

Evaluating UCRP Investment Returns

A Statement by the University of California Academic Senate,
the University Committee on Faculty Welfare, and the
Task Force on Investment and Retirement

March 2009

Abstract

There has been considerable public discussion of the investment performance of the University of California Retirement Plan (UCRP). Much of that discussion has been based on simple comparisons of the realized investment returns of UCRP to those of other pension plans, such as CalPERS. Such comparisons provide no economically meaningful or statistically significant information about the quality of investment management.

The most significant factor in determining the investment return of a portfolio is the allocation of the portfolio among the various asset classes, and the realized returns of those asset classes. The returns of the asset classes in a given time period cannot be predicted in advance, so that portfolios with different asset allocations will exhibit quite different realized returns in different time periods. The effect of the asset allocation should be factored out by comparing the realized return of each asset class within the portfolio to an appropriate benchmark for that asset class. The realized return in each asset class should closely track the class benchmark; this has been the case under the current investment policy adopted by The Regents and implemented by the Office of the Treasurer.

The evaluation of active managers should decompose the realized return into components, in particular the return due to stock selection and the return due to market timing. The evidence indicates that market timing leads to persistent underperformance. It is easy to document persistent underperformance due to stock selection, but the evidence indicates that superior past performance due to stock selection *may* be persistent. The Office of the Treasurer considers the decomposition of investment returns in evaluating active managers, and has replaced managers based on this evidence.

There is no “right” asset allocation for a pension plan; the choice depends on many considerations, so we should expect different pension plans to have different asset allocations, and consequently different returns over different time periods. Although the Academic Senate has not attempted to do a thorough evaluation of UCRP’s asset allocation, the allocation seems appropriate, given UCRP’s characteristics.

1 Introduction

Given the current financial market turmoil, and the prospect of restarting employer and employee contributions to the University of California Retirement Plan (UCRP), people are naturally interested in knowing how well UCRP investments are performing and seeking reassurance that UCRP is well managed.

Many have asked for data comparing UCRP's investment returns to those obtained by other pension plans, such as CalPERS or other state pension plans. However, a simple comparison of realized investment returns provides no economically meaningful or statistically significant information about the quality of investment management, and much of the public discussion based on realized rates of return has been misleading. In particular, the comparison of realized rates of return in portfolios, even over long time periods, is highly sensitive to the specific ending date chosen; it is easy to cherry-pick the ending date to make a given manager look good, or look bad, compared to other managers. In this document, we explain the reasons why realized rates of return do not provide a meaningful basis for evaluating UCRP's performance, and go on to describe the correct way of evaluating UCRP's performance.

2 Statistical Inference from Realized Investment Returns

Even if we are willing to make very strong assumptions, it is very hard to draw any statistically valid inferences on the performance of investment managers simply by looking at their realized rates of return, even over long periods of time. Suppose that we have I portfolio managers $i = 1, \dots, I$. Manager i chooses a portfolio whose *ex ante* expected return is μ_i and whose risk is σ_i . We would like the manager to choose a portfolio with high expected return μ_i and low risk σ_i . There is a tradeoff between risk and expected return, and there will be an upper bound on μ_i that is achievable for any given level of risk σ_i . A portfolio which maximizes μ_i , for a given level of risk σ_i , is called efficient. A good manager will choose an efficient portfolio, while a poor manager will fall short.

Thus, we would like to measure how well the manager has maximized *ex ante* expected return, μ_i , adjusted for risk σ_i . However, we cannot observe μ_i ; we can only observe the *ex post* realized return. Financial returns are very noisy, and the *ex post* realized return is driven primarily by the noise; in other words, the daily *ex ante* expected return is much smaller than the daily volatility or noise. As a consequence, the confidence intervals in estimates of μ_i are huge, even over long periods of time such as ten years. Even though we get roughly 2500 estimates of the manager's daily realized return, our point estimate of *ex ante* expected return depends only on the starting portfolio value and the final portfolio value; we get no additional information from the 2498 other observations of the portfolio value! In practice, if we look only at realized investment returns, we cannot distinguish between a total genius who had a run of bad luck, a competent manager who had average luck, and a total turkey who got lucky.¹ If good returns resulted from luck, there is no reason to think the manager

¹See Appendix A for a formal calculation of the confidence interval on μ_i .

will continue to be lucky in the future. Indeed, simple comparisons of realized returns are highly sensitive to the exact time period chosen; shifting a ten-year period by a year (or even a few months) can easily reverse the ranking of realized investment returns by different managers. It is easy to cherry-pick the ending date of the time period to make a given manager look good or look bad, compared to other managers.

There is a very large empirical literature attempting to measure the extent to which good investment returns are persistent (persistent high returns indicate high skill, while persistent negative returns indicate systematic errors, high transaction costs, or high management fees) or transitory (based on luck). This literature depends on a variety of techniques to tease out additional information, beyond that contained in the realized investment returns. Key ideas include comparing the *ex post* realized return of each asset class in the portfolio to an appropriate benchmark for the asset class, decomposing active returns into stock selection and market timing components, and using factor models. These methods can narrow the confidence intervals and help in differentiating skill from luck; even so, statistical testing remains challenging, and one only occasionally finds superior performance that is statistically significant at conventional levels.

3 Diversification and Idiosyncratic Risk

As noted in Section 2, there is a tradeoff between risk and expected return. There will be an upper bound on expected return μ_i that is achievable for any given level of risk σ_i ; a good manager will achieve this upper bound, while a poor manager will fall short. While some risks are compensated by higher expected returns, other risks are not. Thus, a good manager will take on risk that is well compensated by higher expected returns and avoid risk that is poorly compensated.

The portfolio manager chooses a portfolio of assets. Each individual asset has its own expected return and risk. The expected return of the portfolio is just the weighted average of the expected returns of the individual assets, weighted by the percentage of the portfolio allocated to those assets. However, it is very important to note that the risk of the portfolio is *not* just the weighted average of the riskiness of the individual assets. To the contrary, the risk of the portfolio is determined primarily by the *correlation* among the individual assets in the portfolio. For example, United Airlines and ExxonMobil are both affected by oil prices, but they move in opposite directions in response to changes in the price of oil. A portfolio consisting of both United Airlines and ExxonMobil will respond less to a change in the price of oil than either stock individually. Adding a very risky asset to a portfolio can *lower* the risk of the portfolio.

*Diversification reduces risk, but it has no effect on expected return.*² An undiversified portfolio carries idiosyncratic risk, arising from the idiosyncratic randomness of the performance of individual stocks. This idiosyncratic risk adds to the risk in the portfolio arising from broad risks in the market. *Idiosyncratic risk is not compensated for by higher expected*

²Consider a manager who adopts the following strategy: choose five stocks at random from the Standard & Poor's 500 Index (S&P 500), roughly speaking the largest 500 US domestic stocks, and invest in those stocks in proportion to their weightings in the S&P 500. The expected return of that strategy is equal to the expected return of investing in the S&P 500 index, but its risk is much higher.

return. An index fund is a passively managed fund intended to replicate an index, such as the Standard and Poor's 500 (S&P 500) Index; index funds avoid idiosyncratic risk, incur low transaction costs, and tend to have low management fees.

An active manager chooses a subset of the available assets, based on his/her judgment that those assets will do better than the general market. The manager is accepting additional idiosyncratic risk, which is not compensated for by higher expected return. However, if the manager is sufficiently good at picking stocks, that skill may be enough to raise μ_i sufficiently to compensate for the idiosyncratic risk, as well as the additional transaction costs and management fees inherent in active management. However, for the reasons noted in Section 2, we cannot estimate μ_i well enough simply from realized returns to distinguish between a turkey active manager who got lucky and a genius active manager who got unlucky. In the following sections, we outline additional ways of measuring performance that can be used to provide more economically meaningful and statistically precise measures of portfolio managers.

4 Asset Allocation

A pension plan spreads its investments among a set of asset classes: domestic stocks, domestic corporate bonds, domestic Treasury bonds, foreign stocks, foreign bonds, real estate, hedge funds, commodities, and so on. It is not meaningful to compare the realized performance of different mutual funds or pension funds without taking into account the asset allocation. Two well-managed mutual funds or pension funds with different asset allocations will have significantly different realized returns even over long periods of time such as ten years, and the comparison will be highly sensitive to which particular starting points or ending points are chosen. In other words, shifting a ten-year period by a year (and, in some cases, by a few months) can easily reverse the ranking of realized investment returns for portfolios with different investment allocations. If we are told that three managers earned annual returns of 7%, 8% and 9% respectively over the period 1997-2007, *without information on their asset allocations, we cannot draw any inference about their management skills.* We can say with a high degree of certainty that the manager who earned 9% allocated a larger part of the portfolio to the asset classes which happened to do well during the period.

The effect of the asset allocation can be factored out by comparing the performance of the portfolio's holdings in each asset class to a benchmark: a broad index of all the assets in a given class. For example, the domestic stock portion of a portfolio should be compared to a broad index of domestic stocks, such as the Russell 3000.³ The realized return in each

³The Russell 3000 index consists of all but the smallest US domestic stocks. It includes approximately 98% of the market value of all publicly traded stocks in the U.S. The Regents have chosen a variant, the Russell 3000 Tobacco-Free Index, as the benchmark for domestic stocks. Domestic stocks are often subdivided into smaller asset classes using the Fama-French three-factor model, which makes use of the following factors:

- the overall market, captured by a broad index like the Russell 3000.
- market capitalization, the total value of all common shares of the company issued. Stocks with small market capitalization perform significantly differently than large market capitalization stocks over the course of the business cycle, and over long periods, small market cap stocks seem historically to have had a higher expected rate of return, though there is evidence that this "small stock effect" has

asset class should track the benchmark for that asset class closely. Investment returns for each asset class, and the comparison to the benchmark, are reported to The Regents on a quarterly basis and posted on the Treasurer’s website: <http://www.ucop.edu/treasurer>. For the last five years, the returns in the various asset categories have tracked the benchmarks closely.⁴

5 Decomposing Returns Into Asset Allocation, Stock Selection, and Market Timing

The portfolio’s Asset Allocation Policy specifies the percentage of total assets to be invested in each asset class. It is very helpful in analyzing manager performance to decompose the realized return into four components:

- *Policy*: The return of a passively managed portfolio invested according to the Asset Allocation Policy, and rebalanced as necessary to maintain the asset allocation. This is obtained by computing what the return would have been if each asset class were passively invested in the benchmark index for that class and the fraction of the portfolio invested in each asset class was equal to the long run or strategic asset allocation. This component accounts for the bulk of the variation in the actual portfolio return.
- *Stock Selection*: The return contributed from the selection of individual stocks or other assets within the asset class, as opposed to the benchmark index for the class. This

significantly diminished in recent years. There is less consensus on whether the historically higher expected return was an abnormally high return, or merely appropriate compensation for the risk inherent in small stocks.

- “value,” measured by the ratio of book value of common equity (an accounting measure) to market capitalization; a high ratio denotes “value” and a low ratio denotes “growth;” value and growth stocks perform quite differently over shorter periods, while over long periods, high value stocks seems to have a higher rate of return. There is less consensus on whether this higher expected return is an abnormally high return, or merely appropriate compensation for the risk inherent in distressed stocks, whose low market values lead them to be classified as value stocks.

There is consensus that the Fama-French three factor model has significant explanatory power in individual stock returns. Two well-managed mutual funds or pension funds with different exposures to the three factors will have significantly different realized returns even over long periods of time such as ten years, and the comparison will be highly sensitive to which particular starting points or ending points are chosen. In other words, shifting a ten-year period by a year or two can reverse the ranking of realized investment returns for portfolios with different exposure to the Fama-French factors. Thus, realized returns on portfolios of domestic stocks are often adjusted to reflect the portfolios’ exposure to the Fama-French factors. This provides additional power in measuring manager performance. However, the extent to which superior performance is persistent, after adjusting for the Fama-French factors or other factors, remains controversial. There is consensus that persistent inferior performance is common among active managers, due to transaction costs and management fees.

⁴In 2000, The Regents began phasing in new policies regarding asset allocation and performance measurement, and Treasurer Patricia Small retired. While there has been much public criticism of The Regents’ actions in 2000, the Senate believes the changes in investment policy were appropriate. The returns in the various asset classes now track the benchmarks much more closely than they did prior to 2000, indicating a substantial reduction in idiosyncratic risk through increased diversification.

can be positive or negative in any given period. It is common for an active manager to have a persistently negative stock selection component, if the manager's skill does not outweigh the additional transaction costs and the management fee. There is substantial controversy in the literature about the persistence of positive stock selection returns; in some studies, positive stock selection returns in the recent past are found to predict positive stock selection returns in the future, while other studies fail to find a statistically significant relationship, due to the statistical issues discussed in Section 2.

- *Market Timing*: The return contributed by strategic changes in the weightings on the asset classes. It is tempting to try to predict which asset class will do best in the near future and increase the allocation to that class; for example, one might try to predict recessions, and sell stocks in advance of recessions, then rebuy stocks in advance of the following recovery. However, there is little evidence that those who attempt market timing produce persistent positive returns, and a lot of evidence that attempted market timing produces persistent negative returns.
- *Unallocated*: The difference between the overall realized return and the sum of the other three components. Generally, this component is small.

In evaluating active managers of portions of the UCRP portfolio, it is very useful to carry out this decomposition. Ideally, a manager's market timing component should be close to zero; otherwise, it seems likely that the manager will underperform persistently in the future, even if the market timing component has been positive in the recent past. The stock selection component should ideally be positive in the recent past, as this *may* predict that the stock selection component will be persistently positive in the future; however even a superior manager will exhibit negative realized stock selection returns in some periods.⁵

The Office of the Treasurer considers this decomposition of returns for the outside active managers that are retained to manage portions of the UCRP portfolio. The Treasurer has replaced some outside active managers based on overall performance and the decomposition.

6 Hedge Funds

One asset class, sometimes called the absolute return class, consists of hedge funds. The active asset managers discussed so far typically pick stocks they feel are undervalued, based on fundamental analysis of the underlying strength of various firms; they typically hold stocks for a period of weeks or months, selling them when they feel they are fairly valued

⁵One should look at both the annualized realized stock selection component (denoted α) and also the information ratio IR , defined as α divided by the sample standard deviation of α . α is the best estimate of *economic* significance of the stock selection value added by the manager, while IR is the best measure of the *statistical* significance of the realized stock selection component. One would like to see both α and IR positive. The higher IR is, the more likely a positive result reflects skillful management, which is likely to be persistent, rather than luck, which is unlikely to be persistent. However, it typically takes a very long time to obtain statistical significance at conventional levels. A good manager might obtain an information ratio of 0.5; if so, it would take sixteen years to achieve statistical significance at the 95% confidence level.

or overvalued. By contrast, a hedge fund manager attempts to trade very frequently, potentially hundreds or even thousands of times a day, making a small profit on each trade. Hedge funds initially achieved relatively high expected returns with relatively low risk by using sophisticated statistical models to spot short-term mispricing. As hedge funds have proliferated, it has become harder for them to profit from short-term mispricing. Other strategies, such as proprietary trading and high leverage, are in widespread use by hedge funds. Hedge funds do occasionally fail, sometimes spectacularly; high leverage increases profits in normal times, but increases the risk of a spectacular failure. While most hedge funds were down in 2008, they were on average down significantly less than the domestic and foreign equity markets. There is less consensus on how hedge funds will perform in the future, but based on the evidence to date, it is appropriate to include them as one asset class within the UCRP portfolio. Because hedge funds employ a variety of complicated and opaque strategies, evaluating hedge fund managers is very challenging, and is beyond the scope of this document.

7 Is There a “Right” Asset Allocation Policy?

If we knew the expected return on all the asset classes, and also the correlation of returns among assets classes, we could compute an optimal Asset Allocation Policy that provided the highest expected return, for a given level of risk. Unfortunately, for essentially the same reasons noted in Section 2, we cannot estimate the expected return on an asset class with any precision. Thus, we cannot compute an optimal asset allocation. Moreover, even if we know that one asset allocation outperformed another over a long period, such as ten years, we cannot say that the first asset allocation is “better” than the second; the ranking could easily be reversed in the following ten-year period, or by shifting the ten-year period by a year or two.

As noted above, there is little evidence that market timing produces persistently positive returns, and much evidence that it produces persistently negative returns. Thus, it is preferable to maintain a stable asset allocation, adjusting it slowly and always with an eye to the long term, rather than predictions of short-term performance. This requires periodic rebalancing of the portfolio, selling asset classes that have done well and as a result exceed the target allocation for the class.

The appropriate choice of Asset Allocation Policy for a pension plan like UCRP is complex, and is based on a variety of strategic factors much more than on analyzing past performance of alternative allocations. The strategic choice of an Asset Allocation Policy is beyond the scope of this document. However, to illustrate reasons why different pension plans would choose different Asset Allocation Policies, we note that the following factors would be among the considerations:

- When will the benefit payments need to be made? A plan with a rapidly growing and relatively young workforce, but few retirees, should probably, other things being equal, have a higher allocation in stocks and less in bonds than a plan with a shrinking and relatively old workforce and lots of retirees.
- How well is the plan funded? If the plan is significantly more than 100% funded, one

can argue that it should have a less risky portfolio, other things being equal, than a plan which is significantly less than 100% funded.

- How able is the employer to make up a funding shortfall in the event that investment returns are insufficient to pay the benefits?
- How do the assets correlate with the uncertainty in the benefit obligations?
 - How should the portfolio be allocated between domestic and foreign assets? The inclusion of foreign assets provides diversification benefits, but adds currency exchange rate risk.
 - What inflation-protection does the plan provide to retirees? A plan which provides more inflation-protection might choose to balance that exposure by having a larger allocation to Treasury Inflation-Protected Securities (TIPS).

Different pension plans will have different answers to these questions, and thus *should* have different Asset Allocation Policies. It follows that they will have quite different realized returns, from year to year and even over long periods of time.

On the whole, the asset allocation seems appropriate given UCRP's circumstances, and appears to have served UCRP well; however, the Academic Senate has not attempted to do a thorough analysis of the UCRP asset allocation.

8 Conclusion

We summarize this document by highlighting the main conclusions in bullets:

- Because financial return data is very noisy, simple comparison of realized returns is insufficient to measure performance, even over a long period of time. Comparisons over a long period can be reversed by a small shift in the end date of the period.
- Lack of diversification increases idiosyncratic risk and is generally not compensated for by increased expected returns. Superior stock selection *may* compensate for the increased idiosyncratic risk it produces.
- Returns of two different portfolios cannot be directly compared if they have different asset allocations. The effect of different asset allocations can be removed by comparing performance in each allocation class to a benchmark for that class.
- To analyze manager performance it is useful to look at three main separate components: (a) the return to the underlying asset allocation policy, which accounts for the majority of variation among managers; (b) the return to stock selection; (c) the return to market timing, which should be close to zero.
- Asset allocation policies for any one fund are determined by a complex set of goals and characteristics of that particular fund. There is no right asset allocation that all pension funds should strive to achieve. Hence, different pension funds will have different asset allocations, and consequently will perform quite differently over specific time intervals.

Appendix A: Confidence Interval on μ_i

In this Appendix, we compute the confidence interval on μ_i . Suppose the value of i 's portfolio at date t is p_{it} ; define $r_{it} = \log \frac{p_{it}}{p_{it-1}}$, so we have

$$p_{iT} = p_{i0} e^{\sum_{t=1}^T r_{it}}$$

Here, \log denotes the natural logarithm. The formula assumes no money is flowing into or out of the portfolio; if there are inflows or outflows, the definition of r_{it} needs to be adjusted to take these into account.

We make the following strong assumption: For each i , r_{0i}, \dots, r_{Ti} are independent identically distributed Gaussian with mean μ_i and standard deviation σ_i . The assumption is quite strong. In practice, this assumption fails in several respects:

- Stock returns have fat tails, compared to the Gaussian distribution; thus, the probability of a very large price movement, while small, is considerably higher than one would estimate from typical day-to-day returns.
- Volatility of stock returns, and hence of portfolio returns, varies greatly over time. In particular, high volatility tends to be quite persistent.
- Individual portfolio managers adjust their portfolios, so μ_i should really be μ_{it} and σ_i should really be σ_{it} ; they are moving targets.

Relaxing this assumption makes the problem of comparing managers harder, not easier.

Even under this strong assumption, μ_1, \dots, μ_I *cannot be estimated with any precision, even over very long data periods*. For example, suppose we have access to daily data over a period of ten years. Since there are roughly 250 trading days in a year, this gives us roughly 2500 observations of each manager. Our point estimate of σ_i is the sample standard deviation $\bar{\sigma}_i$; it estimates σ_i quite precisely. Our point estimate of μ_i is

$$\bar{\mu}_i = \frac{\sum_{t=1}^T r_{it}}{T} = \frac{\log \left(\frac{p_{iT}}{p_{i0}} \right)}{T}$$

Note that our point estimate depends only on the starting portfolio value p_{i0} and the final portfolio value p_{iT} ; we get no additional information from the 2498 other observations of the portfolio value! The standard error will be $\frac{\bar{\sigma}_i}{\sqrt{2500}} = .02\bar{\sigma}_i$. The 95% confidence interval is approximately $(\bar{\mu}_i - .04\sigma_i, \bar{\mu}_i + .04\sigma_i)$.

Now suppose we have $\mu_i = \frac{\log 1.075}{250} \simeq .00029$ (i.e. the expected annual rate of return is 7.5%) and $\sigma_i = 0.007$ (in normal times, the daily standard deviation of the S&P 500 is about 1%, so $\sigma_i = 0.007$ is roughly the daily standard deviation of a portfolio invested 70% in a broad index of domestic stocks like the Standard and Poor's 500 and 30% in short-term Treasury securities in normal times; recently, the daily standard deviation of the S&P 500 has been much higher.). If we get really lucky and our point estimate $\bar{\mu}_i$ hits dead on, our confidence interval on the daily rate of return μ_i is $(.00029-.00028, .00029+.00028) = (.00001, .00057)$; expressed as an annual rate of return, the confidence interval is (.25%, 15.3%)! If we observe a realized 7.5% rate of return over ten years, we can't tell whether

it was produced by a total turkey ($\mu_i = 0.00001$) who got lucky, a competent manager ($\mu_i = 0.00029$) who had average luck, or a total genius ($\mu_i = .00057$) who had a run of bad luck.

References

- [1] David Blake, Bruce N. Lehmann and Allan Timmermann, "Asset Allocation Dynamics and Pension Fund Performance," *Journal of Business* 72(1999), 429-461.
- [2] Gary P. Brinson, L. Randolph Hood, and Gilbert L. Beebower, "Determinants of Portfolio Performance," *Financial Analysts Journal* July/August 1986, 39-44. Reprinted in *Financial Analysts Journal* January/February 1995, 133-138.
- [3] Gary P. Brinson, Brian D. Singer and Gilbert L. Beebower, "Determinants of Portfolio Performance II: An Update," *Financial Analysts Journal* May/June 1991, 40-48.
- [4] Mark M. Carhart, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52(1997), 57-82.
- [5] Richard C. Grinold and Ronald N. Kahn, "Active Portfolio Management: A Quantitative Approach for Providing Superior Returns and Controlling Risk," Second Edition. New York: Mc-Graw Hill, 2000.
- [6] Roger G. Ibbotson and Paul D. Kaplan, "Does Asset Allocation Policy Explain 40%, 90%, or 100% of Performance?" *Financial Analysts Journal* January 2000, 26-33.
- [7] Harry Mamaysky, Matthew Spiegel and Hong Zhang, "Improved Forecasting of Mutual Fund Alphas and Betas," *Review of Finance* 2007, 1-42.