the sample. It occurred during the fall of equity markets in January and February 2016. The time to recovery is almost identical for all funds and the drawdown still has not been recovered as of 21.3.2016. The primary reason is that both the equity markets together with bond assets started to plunge in April 2015.
Table 11

Return to risk and benchmark related measures of the pension funds calculated in the third period starting on 28.3.2013 and ending on 21.3.2016. See section (4.2) for the details regarding calculation.

<table>
<thead>
<tr>
<th></th>
<th>Aegon</th>
<th>Allianz</th>
<th>Axa</th>
<th>ING</th>
<th>Postova</th>
<th>VUB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharpe ratio</td>
<td>0.52</td>
<td>0.41</td>
<td>0.96</td>
<td>0.76</td>
<td>0.11</td>
<td>0.67</td>
</tr>
<tr>
<td>Min 252-day Sharpe</td>
<td>-0.74</td>
<td>-0.99</td>
<td>-0.69</td>
<td>-0.93</td>
<td>-1.28</td>
<td>-0.77</td>
</tr>
<tr>
<td>5th perc of 252-day Sharpe</td>
<td>-0.42</td>
<td>-0.61</td>
<td>-0.34</td>
<td>-0.59</td>
<td>-1.18</td>
<td>-0.58</td>
</tr>
<tr>
<td>50th perc of 252-day Sharpe</td>
<td>1.52</td>
<td>1.25</td>
<td>1.98</td>
<td>1.49</td>
<td>1.32</td>
<td>1.52</td>
</tr>
<tr>
<td>95th perc of 252-day Sharpe</td>
<td>2.68</td>
<td>2.53</td>
<td>3.63</td>
<td>3.25</td>
<td>2.50</td>
<td>2.64</td>
</tr>
<tr>
<td>Max 252-day Sharpe</td>
<td>3.05</td>
<td>2.98</td>
<td>4.09</td>
<td>3.61</td>
<td>2.78</td>
<td>2.98</td>
</tr>
<tr>
<td>Sortino ratio</td>
<td>0.73</td>
<td>0.58</td>
<td>1.38</td>
<td>1.08</td>
<td>0.16</td>
<td>0.94</td>
</tr>
<tr>
<td>DVR ratio</td>
<td>0.37</td>
<td>0.24</td>
<td>0.79</td>
<td>0.56</td>
<td>0.04</td>
<td>0.49</td>
</tr>
<tr>
<td>Calmar ratio</td>
<td>0.29</td>
<td>0.21</td>
<td>0.56</td>
<td>0.39</td>
<td>0.05</td>
<td>0.39</td>
</tr>
<tr>
<td>CAR to 5th perc of TS dd</td>
<td>0.46</td>
<td>0.29</td>
<td>0.79</td>
<td>0.52</td>
<td>0.07</td>
<td>0.52</td>
</tr>
<tr>
<td>CAR to 5th perc of list of dd</td>
<td>1.40</td>
<td>0.53</td>
<td>1.45</td>
<td>1.58</td>
<td>0.18</td>
<td>1.04</td>
</tr>
<tr>
<td>Omega (daily returns)</td>
<td>1.12</td>
<td>1.09</td>
<td>1.20</td>
<td>1.15</td>
<td>1.03</td>
<td>1.14</td>
</tr>
<tr>
<td>Omega (monthly returns)</td>
<td>1.62</td>
<td>1.32</td>
<td>1.36</td>
<td>1.72</td>
<td>1.05</td>
<td>1.77</td>
</tr>
<tr>
<td>Return to max loss (daily)</td>
<td>0.0054</td>
<td>0.0043</td>
<td>0.0094</td>
<td>0.0068</td>
<td>0.0016</td>
<td>0.0068</td>
</tr>
<tr>
<td>Return to max loss (monthly)</td>
<td>0.077</td>
<td>0.055</td>
<td>0.132</td>
<td>0.094</td>
<td>0.019</td>
<td>0.090</td>
</tr>
<tr>
<td>Ulcer performance index</td>
<td>1.11</td>
<td>0.62</td>
<td>0.59</td>
<td>0.54</td>
<td>0.09</td>
<td>0.47</td>
</tr>
<tr>
<td>t-statistic</td>
<td>0.96</td>
<td>0.81</td>
<td>1.70</td>
<td>1.37</td>
<td>0.28</td>
<td>1.21</td>
</tr>
<tr>
<td>p-value of t-statistic</td>
<td>0.168</td>
<td>0.210</td>
<td>0.045</td>
<td>0.085</td>
<td>0.390</td>
<td>0.113</td>
</tr>
<tr>
<td>Skewness (daily)</td>
<td>-0.56</td>
<td>-0.58</td>
<td>-0.58</td>
<td>-0.79</td>
<td>-0.86</td>
<td>-0.58</td>
</tr>
<tr>
<td>Skewness (monthly)</td>
<td>-0.34</td>
<td>-0.50</td>
<td>-0.64</td>
<td>-0.43</td>
<td>-0.65</td>
<td>-0.58</td>
</tr>
<tr>
<td>Kurtosis (daily)</td>
<td>5.55</td>
<td>5.19</td>
<td>4.94</td>
<td>5.52</td>
<td>6.12</td>
<td>3.99</td>
</tr>
<tr>
<td>Kurtosis (monthly)</td>
<td>1.54</td>
<td>1.15</td>
<td>0.70</td>
<td>0.19</td>
<td>1.04</td>
<td>0.77</td>
</tr>
<tr>
<td>Fractal efficiency</td>
<td>0.051</td>
<td>0.035</td>
<td>0.085</td>
<td>0.066</td>
<td>0.011</td>
<td>0.057</td>
</tr>
<tr>
<td>K-ratio</td>
<td>0.96</td>
<td>0.71</td>
<td>1.27</td>
<td>1.00</td>
<td>0.49</td>
<td>1.00</td>
</tr>
<tr>
<td>Jensen’s alpha</td>
<td>0.19%</td>
<td>-1.48%</td>
<td>1.97%</td>
<td>1.57%</td>
<td>-3.40%</td>
<td>1.26%</td>
</tr>
<tr>
<td>Beta</td>
<td>0.85</td>
<td>1.58</td>
<td>0.63</td>
<td>0.79</td>
<td>1.11</td>
<td>1.03</td>
</tr>
<tr>
<td>Black-Treynor ratio</td>
<td>0.22%</td>
<td>-0.93%</td>
<td>3.11%</td>
<td>1.98%</td>
<td>-3.06%</td>
<td>1.22%</td>
</tr>
<tr>
<td>Tracking error</td>
<td>2.05%</td>
<td>5.16%</td>
<td>3.06%</td>
<td>2.86%</td>
<td>2.75%</td>
<td>1.76%</td>
</tr>
<tr>
<td>Information ratio</td>
<td>-0.20</td>
<td>0.11</td>
<td>0.19</td>
<td>0.27</td>
<td>-1.15</td>
<td>0.81</td>
</tr>
<tr>
<td>Treynor ratio</td>
<td>4.31%</td>
<td>2.92%</td>
<td>7.31%</td>
<td>6.11%</td>
<td>0.83%</td>
<td>5.30%</td>
</tr>
</tbody>
</table>

With the 3-month Euribor averaging 0.11% p.a. over the third sub-sample period, it was not hard for all of the funds to beat the performance of the risk free rate. This can be observed via several return to risk measures listed in the table (11). This table also contains the benchmark related measures calculated over the aforementioned period.

The highest Sharpe ratio over this most recent period has been achieved by the Axa fund, mainly because of its very low volatility when compared to that of the competing funds. The
minimum, 5th, 50th, 95th percentile and the maximum of the rolling 252-day Sharpe ratio are all highest for the Axa fund as well, which suggests the fund achieved its Sharpe ratio outperformance consistently and not because of the outliers. Although Postova unambiguously recorded the lowest overall Sharpe ratio, its median of the 252-day rolling Sharpe ratio is higher than that of Allianz. This partly confirms that the Postova fund suffered only during the shorter and not the whole time period. The ranking of the funds based on the Sortino ratio is the same as the ranking based on the Sharpe ratio. The same ranking is yielded also when we use the DVR ratio measure. The ranking is similar also for the Calmar ratio, with the only change of ING and VUB switching their (second and third) ranks. The Ulcer performance index is highest for Aegon, because of its lowest sum of the squares of the daily time series of the drawdown (see section (3.3.8) for details). The only fund with the t-statistic significant on the 5% level of significance is the fund managed by Axa with the p-value close to the rejection region (0.045). Skewness has been similar across funds. Kurtosis of the daily returns has been higher than that of the normal distribution while kurtosis of the monthly returns has been lower than that of the normal distribution. The K-ratio and the fractal efficiency are highest for the Axa fund, suggesting the „straightest” price path among funds, and lowest for Postova.

The highest Jensen’s alpha against the synthetic average fund (and also statistically significant with the p-value of 0.0017) has been achieved by Axa. This, together with Postova (which, on the contrary, recorded the most negative alpha) was the only Jensen’s alpha statistically significant at the 5% level. The highest sensitivity to the average fund’s movements, that is the highest beta, belongs to Allianz, in line with its presumably highest risk profile. Allianz also recorded the highest deviations from the benchmark when measured by the tracking error, while the most in line with the average fund (with the lowest tracking error) has been the fund managed by VUB. This is closely related to the information ratio measure (see section (3.5.4)) and puts VUB to the first place when measured by the ratio. The Treynor ratio, however, has been more influenced by the funds’ variation in the beta coefficient (see also section (3.5.5)), which translated into Axa achieving the highest Treynor ratio.

To sum up, some funds do performed notably better than the other in terms of the absolute returns (for example VUB in the whole sample, the second sub-sample and the third sub-sample), low risk (for example Axa in the whole sample and several sub-samples) or return to risk measures (for example VUB in the whole sample, Allianz in the first sub-sample or Axa in the third sub-sample). In spite of this, there is not a clear pattern pointing out to the single fund outperforming the other in all of the measures and all of the sub-periods. It is therefore
necessary for every single investor to thoroughly consider his investment objective and focus on the measures relevant for that investment objective. There cannot be drawn a uniform conclusion out of the historical data and this conclusion depends on the investor’s preferences. Worth mentioning is also an importance of considering the potential, that is future performance of the strategy, which is, however, hard to assess and predict.

4.3.6 Criteria Difficult To Measure

As the criteria difficult to measure (as of the section (3.6)) are hard to be objectively analyzed we devote them only a short paragraph.

Reputation of all of the managers of the pension funds cannot be distinguished clearly. All of the funds have a track record of the same length and none of the managers experienced any managerial error known to the public.

Regarding credit risk we, of course, refer to the credit risk of the securities in which the funds invest, not the credit risk of the pension companies themselves, which, naturally, is hard to analyze. Credit risk has to be analyzed separately for each fund and we recommend a curious reader to study the monthly reports of the funds, public accounting statements or assets and liabilities reports. These may reveal a structure and specific holdings of the portfolios, although with a delay of a few months. In terms of the credit risk an investor should firstly be cautious about the investments in the securities guaranteed by the governments which do have problems with their economy or operate in unstable political regimes and areas. Term deposits made in the banks domiciled in such countries or government bonds issued by such countries are more likely to experience a default. Moreover, in regard to investments in single securities, such as corporate bonds, not only the country of domicile but also a financial and business stability of a single company on its own is of the uttermost importance. Crucial is also a condition of the whole sector in which the company operates.

The same arguments as for the credit risk may be applied to the liquidity risk of assets included in the pension fund portfolios. An investor thus has to study the funds’ reports separately and consider the same risks as mentioned above.

Pricing risk of the pension funds themselves is only very limited, because an investor is able to invest in the funds or switch among funds on a daily basis. On the other hand, the pricing risk of the securities inside the portfolios is again very hard to be measured and is subject to the possible issues mentioned above.

As with presumably any pension fund in the world, all of the Slovak pension funds are exposed
to the market risk of mostly equities and bonds which form the majority of the assets included in the portfolios.

Details regarding investment strategies are in case of the Slovak pension funds impossible to analyze, because no company discloses them. Concentration risk may be analyzed from the above mentioned documents and leverage risk is limited legally by law. Currency risk differs from fund to fund and changes in time and again may be analyzed from funds’ reports.

As of 21.3.2016, costs are identical for all funds and are composed of the 0.3% p.a. management fee and 10% performance fee from the performance achieved above the maximum of the fund calculated from the past 3 years window.

5. Conclusion

The aim of this paper was to describe some of the most useful and common performance measures and to show that each of the measures analyzed may add value to the process of evaluating performance of the investment strategies. We tried to stress throughout the paper that any single performance measure is itself insufficient for the assessment of an investment and should not be analyzed solely without considering other relevant measures. We singled out the common mistakes made by practitioners when evaluating investments or funds, discussed the potential advantages and disadvantages of the listed performance measures and finally applied them to the real data.

We firstly introduced the topic of the evaluation of investment strategies and briefly summarized the relevant literature. We then proceeded to the evaluation of the returns. We explained how a visual inspection may be misleading when performed incorrectly. More specifically, why very long charts may be hard to analyze, why the frequency of returns or prices analyzed does matter significantly and why the correct scaling of the analyzed time series is necessary to avoid false conclusions. We also underlined the importance of incorporating asset-specific data adjustments, such as the correct accumulation of dividends, into the analysis. We attempted to provide simple solutions for the above mentioned issues. We defined the basic return measures such as the simple, annualized or rolling return. We also emphasized that it is always important to analyze the whole distribution of a measure rather than the single point value.

Then we introduced the topic of risk measurement with the main aim to distinguish which risk measures are suited for the different types of risk assessment. We started with the historical volatility, its extensions in form of the the semi-deviation or the downside deviation and
showed why they alone are not sufficient for the assessment of an investment’s risk. We followed with the concept of the maximum drawdown and time to recovery and highlighted their contribution to the risk analysis as well as their potential drawbacks.

The most widely used return to risk measures were covered in the paper in the subsequent sections. We stressed how important is to always analyze the return measures with relation to the risk measures and not each of them separately. We defined the Sharpe ratio, the Sortino ratio and the Calmar ratio and showed how they may add value to the process of the performance measurement. We singled out some common false conclusions to which one may arrive when analyzing only a single of the above mentioned measures. We followed with the introduction of the ratios of the return against percentiles of the drawdown and their role in the evaluation of strategies. We proceeded to the definitions of less frequently employed but useful measures such as the Omega ratio or the Ulcer performance index.

Some of the measures do not fit into the return, the risk or the return to risk category. We discussed those in the subsequent section. We briefly analyzed the ex-post statistical significance of the investment strategies and mentioned why it may not be sufficient with regards to the economical significance of the strategy. Next, we complemented the standard measures of skewness and kurtosis with the less frequently used measures of the fractal efficiency and the K-ratio.

In the next section we focused on the benchmark related measures. These form the main content of the standard performance measurement literature, therefore we analyzed them only briefly. We explained by using the specific examples the concepts of the Jensen’s alpha, beta, the Black-Treynor ratio, the tracking error, the Information ratio and the Treynor ratio. We also emphasized the importance of the correct choice of the benchmark.

In the subsequent section we described some, mostly risk, measures which are hard to quantify but remain important for the evaluation of investment strategies. In there, we provided the reader with examples. An investor has to definitely consider a reputation of a manager, a credit risk involved in an investment, a liquidity risk of assets in a portfolio, a pricing risk regarding the instruments bought and a market risk associated with an investment. Last but not least, the details about the investment strategies such as the degree of concentration, timing, leverage or currency risk and costs are of the uttermost importance.

In the practical section we applied all of the earlier defined measures to the more than 10 year long historical data of the Slovak equity pension funds of the defined contribution, second,
funded pillar. We firstly described how we had prepared the data to avoid any false data manipulation. Next, we performed the visual inspection. We have also divided our sample into three unequal parts corresponding to the different investment policy regimes of the funds. Finally, we calculated all of the measures for each fund in the whole sample as well as in the three aforementioned sub-samples. We analyzed the results and concluded that some funds do perform notably better than the other, be it the whole sample or some sub-sample. Nevertheless, no fund outperforms its competitors in all periods and under all performance criteria. The results naturally vary from measure to measure and from sample to sample. Hence, every investor has to arrive at his own conclusion regarding choice of the fund. Another reason why this is so is that for every investor an investment objective and thus also the relevant performance measures may be different.

Potential future extensions of this paper may involve a more comprehensive list of performance measures, higher focus on an evaluation of strategies against a benchmark or more detailed coverage and implementation of factor models. Prospective work may also elaborate on integration of a fundamental, rather than purely quantitative, approach into an evaluation of the strategies. The same applies for possible integration of a forward looking, i.e. predictive, instead of the backward looking analysis. An extensive and valuable future research may also be conducted in the area of an ex-ante evaluation of investment strategies. That is, in the process of correctly back-testing potential, not yet implemented, strategies.

References


