ORIGINAL SCIENTIFIC PAPER

PERFORMANCE APPRAISAL OF CROATIAN MANDATORY PENSION FUNDS

Pierre Matek, Marko Lukač, Vedrana Repač*

ABSTRACT The goal of this paper is to determine whether managers of Croatian mandatory pension funds have displayed investment skill on a risk-adjusted basis during the 2005-2014 period. We have calculated various risk-adjusted investment performance measures and have then used a number of statistical tools to test the significance of the results. Evidence from our analysis suggests that Croatian mandatory pension funds have reached their investment targets in terms of risk-free rates or benchmarks. Evidence of investment skill was found in some of the funds analysed.

KEYWORDS: *pension funds, risk-adjusted return, performance appraisal*

JEL: G11, G18, G19, G23

1. INTRODUCTION

While performance attribution seeks to identify and quantify the sources of an account's performance relative to its benchmark, the goal of performance appraisal is to provide evidence of investment skill. Feibel (2003, p. 203) defines investment skill as "the ability to deliver value- added above the benchmark over time in a statistically significant manner". Performance appraisal should take into account that investors are risk averse and need additional expected return in order to take on additional risk. Hence, we should adjust portfolio performance for risk. In performance appraisal analysis, we generally measure risk as the account's total risk, measured by its standard deviation, or its market (systematic) risk, measured by its beta.

If the efficient market hypothesis holds true, active investment management cannot add any value for investors (Coggin et al., 1993). Therefore, the only rational choice for a plan sponsor would be to invest in a passively managed market index. As there are many actively managed pension funds, it is not surprising that numerous researchers, particularly in the USA, have analysed the performance of pension funds (e.g. Brinson et

* Pierre MATEK, Croatian Financial Services Supervisory Agency, Miramarska 24b, 10000 Zagreb, Croatia, e-mail: petar-pierre.matek@hanfa.hr
Marko LUKAČ, University College EFFECTUS - College for Finance and Law, Trg J.F.Kennedy 2, 10000 Zagreb, Croatia, e-mail: marko_lukac@live.com
Vedrana REPAČ, University College EFFECTUS - College for Finance and Law, Trg J.F.Kennedy 2, 10000 Zagreb, Croatia, e-mail: vedrana_repac@gmail.com

al. (1986), Ibbotson and Kaplan (2000), Hoernemann et al. (2005)). To summarize, there seems to be little evidence for superior performance of US pension funds relative to passive benchmarks. As regards pension funds in the UK, Blake and Board (2000) concluded that there was "no evidence of superior performance from active fund management that was capable of being sustained over the long life of a pension plan". Likewise, Ammann and Zingg (2009) found no evidence of persistence in the performance of Swiss pension funds and investment foundations. Gallo (2008), on a sample of Italian pension funds, found that "the pension funds do not represent an efficient alternative to the TFR provision" but also found that "a strategy of buying-and-holding on the market portfolio would have represented a more efficient alternative to the TFR."[1]

Less literature is available when emerging or frontier markets are concerned. Antolin (2008) brings a broad overview of pension funds from many countries around the world. The main conclusion from this report is that privately managed pension funds "have obtained a positive premium given the level risk when comparing at least with the short-term alternative investment instrument". In contrast to papers dealing with more mature markets, Stanko (2002) concluded that the gross investment results for the pension fund market in Poland were satisfactory "because the market as a whole and half of the existing sample produced significantly positive results". However, the analysis covered a period of only three years. As far as Croatia is concerned, to our knowledge, only two papers have scrutinized the performance of pension funds. Matek and Radaković (2015) found that the average mandatory pension fund manager in Croatia added 77 basis points annually to the return of a benchmark portfolio over the 2005-2014 period. Novaković (2015) also concluded that Croatian pension funds outperformed their benchmarks, even on a risk-adjusted basis.

Performance appraisal of mandatory pension funds in Croatia is worth undertaking for several reasons. Firstly, research on the topic is an objective contribution to the ongoing public debate on the merits of the pension reform. In our opinion, results of performance analysis should primarily be the basis for strategic decisions about the investment process within fund management companies. Unfortunately, it is often misinterpreted and misused as a "proof" of futility of the pension reform and an argument to return to a one-pillar "pay as you go" government sponsored model. We believe that the performance results of privately managed pension funds are in no case a proof of superiority of the "pay as you go" system. The fact is that this system has been supplemented by privately managed second and third pillar pension funds because of its long-term demographic unsustainability, not because of its poor investment performance (if there is any at all). Secondly, this paper is a logical extension to the work presented in Matek and Radaković (2015). In "Is Active Management of Mandatory Pension Funds in Croatia Creating Value for Second Pillar Fund Members?" Matek and Radaković have evidenced positive results from active management. This paper puts these results in a risk-adjusted context. Finally, publicly scrutinizing performance and portfolios of mandatory pension funds could spur efforts to improve the investment process within the management companies.

[1] The TFR is a government guaranteed scheme based on a quote of monthly salary retained by the employer and linked to inflation.

The paper consists of six chapters. Following the introduction, we give a short overview of the second pillar pension system in Croatia. The next chapter describes the methodological issues and choices we have made in collecting data for the analysis. The fourth chapter presents the results of the calculation of the most common risk-adjusted performance measures. In the fifth chapter, we present tests for statistical significance of the investment results. Finally, in the conclusion we give a summary of our findings regarding evidence on the existence of investment skill in the management of Croatian mandatory pension funds.

2. OVERVIEW OF THE PENSION SYSTEM IN CROATIA

In 2002, a pension system reform in Croatia introduced second and third pillar privately managed mandatory and voluntary pension funds to complement the existing government-sponsored first pillar "pay as you go" system. The second pillar consists of mandatory "defined contribution" pension funds where employers divert a certain percentage of each employee's monthly gross salary into the pension fund. Pension fund members assume the investment risk until retirement. At retirement, they transfer the capital that they have accumulated to a specialized insurance company and become entitled to a life-long annuity.

Today, the mandatory pension funds' market consists of four management companies. Until 2014, each of these companies managed one pension fund and all the participants were assuming the same risk profile. A proxy life-cycle model introduced in 2014 saw the creation of three mandatory pension fund categories with different risk profiles: models A, B and C. Each of the management companies must offer all three models to the system participants. Our analysis covers only investment results of model B funds because they are clearly the successors of the funds created in 2002, in terms of both investment strategy and assets under management. The fund management companies employ their own fund managers and invest funds' assets actively across various asset classes. Although the investment results of the four pension funds vary greatly, the sensitivity of pension fund members to performance is very low. This is evidenced by the extremely low participation of new fund members in the choice of their fund (around 95% of newly employed persons are automatically attributed to a fund management company because they do not choose their fund at employment) and the extremely low rate of switches from one fund to another by existing fund members. Through mandatory pension funds, Croatian citizens have gradually built up individual pension assets invested in capital markets. Mandatory pension funds had 1.7 million members at the end of 2014 and their assets (8.65 billion of euros) represented almost 20% of the GDP of Croatia.

3. METHODOLOGICAL ISSUES AND CHOICES

For our analysis, we have used the data set of monthly gross of fees returns for pension funds constructed in Matek and Radaković (2015). According to Bacon (2008, p. 34) monthly returns are dominantly used as the industry standard. Other authors like Coggin et al. (1993) or Gallo (2008) also use monthly returns. The period analysed covers 10 years from January 2005 to December 2014. We believe that "gross of fees" returns are more appropriate for this analysis for at least two reasons: for the comparability between benchmark and pension fund returns and the comparability between funds themselves across the same or different periods. As an additional argument, the level of fees is set administratively and is beyond the influence of the portfolio managers.

The analysis covers all four mandatory pension funds i.e. 100% of the market from January 2005 to July 2014. After the introduction of three categories of funds in August 2014, we considered funds B to be the successors of the previous funds because 97.3% of the assets remained in funds B and they were the most similar to the previous funds in terms of investment policy. Due to their very short life span, our analysis does not cover the returns of funds A and C. We have given the following codes to the four funds analysed: AZ, EP, PC and RB.

In order to compute risk-adjusted measures of performance, we needed a proxy for the risk-free rate of return and appropriate benchmarks. Many issues arise when determining the risk-free rate for Croatian funds:

- the rating of long-term government debt is below investment grade,
- most domestic government bonds are indexed to the euro/kuna exchange rate,
- from time to time short-term kuna rates undergo periods of extreme volatility,
- the major part of the funds' assets is invested in foreign currency instruments (predominantly euro-denominated) although the NAV of the funds is quoted in kunas,
- Croatian mandatory pension funds' assets under management are several times higher than the total amount of Croatian Treasury bills issued.

After having taken all of this into consideration, we finally opted to use kuna short-term rates as a proxy for the risk-free rate rather than short-term euro rates. We chose to conduct our analysis using two data sets: 3-months Croatian Treasury bill rates at the weekly auction date nearest to the beginning of the month and 1-month Zagreb interbank offered rates (ZIBOR) at the first business day of each month. We have given the code RF1 to the 3-months T-bill rates and the code RF2 to the 1-month ZIBOR rates. Although the investment horizon of pension funds is intrinsically long-term, we consider short-term rates better suited than long-term rates for the risk-free rate because the issues of kuna bonds are scarce and the term premium tends to be quite high. Anyway, the choice of the risk-free rate does not affect the ranking of funds' risk-adjusted performance making this discussion more academic than practical. It is interesting to note that Antolin (2008, p. 12), faced with a similar dilemma decided to use four alternative specifications for the risk-free asset: a short-term local rate, a long-term local rate, a short-term US rate (T-bill) and a long-term US rate (T-bonds).

We have used four different benchmarks in our analysis. The benchmark B1 refers to the pension funds' historical asset allocation between asset classes while the benchmark B2 refers to the regulatory asset allocation as described in Matek and Radaković (2015). The benchmark B3 is the "gross of fees" Mirex[2]. Finally, the benchmark B4 is the "gross of fees" return of the average fund (arithmetic average return of pension funds).

[2] Mirex is the official asset-weighted index of returns of pension funds calculated by HANFA.

Table 1 summarizes key performance and risk indicators used in calculations in Chapter 4.

Table 1

Returns, standard deviations and market share for Croatian mandatory pension funds from 2005 to 2014 (%)

	AZ	EP	PC	RB	B1	B2	B3	B4	RF1	RF2
$\bar{R}^1$	6.23	6.15	4.85	5.87	5.13	5.52	5.89	5.77	3.07	3.77
$\hat{\sigma}^2$	4.40	5.36	4.88	4.44	6.16	5.15	4.48	4.63	0.51	0.96
w^3	40	13	17	30	-	-	-	-	-	-

[1] *annual arithmetic mean return*
[2] *annualized standard deviation*
[3] *average market share*

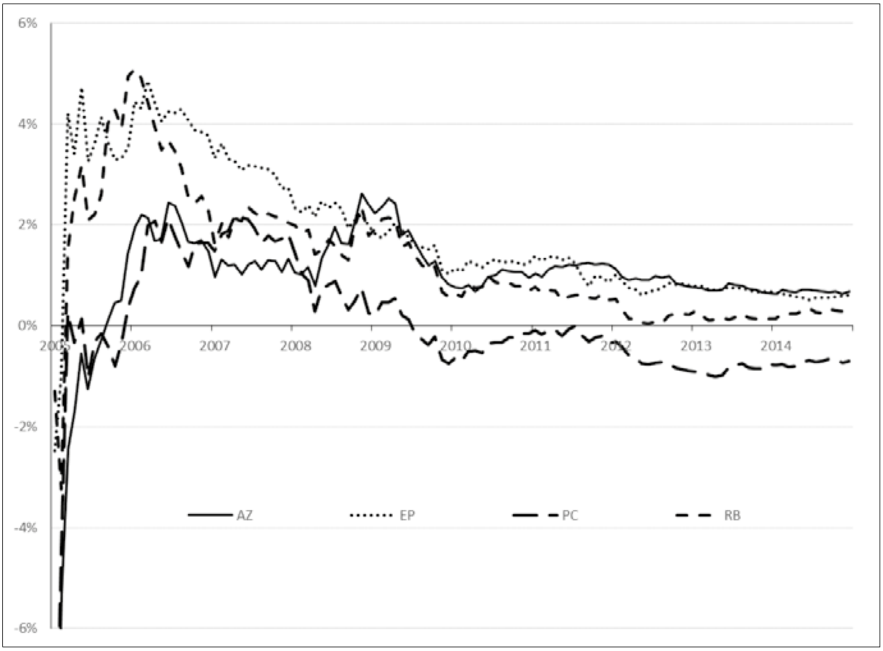
Sources: www.mfin.hr for T-bills returns, Bloomberg for ZIBOR, www.hanfa.hr, Matek and Radaković (2015) and authors' calculations for pension funds' returns

Which benchmark is best suited for performance analysis depends on the performance measure that is used. However, it is clear that none fulfils all of the criteria for a good benchmark: some have not been defined in advance (B1), their structure is not always known in advance (B1, B3, B4), they are not investible (B3, B4) etc. Our perception is that B2 might be the best suited to assess the statistical significance of the investment results of the fund. An underlying assumption of that conclusion is that the regulatory investment limits are in line with the long-term risk-return profile of the pension funds. Figure 1 shows the annualized cumulative value added of the four pension funds compared to B2. We can see that although they are highly correlated, funds have distinctive periods of under or over performance. This is in line with the conclusion from Matek and Radaković (2015) that active management of Croatian pension funds has significantly influenced the performance results.

Another factor affecting the comparability of returns is the use of hold-to-maturity portfolios. Regulation allows Croatian pension funds to use HTM portfolios where assets are valued at amortized cost rather than market prices. This has an obvious impact on both the reported returns and the volatility of the funds. We have not undergone the tedious task of re-pricing portfolios of all the funds and recalculated their returns. Our educated guess would be that the results of such a re-pricing would not have a decisive influence on the results of the analysis. The main rationale behind that argument is that management companies were not using HTM portfolios continuously or on a massive scale. In addition, it is clear from Matek and Radaković (2015) that most of the value-added by active management stems from investments in Croatian equity rather than Croatian government bonds.

Figure 1

Cumulative "gross-of-fees" annualized value-added of Croatian pension funds against benchmark B2 (2005-2014)



Source: author's calculations

4. RISK-ADJUSTED PERFORMANCE MEASURES

We can classify risk-adjusted performance measures into two main categories: measures based on total risk and measures based on systematic or market risk. Total risk is usually measured by the standard deviation of returns, or some downside deviation measure. Market risk is usually measured by beta. Criticisms of risk-adjusted measures described in Feibel (2003) are valid here too:

- it is questionable whether assets are valued according to the CAPM or some other model,
- the use of different market indices as surrogates for the market portfolio can produce significantly different conclusions about performance, and
- ex-post calculations are only estimates of the true parameters, which are often unstable (e.g., the aggressiveness of the portfolio manager may change rapidly).

It is important to bear in mind that Croatian government bonds dominate the portfolios of mandatory pension funds. In practice, this means that due to the exposure to this specific risk, total risk might be significantly underestimated.

However, despite all the shortcomings of risk-adjusted measures, we believe that they give a valuable insight into the quality of the portfolio management process and allow us to rank the funds accordingly.

4.1. Sharpe ratio

In order to compare the risk to return efficiency of two or more portfolios, we could simply divide the average return over a period by the standard deviation of returns. The Sharpe ratio, also known as reward-to-variability ratio, slightly modifies that concept by introducing the idea that we should not be able to earn returns over a risk-free rate without taking on risk.

Traditionally, the formula for the ex-post Sharpe ratio is:

$$S_P = \left(\bar{R}_P - \bar{R}_f\right)/\hat{\sigma}_P \quad (1)$$

where:

S_P is the Sharpe ratio,

$\bar{R}_P$ is the annual arithmetic mean return of the portfolio,

$\bar{R}_f$ is the annual arithmetic mean risk-free return, and

$\hat{\sigma}_P$ is the annualized standard deviation of the portfolio's returns.

The annual arithmetic mean return of the portfolio and risk-free asset are calculated by multiplying the average monthly return by 12.

A manager with a higher Sharpe ratio is producing more average return relative to the risk-free rate per unit of volatility than a manager with a lower Sharpe ratio is. A negative Sharpe ratio implies that the manager has not even been able to earn the risk-free rate. The Sharpe ratio is best used to compare portfolios with similar risk characteristics over the same period. Different time spans and different choices of return frequency and risk-free rates will yield different returns.

According to data presented in Table 1, all four Croatian mandatory pension funds have achieved higher returns than the proxies we have used for the risk-free asset. The 3-months Treasury bills have yielded an annual arithmetic mean return of 3.07% with an annualized standard deviation of returns of 0.51% while the 1-month ZIBOR average annual arithmetic mean return over the same period was 3.77% with an annualized standard deviation of 0.96%. The annual arithmetic mean return of the best yielding fund (AZ) was 6.23% while the worst performing fund (PC) had an annual arithmetic mean return of 4.85%. We can also see that AZ was the best performing fund in terms of both return and risk (measured as the standard deviation of monthly returns). This resulted in the highest Sharpe ratio. The EP fund ranked second in terms of return but came third in terms of Sharpe ratio due to a higher standard deviation of monthly returns.

Table 2

Sharpe ratios and rankings for Croatian mandatory pension funds (2005-2014)

Pension fund	S_P RF1	S_P RF2	Ranking
AZ	0.72	0.56	1
EP	0.57	0.44	3
PC	0.36	0.22	4
RB	0.63	0.47	2

Sources: authors' calculations

In our opinion, from the perspective of a fund member paying mandatory monthly contributions to the fund and for whom it is not allowed to exit the system before retirement, the utility of the standard deviation of monthly returns as a risk measure is questionable. Furthermore, a money-weighted rather than a time-weighted rate of return would probably be a more appropriate basis for any meaningful risk-adjusted performance calculation. On the other hand, from the perspective of the fund manager, all else being equal, a lower standard deviation of periodical returns reflects higher investment skill compared to a higher standard deviation.

4.2. Sortino ratio

The Sortino ratio is a modification of the Sharpe ratio that uses the downside deviation as a denominator and a target return as the hurdle rate in the numerator instead of the risk-free rate.

The formula for the Sortino ratio is:

$$SR_P = (\bar{R}_P - \bar{T}) / \left(\sqrt{\frac{\sum_i^n (R_{Pi} - T_i)^2 \, where \, R_{Pi} < T_i}{N}} * \sqrt{P} \right) \qquad (2)$$

where:

SR_P is the Sortino ratio,

$\bar{T}$ is the annual arithmetic mean target return,

R_{Pi} is the return of the portfolio in period i,

T_i is the target return in period i,

N is the total number of measurement periods, and

P is the number of periods per year.

The Sortino ratio accommodates for two criticisms of the Sharpe ratio. Firstly, we intuitively feel that we should take differently into account volatility created by losses and volatility created by gains. Secondly, standard deviation measures risk relative to the mean return, while investors may consider as risky returns that fall below some periodic minimum acceptable return or target return. The target return can be a fixed percentage or a floating value, e.g. zero, the return of a market benchmark, or the actuarially assumed rate of return.

The above-mentioned criticisms of the Sharpe ratio are particularly relevant when the returns are not normally distributed (when returns are normally distributed, rankings of portfolios based on upside and downside risks will be the same) or if the investor has a target return that is different from the mean return. However, the Sortino ratio is not prone to flaws itself: because downside deviation is calculated, using the actual history of fund returns over the period analysed, it can be very sensitive to the number of return observations and the period selected.

In our analysis, we have used the rate of inflation increased by a margin of 200 basis points per year as the target rate. The rationale behind this choice is that a long-term average return of 200 basis points above the annual rate of inflation was unofficially set as a target at the beginning of the pension reform (Bejaković, 2012). As the input for inflation rates, we have used the month on month consumer price index as published by the Croatian Statistical Bureau. The annual arithmetic mean return for inflation is 2.41% (2.42% annually expressed as the compounded monthly inflation rate) which sets the target rate at the annual arithmetic mean return of 4.35% (4.42% annually expressed as the compounded monthly target rate). As can be seen in Table 3, all the funds have managed to outperform the "unofficial" target return.

As noted by Feibel (2003, p. 200), by ranking portfolios by the Sortino ratio rather than the Sharpe ratio, we might find that portfolios appearing to be the most efficient users of risk are not as appropriate investments for a particular situation as others. However, this is not the case for Croatian pension funds where rankings compared to the Sharpe ratio have remained the same. The only difference is that AZ and RB have switched places in terms of observed risk.

Table 3

Annual excess returns (%), downside deviations (%), Sortino ratios and rankings for Croatian mandatory pension funds (2005-2014)

Pension fund	$\overline{R}_P - \overline{T}$	Annualized downside deviation	SR_P	Ranking
AZ	1.88	3.47	0.54	1
EP	1.80	4.35	0.41	3
PC	0.50	4.15	0.12	4
RB	1.52	3.42	0.44	2

Sources: www.dzs.hr for monthly CPI, www.hanfa.hr and authors' calculations for pension funds' returns

4.3. Information ratio

If we want to compare the risk-adjusted return of a portfolio and its benchmark, we can compute the Sharpe ratio for the benchmark and for the account. A higher Sharpe ratio for the account than for the benchmark indicates superior performance. Another way to approach this problem is to use the Information ratio. The Information ratio is a modification of the Sharpe ratio that incorporates both risk-adjusted returns and a benchmark.

The formula for the Information ratio is:

$$IR_P = (\bar{R}_P - \bar{R}_B)/\hat{\sigma}_{P-B} \tag{3}$$

where:

IR_P is the Information ratio,

$\bar{R}_B$ is the annual arithmetic mean return of the benchmark, and

$\hat{\sigma}_{P-B}$ is the annualized standard deviation of the difference between the returns on the portfolio and the returns on the benchmark.

The Information ratio measures the reward earned by the account manager per incremental unit of risk created by deviating from the benchmark's holdings. The numerator is often referred to as the active return and the denominator is referred to as the active risk or tracking error. As described by Feibel (2003, p. 202) the term Information ratio refers to the idea that the manager would depart from the benchmark only if he had some special information not already priced into the market, which presumably would lead to value-added over the benchmark return. An annualized Information ratio above one, using a long enough series of observations, is commonly interpreted as an indication of skill on behalf of the investment manager. A ratio of 0.5 is considered an adequate measure. A negative Information ratio indicates that a fund underperformed its benchmark.

In our opinion, B1 and B2 are not quite adequate for the calculation of the Information ratio, as these benchmarks were not defined in advance. The Mirex might be more appropriate in this context. Firstly, because it is publicly available. Secondly, because it is used as the benchmark for the activation of the guarantee scheme: according to Croatian regulation, if a type B fund underperforms the three years return of the Mirex for more than 600 basis points at the end of any calendar year, the management company will refund the members for the difference. Although the structure of the Mirex is also not fully available in advance, portfolio managers can approximate it using publicly available data on portfolio composition and are motivated not to deviate too much from the asset weighted portfolio composition of their peers. However, the Mirex being composed of the four funds we are analysing, its use has itself numerous flaws.

As can be seen from Table 1, all funds with the exception of PC have managed to beat benchmarks B1, B2 and B4. PC and RB have underperformed the B3 benchmark, although RB underperformed it for only two basis points. The results of our calculations of the Information ratio are shown in Table 4. None of the portfolio managers recorded an Information ratio above 1. In fact, all the Information ratios were below 0.5. In terms of ranking, EP ranked first when compared with benchmarks B1 and B2. This is due to lower tracking error. AZ ranked first when compared with benchmarks B3 and B4. This is does not come as a surprise because AZ has a dominant market position and therefore influences very significantly the performance of the Mirex index. For the same reason, the two largest funds, AZ and RB recorded the lowest tracking errors in relation to B3.

Table 4

Tracking error (%), Information ratios and rankings for Croatian mandatory pension funds (2005-2014)

Fund	TR*	IR_p	Rn	TR*	IR_p	Rn	TR*	IR_p	Rn	TR*	IR_p	Rn
		B1			B2			B3			B4	
AZ	3.08	0.36	2	2.27	0.31	2	0.95	0.36	1	1.22	0.37	1
EP	2.28	0.45	1	2.01	0.31	1	1.61	0.16	2	1.31	0.29	2
PC	2.67	-0.11	4	2.32	-0,29	4	1.27	-0.82	4	1.04	-0.89	4
RB	3.22	0.23	3	2.48	0.14	3	1.14	-0.01	3	1.26	0.08	3

** annualized tracking error*

Sources: www.hanfa.hr; Matek and Radaković (2015) and authors' calculations for pension funds' and benchmark returns

4.4. M^2

The fourth measure we are considering also uses the standard deviation as a measure of risk. M^2, named after its authors Franco and Leah Modigliani, is the mean incremental return over a market index of a hypothetical portfolio formed by combining the account with borrowing or lending at the risk-free rate in order to match the standard deviation of the market index. In other words, M^2 measures what the account would have earned if it had taken on the same total risk as the market index, where the total risk of the account is rescaled by borrowing (to increase the level of risk) or lending (to decrease the total risk) at the risk-free rate. A skilful manager will generate an M^2 value that exceeds the return on the benchmark.

The formula for M^2 is:

$$M_P^2 = \bar{R}_f + \left(\left(\bar{R}_A - \bar{R}_f\right)/\hat{\sigma}_P\right) * \hat{\sigma}_B \qquad (4)$$

where:

M_P^2 is the M^2 return, and

$\hat{\sigma}_B$ is the annualized standard deviation of the benchmark.

Since we have used two different measures for the risk-free rate and four different benchmarks, we have calculated as much as eight M² returns for each of the four mandatory pension funds.

As we can see in Table 5, the choice of benchmark or risk-free rate did not influence the results in terms of ranking and, as expected, produced the same conclusions as using the Sharpe ratio. However, it did influence the assessment of skill, when we compare M² to the annual return of benchmark portfolios.

Table 5

M2 returns (%) and rankings for Croatian mandatory pension funds (2005-2014)

	Annual return of benchmark portfolios				
	B1	B2	B3	B4	**Ranking**
	5.13	5.52	5.89	5.77	
Fund	**M² using RF1**				
AZ	**7.49**	**6.77**	**6.29**	**6.40**	1
EP	**6.61**	**6.03**	5.64	5.73	3
PC	**5.31**	4.95	4.70	4.76	4
RB	**6.95**	**6.32**	5.89	**5.99**	2
Fund	**M² using RF2**				
AZ	**7.21**	**6.65**	**6.27**	**6.36**	1
EP	**6.50**	**6.06**	5.76	**5.83**	3
PC	5.13	4.91	4.76	4.79	4
RB	**6.68**	**6.20**	5.89	**5.96**	2

Sources: www.mfin.hr for T-bill auction results, Bloomberg for ZIBOR rates, www.hanfa.hr, Matek and Radaković (2015) and authors' calculations for pension funds' and benchmark returns

4.5. Jensen's alpha

Jensen's alpha, also known as ex-post alpha, uses systematic risk (beta) rather than total risk (standard deviation) for the calculation of risk-adjusted performance. The formula for the calculation of the ex-post alpha is derived from an ex-post version of the CAPM relationship:

$$\bar{\alpha}_P = \left(\bar{R}_p - \bar{R}_f\right) - \beta_P * \left(\bar{R}_B - \bar{R}_f\right) \quad (5)$$

where:

$\bar{\alpha}_P$ is the annual average Jensen's alpha, and

β_P is the beta of the portfolio relative to the benchmark.

The beta of the portfolio relative to its benchmark is the covariance of the fund and benchmark excess returns over the risk free rate divided by the variance of the benchmark excess returns.

According to Feibel (2003, p. 197), the Jensen's alpha is the excess return not predicted by the CAPM equation. In other words, Jensen's alpha is the factor that reconciles actual returns to those predicted by the CAPM. Positive alpha is a sign of a portfolio manager's skill. If alpha is greater than zero, the fund had a return higher than expected by the CAPM. A negative alpha indicates that the fund performed worse than predicted given the market risk taken.

Table 6 shows the results of our calculations for Jensen's alpha with RF1 and benchmarks B1 and B2. We are not showing here results using B3 or B4 because we do not find these benchmarks appropriate for Jensen's alpha calculation. This is particularly true for the Mirex index where fund's performance is not only partially calculated against itself, but where as much as 40% (in case of fund AZ) of the index constitutes of one fund's own performance.

Table 6

Jensen's alpha (%) and rankings for Croatian mandatory pension funds (2005-2014)

	B1			B2		
Fund	**beta**	**Annual Jensen's alpha**	**Ranking**	**beta**	**Annual Jensen's alpha**	**Ranking**
AZ	0.64	1.84%	1	0.77	1.26%	1
EP	0.81	1.40%	3	0.96	0.71%	3
PC	0.73	0,28%	4	0.85	-0.32%	4
RB	0.63	1.50%	2	0.77	0.93%	2

Sources: www.mfin.hr for T-bill auction results, www.hanfa.hr, Matek and Radaković (2015) and authors' calculations for pension funds' and benchmark returns

The choice of benchmark used did not affect rankings. Fund AZ exhibits the highest Jensen's alpha while fund PC exhibits the lowest alpha. Jensen's alpha for fund PC is negative when compared to B2.

4.6. Treynor measure

The Treynor measure also relates an account's excess returns to the systematic risk assumed by the account. The calculation of the Treynor measure, also known as reward-to-volatility or excess return to non-diversifiable risk is:

$$T_P = \left(\bar{R}_P - \bar{R}_f\right)/\beta_P \qquad (6)$$

where:

T_P is the Treynor measure.

The results of our calculations of the Treynor measure for B1 and B2 using RF1 are shown in Table 7.

Table 7

Treynor measure (%) and rankings for Croatian mandatory pension funds (2005-2014)

	B1			B2		
Fund	**beta**	**Treynor measure**	**Ranking**	**beta**	**Treynor measure**	**Ranking**
AZ	0.64	4.96%	1	0.77	4.08%	1
EP	0.81	3.78%	3	0.96	3.18%	3
PC	0.73	2.44%	4	0.85	2.07%	4
RB	0.63	4.43%	2	0.77	3.66%	2

Sources: www.mfin.hr for T-bill auction results, www.hanfa.hr, Matek and Radaković (2015) and authors' calculations for pension funds' and benchmark returns

4.7. Summary of results

Jensen's alpha and the Treynor measure will always give the same assessment of the existence of investment skill. However, it is possible that measures of risk-adjusted performance based on systematic risk produce opposite conclusions compared to measures based on total risk. This is most likely to happen if the manager takes on a large amount of non-systematic risk. Increased non-systematic risk will increase the total risk but will leave the systematic risk unchanged. Consequently, this will produce a lower Sharpe ratio and M^2 but will have no impact on the Treynor measure or Jensen's alpha.

In the case of Croatian pension funds, the Sharpe, Sortino, M^2, Jensen's alpha and Treynor measures all gave the same rankings in terms of risk-adjusted performance. The only exception is the Information ratio where results were different from the other measures and changed when we applied different benchmarks: when compared to B1 and B2, fund EP exhibited the best results, but it was ranked second when we applied B3 and B4. Fund RB ranked second with all measures except the Information ratio where it ranked third. This result is very interesting when considering that fund EP had an average annual return much higher than fund RB. Fund PC consistently ranks fourth. Table 8 summarizes the results described above.

Table 8

Rankings based on risk-adjusted performance measures for Croatian mandatory pension funds (2005-2014)

Pension fund	S_P	SR_P	IR_P		M^2	$\bar{\alpha}_P$	T_P
			B1, B2	**B3, B4**			
AZ	1	1	2	1	1	1	1
EP	3	3	1	2	3	3	3
PC	4	4	4	4	4	4	4
RB	2	2	3	3	2	2	2

Sources: authors' calculations

5. ASSESSING THE EXISTENCE OF SKILL

According to Feibel (2003, p. 203) "skill can be defined as the ability to deliver value-added above the benchmark over time in a statistically significant manner". We need to observe performance over many periods to be able to say that there is some evidence of investment skill. It is important to note three important shortcomings of such statistical analysis. Firstly, even when results show evidence of skill, it is possible that the investment results of a particular investment manager are completely due to chance. This is a fact inherent to the investment process itself. Secondly, we need very long series of historical observations to assess performance. As Feibel states (2003, p. 203): "if we consider a year the shortest single period over which we wish to evaluate the performance of a manager, we would need many yearly returns to statistically isolate skill from luck. If we shorten the measurement period to quarters or months, we artificially introduce a short-term horizon into a situation where it is probably the long-term results that matter". Finally, the truth is that we are using past performance to find evidence of past skill and superior past performance is no guarantee of skill in the future.

Nevertheless, despite all these shortcomings, careful interpretation of results can give relevant insight into the quality of the portfolio management process within asset management companies. In this paper, we have used different methods to assess the existence of investment skill on the example of Croatian mandatory pension funds: the t-statistic, quality charts and runs test.

5.1. T-statistic

As explained earlier, subtracting the periodic benchmark returns from the funds' return will give us the value-added. We can quantify whether or not the value-added was significant (or not) by calculating the t-statistic for the value-added. To do this, we set up the null hypothesis that the manager has added no value over the period; the alternative hypothesis is that the manager did add value, and we then use the t-statistic to try to prove the null hypothesis false. The formula for the t-statistic is:

$$t_s = ((R_P - R_B) - 0)/\left(\sigma_{P-B}/\sqrt{N}\right) \tag{7}$$

where:

t_s is the t-statistic,

R_P is the arithmetic mean monthly return of the portfolio,

R_B is the arithmetic mean monthly return of the benchmark, and

σ_{P-B} is the standard deviation of the difference between the monthly returns on the portfolio and the returns on the benchmark.

Table 9 shows the results of our calculations of the t-statistic. With a significance level of 5% and 119 degrees of freedom, we would reject the null hypothesis if the t-statistic had a value lower than -1.9801 or higher than 1.9801. Because it is not, we cannot reject the null hypothesis for B1 and B2 for any fund. When B3 and B4 are considered, for fund

PC, we can reject the null-hypothesis that the manager has added no value over the period i.e. that the results are due to chance. In fact, since the results are negative, and the results are statistically significant, we can say that the manager of fund PC has negative skill when compared to its peer group benchmarks B3 and B4.

Table 9

T-statistic for Croatian mandatory pension funds (2005-2014) at 95% confidence level

Fund	B1	B2	B3	B4
AZ	1.1275	0.9863	1.1353	1.1822
EP	1.4117	0.9900	0.5144	0.9042
PC	-0.3357	-0.9160	**-2.5946**	**-2.8066**
RB	0.7248	0.4473	-0.0470	0.2416

Sources: authors' calculations

There are two problems with the calculation and the interpretation of t-statistics. Firstly, we intuitively understand that we should be interested in the performance of a fund over single periods longer than a month. Secondly, there is a problem with non-stationarity, as the statistics calculated will not remain stable if in the following period portfolio managers dramatically outperform or underperform their benchmarks. In our opinion, the t-statistic as calculated above primarily demonstrates that pension funds have not implemented an investment strategy designed in a way that consistently creates value added over the benchmark return on a monthly basis. Since the investment horizon of a pension fund is typically longer than 10 years this does not come as a surprise. The interpretation of the t-statistic in case of B3 and B4 is not clear-cut either, in particular for B3 where larger funds have a larger weight in B3 and will automatically perform closer to B3.

5.2. Quality Control Charts

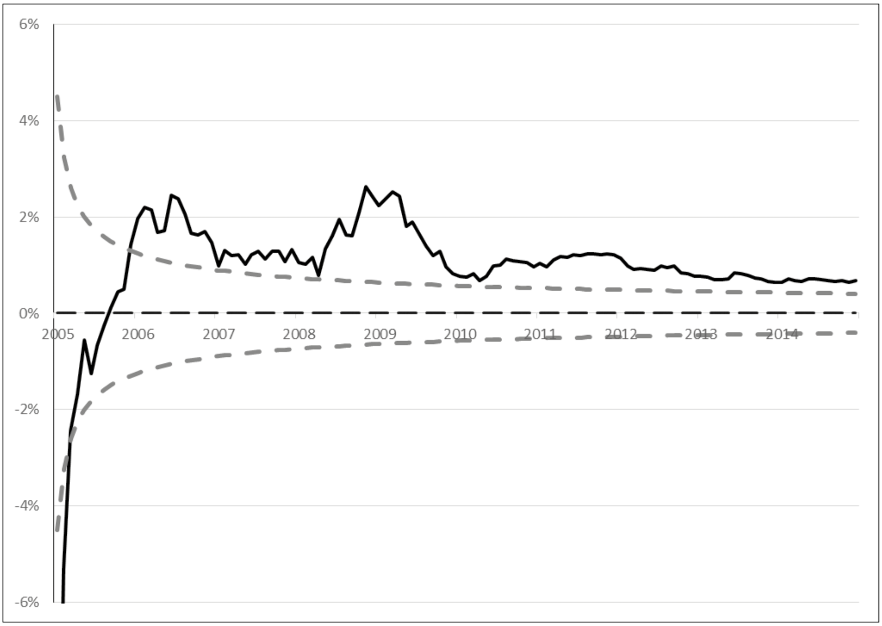
Besides t-statistics calculated on monthly value-added returns data, we can assess the statistical significance of the value-added returns generated by the pension funds' managers using cumulative data. As described in Bailey et al. (2009), one effective means of presenting such data is with quality control charts.

Quality control charts in Figure 2 show the performance of the four Croatian mandatory pension funds versus the B2 benchmark. We have chosen B2 for our analysis because it has recorded the highest total return over the 10-years period analysed. Obviously, we could have conducted the same analysis with any other benchmark. The jagged line is the funds' cumulative annualized performance relative to the benchmark. The funnel-shaped lines surrounding the horizontal lines form the confidence band (the annualized standard deviation of annualized cumulative value-added returns decreases at a rate equal to the square root of time, hence the funnel-shaped lines). We have chosen a 95% confidence band. The null hypothesis of our quality control chart is that the manager has no investment skill; thus, the expected value-added return is zero (represented by the straight horizontal line emanating from the vertical axis at zero). The null hypothesis suggests that ex post differences between benchmark and portfolio returns have no directional biases and are entirely due to random chance.

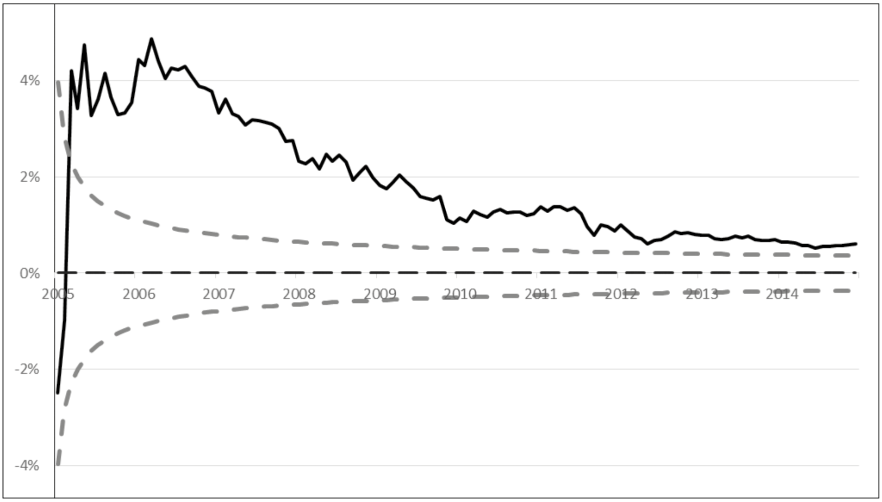
Figure 2

Quality control charts for Croatian mandatory pension funds versus a benchmark based on regulatory limits (2005-2014) with a 95% confidence band

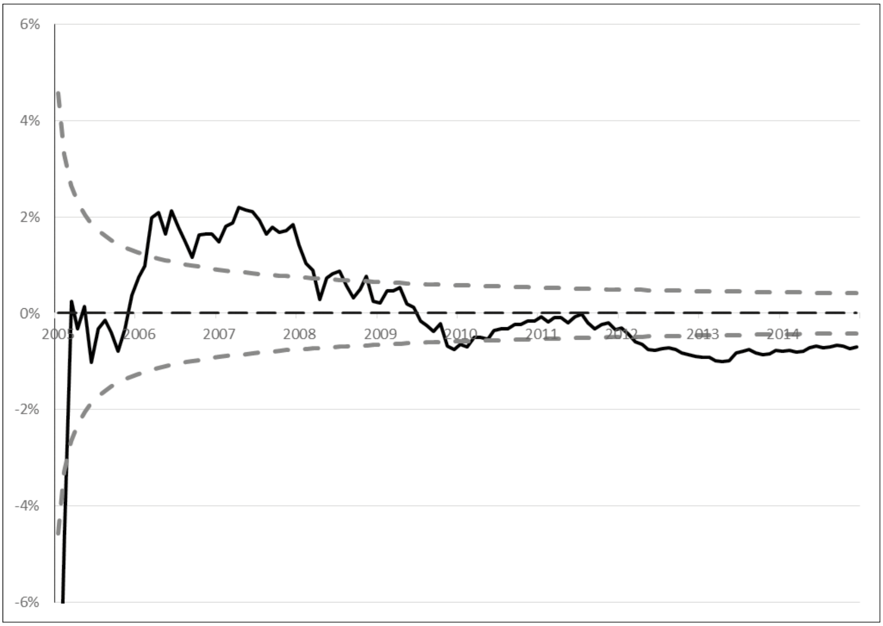
Fund AZ



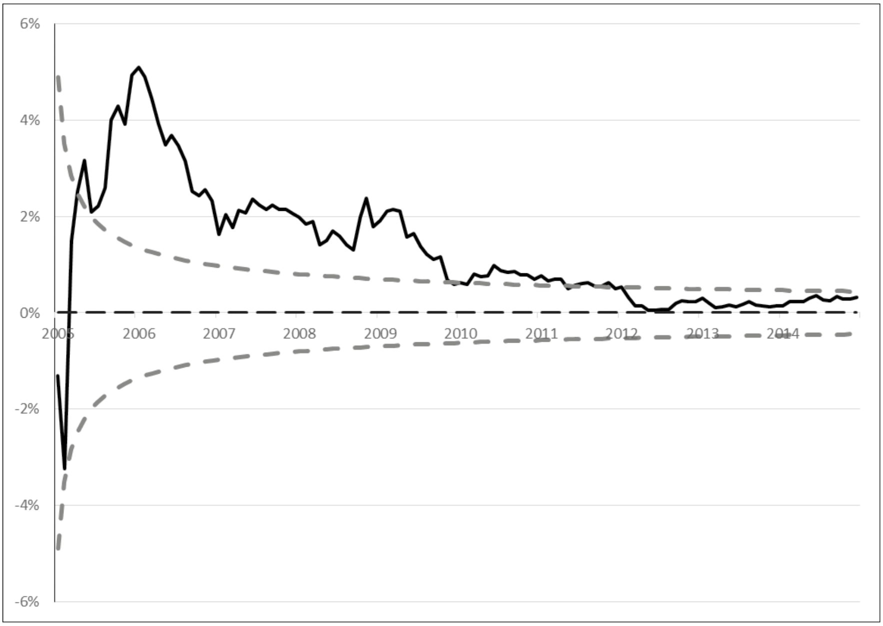
Fund EP



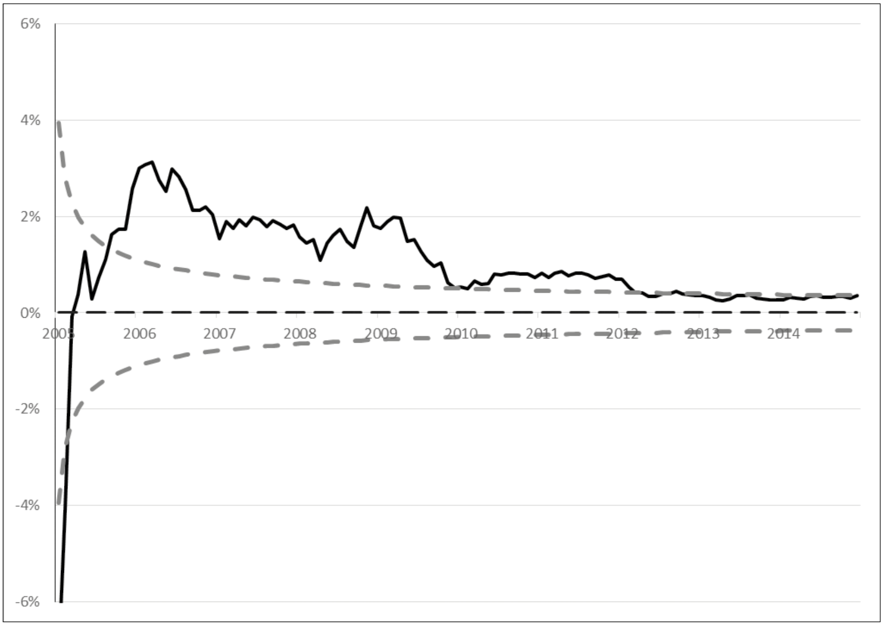
Fund PC



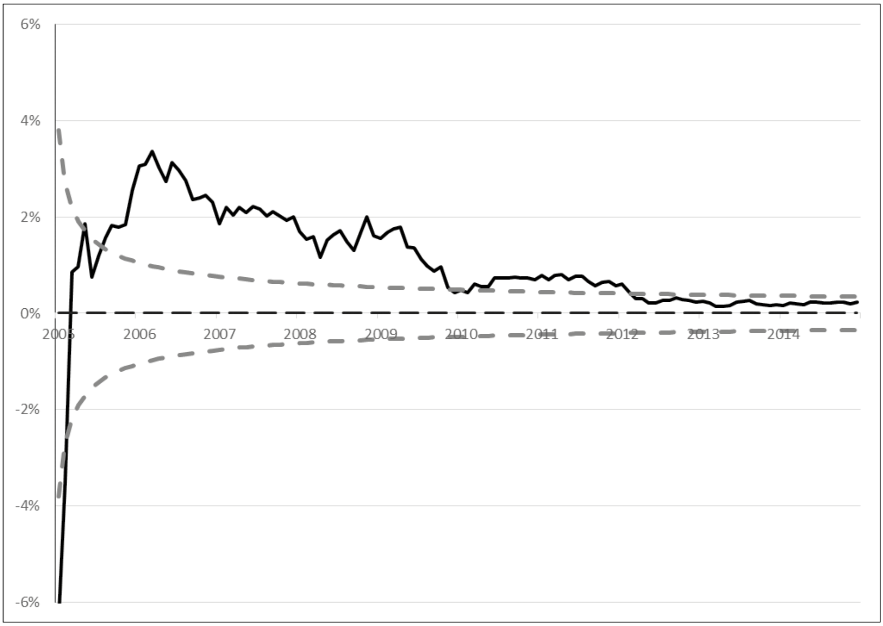
Fund RB



Mirex



Average fund



Besides the null hypothesis, we rely on two other assumptions:

- the manager's value-added returns are independent from period to period and normally distributed around the expected value of zero, and
- the manager's investment process does not change from period to period implying that the variability of the manager's value-added returns remains constant over time.

If such assumptions hold, the confidence band indicates the likely range of value-added return results for a manager who possesses no skill. In other words, based on the properties of a normal distribution, we know that 1.96 standard deviations around the mean will capture ex ante 95% of the possible outcomes associated with a normally distributed random variable and a specified level of value-added return variability.

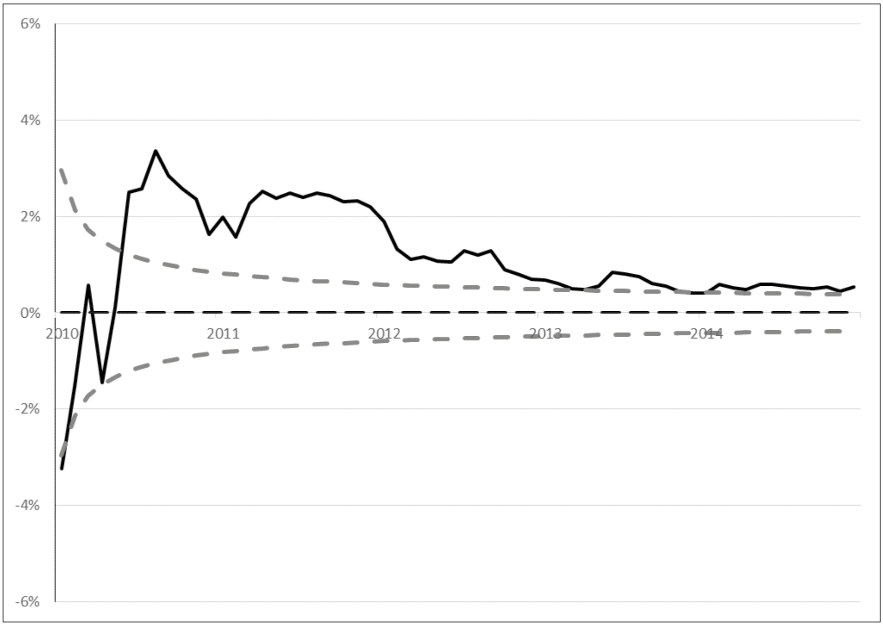
Based on the quality charts in Figure 1, for funds AZ, EP and PC we reject the null hypothesis that the manager's expected value-added return is zero. That is, there is only a 5% chance that a zero-value-added return manager would produce results that lie outside the 95% confidence band. Investment results for funds AZ and EP have breached the confidence band on the upside. The unstated implication is that managers of fund AZ and EP have displayed investment skill during the period of our analysis. Fund PC pierced the confidence interval on the downside. The implication is that the manager is an "underperformer". The investment results of fund RB are positive but within the confidence band. Hence, we cannot reject the null hypothesis that the manager of fund RB has no skill. Even if the manager produced around 41 basis points annually of value-added over the 10 years analysed, as stated in Bailey et al. (2009), "his or her talents are obscured by the variability of his or her short-run results". We have also shown quality control charts for the Mirex (B3) and the arithmetic average fund (B4) - the results are inconclusive.

Because the equity market experienced an extraordinary bull market in 2006 and 2007, followed by an extraordinary bear market in 2008 and 2009, we excluded the first five years and applied the same analysis on the 2010 to 2014 period. The results of this alternative analysis are shown in Figure 3 with very similar results. In the 2010-2014 period, AZ fund breached the confidence band on the upside, funds EP and RB are positive but within the confidence band, and fund PC breached the confidence band on the downside.

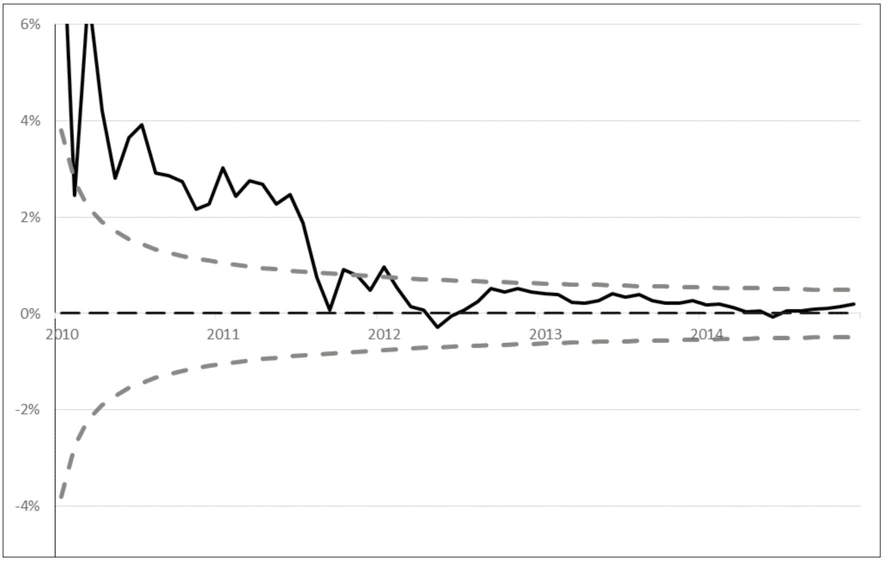
Figure 3

Quality control charts for Croatian mandatory pension funds versus a benchmark based on regulatory limits (2010-2014) with a 95% confidence band

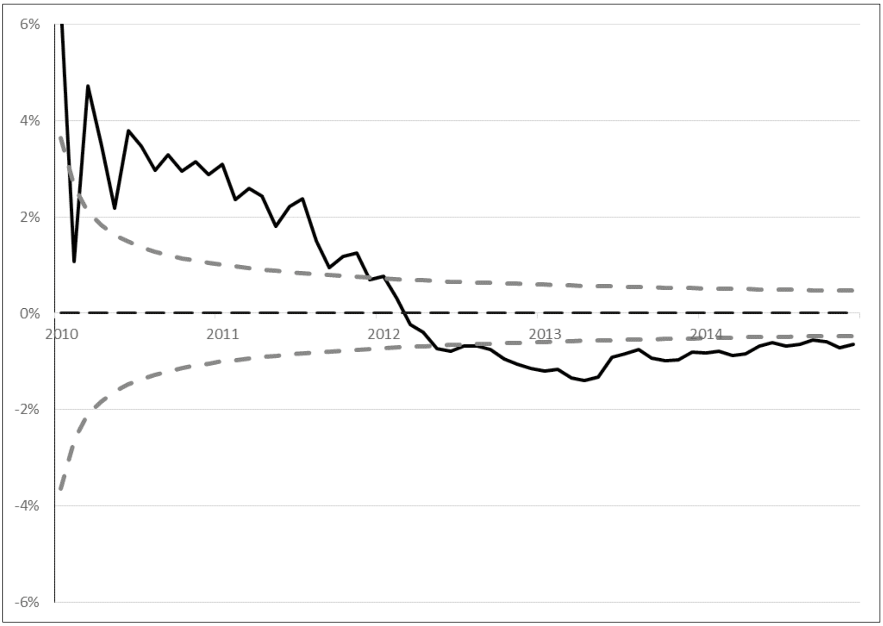
Fund AZ



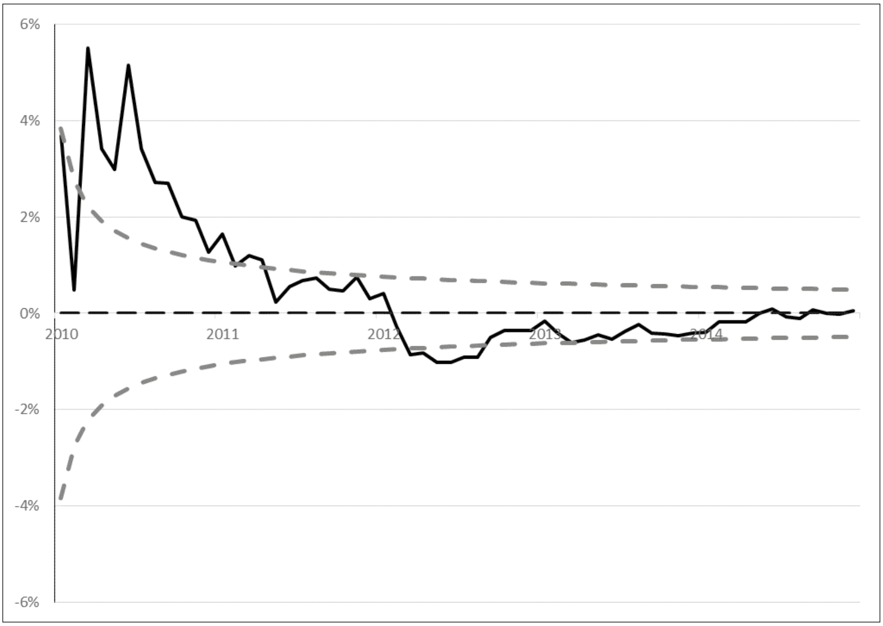
Fund EP



Fund PC



Fund RB



5.3. Runs test

According to Feibel (2003, p. 207), runs tests are used to determine whether the value-added over the benchmark is the result of a random process or there is evidence of consistency. In our calculations, we have first marked each monthly period according to whether there was value-added over the period or not. We have then looked for "runs" in the data. When value-added remains positive or negative from one month to the next, there is a run. When it turns from positive to negative or vice versa, then there is a break in the run. Runs tests are based on the idea that if the pattern of value-added were random, we would expect half of the months to reverse. If there were periods of consistent value-added, we would expect a smaller number of reversals. We can divide the number of runs by the possible number of runs to derive a run ratio. A random process would exhibit a ratio of runs to expected runs of 0.5. A ratio near zero would indicate that there was a pattern of trending value-added, be it positive or negative. Although a ratio near one would indicate that there were no trends, the process was not entirely random either.

Our calculations for all pension funds and for all benchmarks used in this analysis resulted in ratios of runs to expected runs ranging from 0.44 to 0.57. We can conclude that there is no clear pattern of consistency in funds' monthly value-added to the benchmarks we have selected, whether these benchmarks are based on selected asset class proxies or peer group returns. Such results are in line with the t-statistic tests.

One obvious problem with the runs test as a measure of performance consistency is that it accounts for whether or not there was value-added, but not for the size of the value-added.

6. CONCLUSION

Croatian mandatory pension funds have all achieved a rate of return superior to the risk free-rate and have consequently exhibited positive Sharpe ratios. This is in line with the findings from Antolin (2008) and Novaković (2015). They have also achieved positive Sortino ratios, i.e. a rate of return superior to the inflation rate increased by a margin of 2 percentage points per year. This rate was set as an unofficial target at the beginning of the pension reform (Bejaković, 2012). From that point of view, we can conclude that they have achieved their minimum investment targets. When comparing the returns of pension funds with the four benchmarks we have assigned, the results are less conclusive, but in general, we can say that the rates of return were on average superior to the rates of return of the four designated benchmarks. This is in line with the findings from Stanko (2002) but opposed to the findings of other authors, which analysed returns on more mature markets. It is interesting to note that the adjustment for risk did not influence the rankings of the funds. This is mostly because the fund with the highest return also had the lowest standard deviation of monthly returns. The two largest funds displayed the lowest standard deviations of monthly returns, which is in line with our expectations because the funds' size prevented them to achieve a high turnover rate and to take on specific investment bets. The two smaller funds have divergent results: one displayed increased volatility and returns while the other displayed poor investment results combined with a relatively high standard deviation of returns. When assessing the existence of investment skill, the t-statistic based on monthly returns and the runs test lead us to the conclusion that portfolio managers, in general, did not exhibit significant investment skill. However, in our analysis based on cumulative annualized returns, which we find more relevant in the particular case, two or three funds, depending on the period observed, exhibited statistically significant results. In one case, the fund managers exhibited statistically significant investment skill. In another case, the results suggest lack of skill.

REFERENCES

1) Ammann, M. and Zingg, A., 2008. *Investment Performance of Swiss Pension Funds and Investment Foundations*, Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1091165 [Accessed 15 November 2015]
2) Antolin, P., 2008. *Pension Fund Performance*. OECD Working Papers on Insurance and Private Pensions, No. 20, OECD Publishing, doi:10.1787/240401404057
3) Bacon, C. 2008. *Practical Portfolio Performance Measurement and Attribution*. 2nd Ed., John Wiley & Sons, Ltd,.
4) Bailey, J.V., Richards, T.M., Tierney, D.E., 2009. *Evaluating Portfolio Performance*. In: Investment Performance Measurement - Evaluating and Presenting Results, Lawton P., Jankowski T. (eds). John Wiley & Sons, Ltd,.
5) Bejaković, P., 2012. *Obilježja i učinci kapitalizirane obvezne i dobrovoljne mirovinske štednje u Hrvatskoj.* Rev. soc. polit., god 19, no 2, pp 219-230. doi: 10.3935/rsp.v19i2.1089
6) Blake, D. and Bord, J. 2000. *Measuring Value Added in the Pensions Industry*. Discussion Paper PI-0005, The Pensions Institute, Cass Business School. Available from: http://www.pensions-institute.org/workingpapers/wp0005.pdf [Accessed 15 November 2015]
7) Brinson, G.P., Hood, R., Beebower, G.L., 1986. *Determinants of Portfolio Performance.* Financial Analysts Journal, July-August 1986: 39-44
8) Coggin, T.D., Fabozzi, F. J., Rahman, S., 1993. *The Investment Performance of U.S. Equity Pension Fund managers: An Empirical Investigation*. The Journal of Finance, Vol. 48, No.3, July 1993: 1039-1055
9) Dowd, K., 2000. *Adjusting for risk: An improved Sharp ratio,* International Review of Economics and Finance, no 9(2000): 209-222
10) Feibel, B.J. 2003. *Investment Performance Measurement*. John Wiley &Sons, Inc.
11) Gallo, A., 2008. *The performance of pension funds: the case of Italy,* Investment Management and Financial Inovations, vol 5, no 4: 25-32
12) Hoernemann, J. T., Junkans, D. A. and Zarate C. M., 2005. *Strategic Asset Allocation and Other Determinants of Portfolio Returns,* The Journal of Wealth Management, 8(3): 26-38. doi: 10.3905/jwm.2005.598420
13) Ibbotson, R. G. and Kaplan, P. D., 2000. Does Asset Allocation Policy Explain 40, 90, or 100 Percent of Performance? Financial Analysts Journal, 56(1), pp. 26-33. doi: 10.2469/faj.v56.n1.2327
14) Matek, P. P. and Radaković, M., 2015. *Is active management of mandatory pension funds in Croatia creating value for second pillar fund members?,* Financial Theory and Practice, vol. 39, no 3: 245-278. doi: 10.3326/fintp.39.3.1
15) Novaković, D., 2015. *Evaluation of the financial performance of pension funds in Croatia*, Econviews, vol 28, no 1: 199-212
16) Stanko, D., 2002. *Performance Evaluation of Polish Pension Funds* [pdf], Available at: http://www.researchgate.net/publication/2562092_Performance_Evaluation_Of_Polish_Pension_Funds [Accessed 17 November 2015]