Performance Persistence of Dutch Pension Funds

Xiaohong Huang and Ronald Mahieu*

January 15, 2008

Abstract

This paper studies the performance of pension funds with regard to their investment portfolios. We focus on Dutch industry-wide pension plans, which are obliged by law to report their investment performance according to the so-called *z-score*, which is a risk-adjusted performance measure. The benchmarks in the z-score can be chosen by the trustees a priori. Our results show that pension funds as a group cannot beat their self-selected benchmarks. We also find that no persistence exists in performance, reflecting that trustees are incapable of selecting quality asset managers. Cross-sectionally, it turns out that large plans are better able to beat their benchmarks persistently than smaller plans.

1 Introduction

The aggregated market value of Dutch pension fund investment portfolios is enormous. At the end of 2006 the total asset size of Dutch pension funds at year end 2006 was around €691 billion, while the assets from sources other than pension funds and insurance company, managed by collective investment schemes such as mutual fund and hedge fund is only about 117 billion at the same period.[1] Most of these assets are associated to mandatory industry-wide pension funds (€470 billion). The sheer size of the pension fund asset management sector warrants a careful investigation of the performance of their investment portfolios .

This paper provides a cross-sectional and longitudinal description of the investment performance of Dutch mandatory industry-wide pension plans. Beside its relevance to pension

*Huang and Mahieu are at RSM Erasmus University. Mahieu is also affiliated with Netspar. Corresponding author is Huang. E-mail addresses are xhuang@rsm.nl and rmahieu@rsm.nl. We thank Joop Huij, and Hao Jiang for valuable comments. We gratefully acknowledge IQ Info (www.iqinfo.com) for providing us access to their pension fund database. The usual disclaimer applies.

[1]See tables 3.2 and 3.3.1 on the website of the Dutch central bank (DNB): http://www.statistics.dnb.nl/index.cgi?lang=uk&todo=KapMarkt.

plans themselves, the study also has implications to asset management. In the Netherlands mandatory industry-wide pension plan are multi-sponsor pension plans providing defined benefit pension services to all employees of the companies affiliated to a particular industry. Employees of these companies are obliged to participate in these schemes. The mandatory feature of this pension sector necessitates performance evaluation, as only sufficient investment results can justify mandatory participation.

As a pension portfolio consists of various asset classes, the study of investment performance can be performed both on the level of the individual portfolio classes and on the overall pension plan investment portfolio. Previous empirical studies focus on the former or some aggregate in between. Among them are Ippolito & Turner (1987), Lakonishok, Shleifer & Vishny (1992), Tonks (2005), Bauer, Frehen, Lum & Otten (2007) and Busse, Goyal & Wahal (2006), with mixed results. Ippolito & Turner (1987) and Lakonishok et al. (1992) find that equity portfolios underperform their benchmarks, while Tonks (2005) finds outperformance with a UK sample. Busse et al. (2006) extends the analysis to fixed income portfolios and international equity portfolios but still examines performance on the individual portfolio level. At the overall pension portfolio level, Bauer et al. (2007) find close-to-benchmark performance of aggregate equity portfolios but this paper is silent about other asset classes in respect to the overall pension plan portfolio.

No doubt, performance evaluation of asset class portfolios within a pension plan is important as it aids trustees to hire or fire asset managers (Goyal & Wahal (2004)). Yet to the participants, the ultimate beneficiaries of the plan, the investment performance of the plan's entire asset pool matters more, as it will directly influence the premium they have to pay or the indexation they can receive for their pension rights. Typically the portfolio of a pension plan is prudently designed by the pension trustees or by the investment house delegated by the trustees[2]. Studying the overall portfolio performance can help us in identifying trustees'

[2]In some cases the allocation decision is delegated to a fiduciary asset manager who then select other asset managers for particular asset classes. In this case, the paper shows the selection ability of asset managers, but still indirectly reflects trustees' ability of delegating to the right asset manager.

ability of selecting a capable group of asset managers.

The lack of empirical studies on the total pension plan investment performance is related to the fact that there is little detailed information available on asset allocation and returns for individual component of the investment portfolio of pension funds. Fortunately we can exploit a unique regulation for Dutch mandatory industry-wide pension plans and try to fill this blank area in the empirical research on overall pension portfolio performances.

We expect that trustees have a better ability to select asset managers than the average investor. But our study finds that over time an average pension plan cannot earn above-benchmark returns. Furthermore, no significant persistence of any out- or under-performance is found. This reveals that on average pension plan trustees are not able to selecting superior asset managers to outperform their passive benchmarks. Cross-sectionally, however, big plans are able to beat the investment benchmarks than their smaller peers, which implies that trustees of large plans somehow can better select asset managers than the ones of small plans do.

2 Investment performance and the z-score

Before we address performance evaluation, we provide a brief description of the investment process of a defined benefit Dutch pension plan. In general the investment policy is made by the trustees, who consequently delegate the execution of this policy to asset managers. The investment policy is often motivated by an Asset and Liability Management (ALM) study, which is an integral risk management study of the fund, taking into account the long-term and short-term objectives of the fund. Trustees make a decision on the strategic asset allocation, which is a portfolio based on the fund's (subjective) view of expected returns and risks of each asset class and an estimation of the plan's liabilities from a long-term perspective. The strategic asset allocation is often reviewed every 3 or 5 years to reflect major changes in the underlying assumptions regarding the assets and the liabilities. From the strategic asset allocation, trustees define an investment plan that can be implemented by asset managers, often on an annual basis. It reflects a short-term view on the risk-return profile of each asset class and tries to exploit forecasting skills. The plan typically consists of weights that differ from the weights implied from the strategic allocation. According to the annual investment plan, trustees assign mandates for each asset class to a selection of asset managers.[3] Note that these managers can be either in-house or external, one or multiple, passive or active. The main reasons for delegation are mainly related to the expertise of a manager in a particular asset class. Other reasons are economies-of-scale in trading and record keeping (Sharpe (1981)).

The way investments are arranged in the pension fund industry shows that the investment returns are generated from three sources. One is from the long-term strategic allocation using the portfolio weights and returns per asset class from the ALM study. The second source is from executing the annual investment plan, but measured against the benchmark returns. This part of the returns measures the added value from over- and under-weighing the strategic benchmark. The last source results from the actual execution of the investment plan with the actual allocations and the actual returns. More details on return attributions can be found in Good (1984) and **?**. This paper is on the return difference between the second and third sources. The difference tells us how aggregate asset managers perform given an annual investment plan.

To obtain the difference between the actual returns and the returns attainable from a strict adherence to the annual investment plan, we need to know the benchmark portfolio that represents the annual investment plan. This benchmark portfolio is a hypothetical portfolio, which is "structurally identical to the investment strategy without whatever active management takes place" (Logue & Rader (1998) p168) or a "passive mix with the same

[3] Some practice is that decisions on asset allocation among different styles such as value stocks, growth stocks, government bonds and corporate bonds are also delegated to asset managers (*vermogensbeheerders* in Dutch), often called "decentralized asset management" as in Sharpe (1981) and Binsbergen, Brandt & Koijen (2006), etc. Nevertheless this does not influence our performance evaluation procedure. Our results tell us something about the selection ability of the entity who selects the individual asset managers, be it trustees or delegated asset managers.

style" (Sharpe (1992)).

In the Netherlands, the benchmark is sometimes called the "norm portfolio". The norm portfolio has the same investment style/category of the pension plan's annual investment plan and uses the index in each style/category as the return benchmark. The overall return from the norm portfolio represents the return that can be obtained from a passive management of the pension plan portfolio. One example can be found in Table 1. The norm portfolio has a twofold purpose. First, the index for each component portfolio can be used by trustees to evaluate the performance of individual asset managers for a particular asset class. Second, the overall passive return from the norm portfolio can serve as the investment objective and can be used to evaluate whether trustees have selected capable asset managers, either internally or externally. In our study, we use the norm portfolio for the second purpose.

Since 1998 every Dutch mandatory industry-wide pension plans must compute a so-called z-score to reflect their investment performance. It is the difference between the actual return and the return on a predefined norm portfolio, net of expenses, and normalized by the riskiness of the portfolio, as in the following equation:

$$Z_{i,t} = \frac{(R_{p,i,t} - c_{p,i,t}) - (R_{b,i,t} - c_{b,i,t})}{E_{i,t}}$$

where $R_{p,i,t}$ and $c_{p,i,t}$ are the gross investment return and internal investment cost of pension plan i at time t respectively. $R_{b,i,t}$ is the plan i's norm portfolio return using market index in the respective invested asset category at time t. $c_{b,i,t}$ is the associated investment cost of the norm portfolio which depends on the percentage of equity in the portfolio.[4] The norm portfolio is determined by the trustees at the beginning of each year and fixed for one year. $E_{i,t}$ is the risk of the portfolio as a function of the asset mix. The risk percentages for equity and fixed income are fixed by law at 2.6% and 0.6%, respectively[5] For example, if a plan has an asset mix of 60% equity and 40% fixed income, then $E_{i,t}$ is $0.6 * 2.6\% + 0.4 * 0.6\%$. An

[4]These cost are presented in Bpf (2000), and range from 0.10% to 0.22%.

[5]The riskiness of equity and fixe income is fixed across plans and over time. See Bpf (2000)

example of a norm portfolio is presented in Table 5.

The way the z-score is constructed reveals that it is not a measure to evaluate the effectiveness of the investment policy, but a measure of the quality of the implementation of the investment policy. A positive (negative) z-score means that the asset managers selected by trustees can collectively beat (be beaten by) the benchmark set by the trustees. The z-score also reflects trustees' ability to select capable asset managers in general. An unsatisfactory z-score for one year implies that the selected asset managers are not doing a good job to beat the benchmark for that year.[6] If the z-score is persistently low or negative the trustees are incapable of selecting good asset managers. In that case individual companies/sectors may decide to leave the pension plan. A statistical test (*performance test* in Dutch) is used to support this decision. The test statistic is based on the five-year geometric average of z-scores:

$$P_{5\text{ year}} = \frac{\sum_{t=1}^{5} Z_{i,t}}{\sqrt{5}}.$$

The critical value of the test is -1.28, based on a confidence level of 90%. So, if the five-year geometric average of the z-score is less than -1.28, a sponsor can choose to fire the trustees by opting out of this industry-wide pension plan. Subsequently they can either form a new pension plan or they may join another industry-wide pension plan.

There is a tremendous amount of criticism on the z-score. First of all the z-score does not reflect investment performance properly. This is true as investment performance is largely determined by the strategic asset mix and subsequently marginally determined by the operational execution of the strategic investment plan. The z-score can only evaluate the quality of the implementation of this investment policy. This is also our purpose in this paper. We do not indicate which plan has a better investment policy, but we do give an indication which plan can better employ asset managers to reach the a priori investment goals. In this way our study tells whether pension plan trustees do a good job in delegating investments.

[6] An important condition is assumed in practice that there exists good asset managers in the market. Therefore the result is interpreted as trustees' ability of selecting good asset managers.

A second critique is that the norm portfolio is a static benchmark, which is fixed for one year. So any changes in the investment policy cannot be captured by the norm portfolio, but these changes may impact the results on the actual portfolio implementation. This can hamper a fair evaluation of asset managers, because part of the deviations may be from the norm portfolio and are not caused by inferior asset management execution, but by the staleness of the norm portfolio itself. Our paper can not remedy this problem, but in our sample there are plans that use a floating benchmark moving with portfolio development.[7] In addition, allowing for dynamic investment management, which leads to a changing benchmark, is only taking place around 2006 when new regulation on market valuation of liabilities is announced.

A third criticism concerns the risk adjustment in the z-score, where riskiness of equity and fixed income are the same for all plans and for all time. This will also cause an unfair presentation of the plan's ability to beat the benchmark. For example, even though two plans have the same asset mix, one plan may involve higher risk due to investments in higher risk securities within an asset class. So plans with a risk higher than the average is better off by using the fixed risk adjustment.

In this paper we analyze the time series and cross-section of reported z-scores to study the performance of Dutch industry-wide pension funds. We will check the persistence of the z-score to see whether trustees are doing their job of selecting asset managers well. As the z-score is normalized by the riskiness of a plan's asset mix, the comparison of the z-scores among different pension plans allows us to identify plans with better ability of selecting asset managers, and we will link this ability with plan characteristics such as size.

[7]One example is pension fund *Vervoer*, see p.23 in its 2006 annual report, to be obtained from `www.iqinfo.com`.

3 Data

We use the publications of the Dutch industry-wide pension fund association.[8] In addition we obtained data from *pensioninfo* which collects and composes aggregate financial information of companies and organizations including pension funds.[9] We merged and verified data from both sources. When there appeared to be a discrepancy between the z-scores from the two sources, we used the z-score reported in a plan's annual report(s), if available.

Our sample of z-scores runs from 1998 through 2006 and covers the entire history of reported z-scores by Dutch industry-wide pension funds. We include 61 mandatory industry-wide pension funds, either as a pension fund or a pre-pension fund.[10] The sample varies between 57 to 61 over time as some funds become mandatory plans after 1999. Also some funds start after 2000, while some funds merged. Interestingly, our sample covers almost all active mandatory industry funds in the Netherlands.

4 Empirical results

The z-score is based on the plan-specific benchmark, the norm portfolio. As a result the performance analysis in our paper focuses on a a plan's ability to beat its own benchmark. Descriptive statistics in Table 2 show that during the period of 1998 through 2006 the average z-score varies around 0 over time. In the years 2002 and 2004 the average plan underperforms its benchmark reflected by a negative average z-score. In total, an average pension plan does slightly better than its benchmark.

We perform a t-test to examine the statistic significance of the above results. During the buoyant period of 1998 through 2000 and the recovering period of 2005 and 2006, the z-scores are positive at 5% significance level, while in 2002 and 2004 the z-scores are negative at a 5% significance level. When pooled together, the z-score is not significantly different from 0. In

[8] In Dutch it is called the *Vereniging van Bedrijfstakpensioenfondsen* (VB). See their website at `www.vb.nl`.
[9] See their website at `http://www.iqinfo.com`.
[10] A pre-pension fund is a special vehicle that allowed for saving for early retirement.

total we can not reject the hypothesis that industry-wide pension plans as a group over time are not able to beat their own benchmarks.

The test shows that the average plan is performing well under good market conditions, the years 1998-2000, and 2005-2006, but badly under poorer market conditions, the years 2001-2004. This finding seems to indicate that plans tend to take more risk than their benchmark portfolios. As a consequence, in the z-score the systematic risk of the plan portfolio is not fully adjusted by the norm portfolio. Note that we cannot use the z-score directly to say anything of the generated alpha by the funds in our sample. We can only evaluate the performance with respect to the self-selected norm portfolio.

4.1 Cross-sectional performance

The descriptive statistics show that the *average* pension plan is not able to beat its benchmark persistently over time. In this section we will focus on the cross-sectional performance characteristics of the pension funds in our sample. We present the results from three methods.

Following the methodology of Fama & MacBeth (1973), we first regress the future z-score on the past z-score on a yearly basis as in

$$z_{i,t} = a_t + b_t z_{i,t-1} + \epsilon_{i,t},$$

for every year 1999-2006. Using standard OLS we obtain a time series of coefficient estimates. A t-test based on the estimated coefficients, as shown in Panel A in Table 3, gives a positive test statistic of 0.55, which is not significantly different from zero at the 5% significance level. This result (again) implies that as a group pension plan trustee do not exhibit a persistent ability in beating their own investment benchmarks.

To circumvent distributional assumptions on the z-scores, we also apply a Spearman rank correlation test for persistence. In this test we only use the plans with a complete set of z-scores in all 9 years, which results in a sample of 58 plans. In each year we give a rank to each plan based on its z-score. The Spearman rank correlation coefficient for two consecutive periods is then computed as

$$\rho_{t,t-1} = 1 - \frac{6\sum_{t=1}^{N} d_{t,t-1}^2}{N(N^2-1)},$$

where $\sum_{t=1}^{N} d_{t,t-1}^2$ is the sum of squared differences between plans' ranks over two consecutive periods. N is the number of funds (or ranks), i.e. N=58 in our case.

For our 9-year sample, we obtain 8-year time series of correlation coefficients. As in the previous regression test, we apply a t-test using the average and the standard deviation of the time series of correlation coefficients as shown in Panel B in Table 3. The test statistic (0.262) is not significantly different from zero, which is consistent with earlier results.

There might be concerns that the above results may be subject to noise from a few individual plans. Therefore we construct 3 (and 5) portfolios based on their past performance to see how these portfolios perform in the future. Every year, 3 (and 5) portfolios are formed based on the previous year's z-score. For each individual portfolio the average z-score is computed in every year. Repeating this for each year, we obtain a times series of z-scores for 3 (and 5) portfolios. See Table 4. When performance is persistent the table should show that the best-performing portfolios provide the best performance in the subsequent year again. However, the results in the table show that in some years the best performing portfolio from the past year provides the worst performance in this year. When the performance is compared using paired sample t-tests, as presented in Table 5, we find that there is no significant pattern in the performances from one year to the next. In fact, none of the test statistics in Table 5 is statistically different from zero. This again confirms no persistence of plan performance over time.

In order to understand the non-persistence better, we focus further on the composition of the portfolios over time, by applying the methodology of Fama & French (2007). In this analysis we only use the three-portfolio division. Table 6 reports the percentages of number of plans in the current portfolio that originated from the previous year's top, mid and bottom portfolio, respectively. Funds move frequently among these three portfolios. Of the current top portfolio 41% are plans that were also in previous year's top portfolio, but 30% comes from the bottom portfolio of last year. Of the current bottom portfolio only 27% are the plans that were also in the bottom portfolio in the previous year. We test the hypothesis of random migration of plans for the three portfolios. The null hypothesis is that the migration probabilities are all equal to 1/3. The test statistics show that we cannot reject the hypothesis except for one category. This near-random movement of plans among the three portfolios underlines the lack of persistence that we found earlier.

In addition to the information how plans migrate from one portfolio to another over time, we also investigated what the contributions to the current z-score are from the migrating plans. Results are presented in Table 7. In 1999, a large part (-0.39) of the z-score of the bottom portfolio is contributed by the plans that used to be in the top portfolio in the past, while the top portfolio obtains a big chunk of its z-score from the plans in the previous bottom portfolio. Similar patterns can be found in year 2001 and 2005, where the current bottom portfolio's negative z-score is mostly contributed by the plans in the past top portfolio (-0.32 and -0.10 respectively.) In the years 2000 and 2003 the current top portfolio obtains a high contribution to its z-score from the plans that used to be in the past bottom portfolio. Such dramatic changes of performance attribution between years again confirms our previous results that past performance does not tell us much about future performance.

4.2 Performance and plan size

The previous analysis shows that as a group pension plans do not show any out- or under-performance with respect to their benchmarks. It is interesting, however, to investigate the cross-sectional difference among plans. Blake & Timmermann (1997) find that fund size accounts for an important fraction of the cross-sectional variation in equity returns. Following this lead we test whether pension plan size is relevant for explaining performance. Since there is no big change in the relative sizes of the pension plans in our sample, we use the amount of invested assets in 2006 as a proxy for size. We perform a regression of the time-averaged z-score on the plan's size. The sample included for this part is 57 plans after excluding those funds with incomplete z-scores over time and those funds that have merged with other funds. Table 8 shows that size indeed matters. Size explains almost 28% of the variation in a plan's average z-score. The larger plans have a higher average z-score than the smaller plans. This finding implies that larger plans are more capable or better equipped than smaller plans in selecting quality asset managers.

We also computed z-scores on portfolios sorted on size. Results are presented in Table 9. For tertile portfolios, the size effect is not obvious, but in the quintile portfolios we see a clear difference in the z-score between the largest and the smallest plan. In the range of middle-sized portfolio, there seems no clear difference in the z-score.[11] We apply paired sample t-tests for the largest and the smallest portfolio in Table 10. We find that the difference in the z-scores between the largest and the smallest size portfolios is statistically significant. In accordance with the size effect in equity portfolio returns found by Blake & Timmermann (1997), we also find a size effect for pension funds in their ability in beating their self-selected benchmarks. The largest plans can better manage the performance of their investment portfolios than the smallest plans.

5 Discussion and conclusions

Pension plans make decision on asset allocation and decision on selecting asset managers to manage their investments. The first decision determines the majority part of the investment return, and the second decision determines wether selected asset managers can deliver the

[11] Our sample of pension funds contains two very large funds, ABP and PGGM with asset sizes of €208 billion and €81 billion, respectively.

benchmark performance. Our paper focus on the effectiveness of the second decision. Dutch industry-wide pension plans publish z-scores to show their investment performance relative to a priori self-selected benchmarks. These scores reveal asset managers' joint ability to beat the benchmark and thus reflects plan trustees' ability of selecting good asset managers, either internally or externally. After a study of these z-scores on a comprehensive set of industry-wide pension funds in the Netherlands, we find that an average pension plan cannot generate investment return above their benchmarks over time. Using a number of tests we find no performance persistence revealed by the pension plans in our sample. Moreover, worst-performing pension plans in the past seem to make an effort to end in the top performing portfolio in future. We do find that cross-sectionally, large plans can are better able to beat their benchmarks than smaller plans.

References

Bauer, Rob, Rik Frehen, Hubert Lum & Roger Otten (2007), The performance of us pension funds: New insights into the agency costs debate, Technical report, Maastricht University.

Binsbergen, Jules H. Van, Michael W. Brandt & Ralph S.J. Koijen (2006), Optimal decentralized investment management, Technical report, EFA 2006 Zurich Meetings.

Blake, D, Lehmann BN & A Timmermann (1997), Performance measurement using multiple asset class portfolio data: a study of uk pension funds, Technical report, Discussion Paper PI-9704, Birkbeck College.

Bpf, Vrijstellingsbesluit Wet (2000), 'Besluit van 21 december 2000'.

Busse, Jeffrey A., Amit Goyal & Sunil Wahal (2006), 'Performance persistence in institutional investment management', *EFA 2006 Zurich Meetings Paper* .

Fama, E. & J. MacBeth (1973), 'Risk, return, and equilibrium: Empirical tests', *Journal of Political Economy* **81**, 607–636.

Fama, Eugene F. & Kenneth R. French (2007), 'Migration', *Financial Analysts Journal* **63**(3), 48–58.

Good, Walter R. (1984), 'Accountability for pension fund performance', *Financial Analysts Journal* **40**(1), 39–42.

Goyal, Amit. & Sunil Wahal (2004), The selection and termination of investment managers by plan sponsors, Technical report, EFA 2005 Moscow Meetings Paper.

Ippolito, R.A. & J. A. Turner (1987), 'Turnover, fees and pension plan performance', *Financial Analyst Journal* **43**, 16–26.

Lakonishok, J, A Shleifer & R Vishny (1992), 'the structure and performance of the money management industry', *Brookings Papers on Economic activity. Microeconomics* **1992**, 339–391.

Logue, Dennis E. & Jack S. Rader (1998), *Managing pension plans: A comprehensive guide to improving plan performance*, Harvard Business School Press, Boston, MA.

Sharpe, W.F. (1981), 'Decentralized investment management', *Journal of Finance* **36**, 217–234.

Sharpe, W.F. (1992), 'Asset allocation: Management style and performance measurement', *Journal of Portfolio Management* pp. 7–19. winter.

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

Table 1: An example of a norm portfolio

This is a reproduction of a norm portfolio. Source: 2006 annual report of the Agriculture and Food Supply Pension Plan, which can be found via `www.iqinfo.com`.

Assets	Weight	Range	Index
Fixed income	**75%**	**65%-85%**	
Governments	70%	60%-80%	Citigroup Gov Bond Index
Corporates	15%	10%-20%	Citigroup non-EGBI EMU index
Private Loans	15%	10%-20%	Customized Private Loan Index
Equity	**15%**	**5%-25%**	
Europe	40%	3O%-50%	MSCI Europe
USA	20%	1O%-30%	MSCI North America
Pacific	15%	5%-25%	MSCI Pacific
EM Global	25%	15%-35%	MSCI EM Global
Real estate	**5,0%**	**0%-10%**	
Residential	5O%	25%-75%	ROZ- IPD Woningen
Shops	5O%	25%-75%	ROZ- IPD Winkels
Alternatives	**5,0%**	**0%-10%**	
Commodities	50%	0%-100%	DJ-AIG Commoditie Index
Hedge Fund	50%	0%-100%	Euro 7-day Libid

Table 2: Descriptive statistics of z-scores

Descriptive statistics for the z-scores of 61 Dutch industry-wide pension plans over the period of 1998-2006. With a degrees of freedom equal to 60, critical values of 10%, 5%, 1% significance level are 1.67, 2, and 2.39, respectively. Accordingly, (*), (**), and (***) indicate a significance at the 10%, 5% and 1% levels.

	1998	1999	2000	2001	2002	2003	2004	2005	2006	Pooled
Mean	0.26**	0.27*	0.29***	0.08	-0.88***	0.12	-0.37***	0.30 ***	0.30***	0.04
Median	0.14	0.19	0.28	-0.09	-0.98	0.04	-0.36	0.24	0.14	0.02
Maximum	2.25	3.43	3.44	3.84	0.80	1.74	1.34	2.30	2.27	3.84
Minimum	-3.07	-1.22	-1.59	-2.25	-2.91	-1.14	-1.79	-0.87	-0.58	-3.07
Std. Dev.	0.89	0.92	0.81	0.90	0.80	0.54	0.57	0.63	0.57	0.84
Skewness	-0.38	0.84	0.59	0.98	-0.29	0.61	0.08	1.12	1.00	0.22
Kurtosis	5.65	4.20	6.17	6.88	3.36	4.41	3.69	4.64	4.26	5.31
Obs.	59	59	60	60	61	61	61	61	60	542
t-statistic	2.21	2.25	2.75	0.67	-8.63	1.66	-5.13	3.77	4.17	1.04

Table 3: Persistence tests 1 and 2: regression and ranking

Panel A reports the average coefficients from the cross-sectional Fama-MacBeth regression $z_{i,t} = a_t + b_t z_{i,t-1} + \epsilon_{i,t}$. $\bar{\hat{a}}_t$, and $\bar{\hat{b}}_t$ are the time-averaged values of the estimated coefficients $\hat{a}_t$ and $\hat{b}_t$ from the Fama-MacBeth regressions. Panel B reports the Spearman rank correlation coefficient over time, and a t-test on the average coefficients. t-statistics are within brackets.

Panel A: Regression	$\bar{\hat{a}}_t$		$\bar{\hat{b}}_t$		R^2
coefficient	0.0082	(0.052)	0.0584	(0.0549)	0.09

Panel B: Ranking								
Year	98-99	99-00	00-01	01-02	02-03	03-04	04-05	05-06
$\rho_{t,t-1}$	-0.19	-0.28	-0.02	0.22	-0.29	0.15	0.15	0.44
Correlation coefficients	0.262	(0.263)						

Table 4: Persistence test on pension fund portfolios

This table reports the z-score in each year of a portfolio formed on the previous year's z-score. A sample of 58 plans with complete sets of z-scores are used. Panels A and B shows 3- and 5-portfolio divisions, respectively.

Panel A	3 tertile portfolios							
	1999	2000	2001	2002	2003	2004	2005	2006
1(Best past performer)	-0.08	0.18	0.09	-0.78	0.11	-0.30	0.60	0.80
2	0.57	0.21	0.05	-1.02	-0.04	-0.35	0.22	-0.01
3(Worst past performer)	0.32	0.46	-0.09	-0.96	0.31	-0.40	0.14	0.09
Panel B	5 quintile portfolios							
1(best past performer)	-0.10	-0.03	-0.17	-0.63	0.16	-0.11	0.64	0.87
2	0.02	0.31	0.35	-1.01	-0.05	-0.52	0.43	0.50
3	0.59	0.48	0.05	-0.87	-0.03	-0.22	0.10	-0.07
4	0.23	-0.03	-0.03	-1.01	0.11	-0.34	0.19	-0.01
5 (worst past performer)	0.52	0.63	-0.09	-0.99	0.42	-0.53	0.25	0.22

Table 5: Paired sample t-tests on pension fund portfolios

The table reports the paired sample t-test for mean differences of z-scores of various portfolios from Table 4. Panels A and B shows 3- and 5-portfolio division, respectively.

Panel A:	Mean of paired difference	Std. Deviation	t-test	df	Sig. (2-tailed)
Top - Mid	0.12	0.41	0.85	7	0.42
Mid - Bottom	-0.03	0.20	-0.42	7	0.69
Top - Bottom	0.09	0.38	0.70	7	0.51
Panel B:					
Top1 - Top2	0.08	0.36	0.59	7	0.57
Top1 - Mid3	0.08	0.54	0.39	7	0.70
Top1 - Bottom4	0.19	0.38	1.41	7	0.20
Top1 - Bottom5	0.03	0.50	0.14	7	0.89
Top2 - Mid3	0.00	0.38	0.00	7	1.00
Top2 - Bottom4	0.12	0.29	1.14	7	0.29
Top2 - Bottom5	-0.05	0.35	-0.40	7	0.70
Mid3 - Bottom4	0.12	0.23	1.44	7	0.19
Mid3 - Bottom5	-0.05	0.25	-0.56	7	0.59
Bottom4 - Bottom5	-0.17	0.27	-1.76	7	0.12

Table 6: Migration statistics

This table reports performance results from sorting portfolios. Every year portfolio is formed into top, mid and bottom portfolio according to their z-scores in that year. The table decomposes the current portfolio to reflect the fraction of plans that comes from the past top, mid or bottom portfolio respectively. In brackets are the t-statistics testing wether the fraction is equal to 1/3 for a sample of 58 plans (degree of freedom = 7). (*) indicates a significant level of 10%.

	Portfolio based on current performance		
Portfolio based on past performance	Top	mid	Bottom
Top	41%(1.14%)	31%(-0.89%)	31%(-0.41%)
Mid	29%(-1.02%)	28% (-1.36%)	42%(-1.64%)
Bottom	30%(-0.54%)	41% (1.65%)	27% (-1.74%*)
Total	1	1	1

Table 7: Z-score decomposition of portfolios over time

Portfolio is formed based on current year's z-score as in Table 6. The z-score of each current portfolio during year 1999:2006 is decomposed into the z-score contributed by the plans from different past portfolios.

	Decompose portfolio's z-score per year							
	1999	Top	Mid	Bottom	2000	Top	Mid	Bottom
Top		0.23	0.07	-0.39		0.23	0.09	-0.15
Mid		0.51	0.08	-0.05		0.30	0.11	-0.22
Bottom		0.49	0.02	-0.22		0.50	0.07	-0.14
	2001				2002			
Top		0.43	-0.04	-0.32		-0.04	-0.22	-0.52
Mid		0.33	-0.03	-0.27		0.00	-0.39	-0.63
Bottom		0.11	-0.06	-0.15		-0.04	-0.35	-0.58
	2003				2004			
Top		0.24	0.02	-0.16		0.10	-0.14	-0.26
Mid		0.12	0.00	-0.17		0.04	-0.06	-0.32
Bottom		0.30	0.01	-0.02		0.10	-0.14	-0.36
	2005				2006			
Top		0.61	0.06	-0.10		0.71	0.05	0.00
Mid		0.28	0.01	-0.09		0.09	0.02	-0.13
Bottom		0.05	0.17	-0.08		0.11	0.10	-0.12

Table 8: Pension fund performance regressions and size

The dependent variable is the time-averaged z-score for each plan. The independent variable is the logarithm of a plan's invested assets in 2006.

Variable	Coefficient	Std. Error	t-Statistic
Constant	-1.20	0.26	-4.59***
Log(assets)	0.14	0.03	4.76***
R-squared	0.54	Adj. R-squared	0.28

Table 9: z-scores of size portfolios over time

3 and 5 portfolios are formed based on the plan's size in 2006. The tables reports the equally-weighted z-score for each portfolio over time. The sample includes 57 plans that have complete z-score over our sample period 1998-2006 and have not merged with other funds.

Panel A	3 size (tertile) portfolios								
Year	1998	1999	2000	2001	2002	2003	2004	2005	2006
1 (largest plan)	0.31	0.31	0.27	0.09	-0.46	0.08	-0.20	0.64	0.51
2	0.48	0.09	0.48	-0.17	-0.83	0.19	-0.37	0.14	0.25
3(smallest plan)	0.06	0.43	0.09	0.13	-1.41	0.11	-0.45	0.18	0.09
Panel B	5 size (quintile) portfolios								
1(largest)	0.49	0.51	0.13	0.10	-0.42	0.19	-0.11	0.68	0.65
2	0.43	-0.04	0.22	0.02	-0.48	-0.07	-0.38	0.46	0.39
3	0.22	0.22	0.52	-0.22	-0.96	0.20	-0.36	0.08	0.21
4	0.15	0.09	0.55	0.17	-1.43	0.24	-0.62	0.10	-0.04
5(smallest)	0.10	0.60	-0.01	-0.02	-1.31	0.08	-0.26	0.25	0.17

Table 10: Paired sample t-tests of z-scores on size portfolios

The table reports the paired sample t-test for z-score difference between the top portfolio and the bottom portfolio in the respective 3 (tertile) and 5 (quintile) portfolio divisions. Portfolios are formed on size, which is measured by the investment amount in 2006. (*)(**)(***) indicates a significant level at 10%, 5% and 1% respectively.

	Mean of paired difference	Std. dev.	t-test	df	Sig. (2-tailed)
Within 3 portfolio	0.26	0.33	2.34**	8	0.05
Within 5 portfolio	0.29	0.29	3.00***	8	0.02

Figure 1: Pension fund Z-scores: 1998 - 2006

In this figure we report the box plots on sector pension fund z-scores for each of the years in our sample 1998-2006. The boxes around the median line represent the interquartile range. The dotted lines extend to the most extreme data values within 1.5 times the interquartile range. '+'s denote further outlying observations.