Modeling Defined-Benefit Pension Plans: Basic Metrics

By Robert T. McCrory, FSA
EFI Actuaries
1532 East McGraw Street
Seattle, WA 98112

Abstract

The purpose of this paper is to explore pension plan metrics—measuring the behavior of a defined-benefit pension plan in an uncertain environment. Our goal is to measure and evaluate the quantitative impact of competing policy choices. We proceed by investigating some basic pension metrics, evaluating the characteristics of each using a simulated pension plan. We find that some typical metrics—such as the mean and standard deviation—are useful only in certain circumstances and that other less traditional measurements are sometimes better suited to evaluating policy choices.

Keywords

Pension plans, actuarial funding, stochastic system dynamics, risk analysis, risk metrics, policy analysis, optimization

1. Introduction

Defined-benefit pension plans[1] are sponsored by employers to provide retirement security to their employees. The administrator of one of these plans posed a simple question: "What do we manage to?" The purpose of this paper is to answer that question by presenting a set of potential measurements, or metrics, of pension plan performance. We will apply these metrics to some sample policy issues to see if they represent a useful framework for evaluating the choice among different management policies.

2. Methodology

2.1 The Model Plan and Model Economy

In this paper we take an empirical approach: We will report on the results of a number of simulation experiments conducted on a Model Plan in a Model Economy. Both the Model Plan and the Model Economy will be highly simplified; nonetheless, they are representative of actual pension plans in operation, and we can learn a great deal from these simple models about the dynamics of actual pension plans in a real economy. The Model Plan and its dynamic behavior in the Model Economy were reviewed in a prior paper.[2]

2.2 Actuarial Funding

While complicated in detail, actuarial funding of a pension plan is conceptually simple, involving three parts. First, based on assumptions regarding member behavior—probabilities of retirement, death and so on—and on assumptions regarding economic behavior—rates of interest, investment return, inflation and so on—a target level of assets is established. This asset target is called the actuarial accrued liability. Second, an annual contribution, called the normal cost, is computed that is sufficient to pay benefits and maintain the funding target. Third, since all of the assumptions will not be met, assets will differ each year from the asset target. Therefore, an additional payment—which could be negative if assets exceed the target—is added to the normal cost to cause the actual asset level to converge over time to the asset target. This third part of funding is called the amortization of the unfunded—or overfunded—actuarial accrued liability. The employer contribution to the plan is the sum of the normal cost and the amortization of the unfunded actuarial accrued liability.

2.3 Cost Smoothing

Most of the variation in the annual contribution to the plan results from the uncertainties in the investment markets. Actuaries have developed techniques to smooth the impact of investment market fluctuations on employer costs. The method most frequently used spreads unexpected investment gains and losses over a number of years. A

---

[1] Allen, E., J. Melone, J. Rosenbloom and D. Mahoney. 2002. *Pension Planning*. New York: McGraw-Hill/Irwin.

[2] McCrory, R. 2012. "Modeling Defined Benefit Pension Plans: Basic Dynamics," in *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*, Orlando, Fla. (submitted).

smoothed or actuarial value of plan assets is computed, and it is this smoothed asset value rather than the market value of assets that is used to compute the amortization of the unfunded actuarial accrued liability. These methods are described in various actuarial publications.[3]

2.4 Metrics

In this paper we will discuss a number of possible measurements of the behavior of the Model Plan. Our goal in these measurements is to compare the measured performance of the Model Plan under alternative policies. The competing policies may concern plan design, investment strategy, or the actuarial funding method. In doing so, we desire metrics that are intuitively meaningful and that discriminate among proposed policies; a metric with roughly the same value for all proposed approaches is of little value.

We will use a simple test policy to illustrate the behavior of our metrics. One of the features frequently found in a cost smoothing method is a corridor around market value constraining how far the smoothed asset value may deviate from market value. A typical corridor value might be 20 percent, forcing the smoothed value of assets to differ from market value by no more than 20 percent of market value. As a simple test policy, we will measure the behavior of a cost smoothing method with a smoothing period of five years and differing corridor widths. In particular, we will be interested in which metrics seem to be useful in selecting an optimal corridor width.

3. Cost Metrics

We begin with measurements of the plan's actuarial cost; any proposed change in management policy will be assessed in terms of its impact on plan cost. A policy may affect both the level and variability of plan cost.

3.1 Level of Cost

Figure 1 below shows the mean simulated actuarial cost by year for each of three corridor widths: A 0 percent corridor, which makes smoothed value of assets equal market value, removing smoothing altogether, a 20 percent corridor, and a 50 percent corridor. Cost is expressed as a fraction of active member payroll. Cost is simulated for 100 years using 1,000 trials for each smoothing policy option. In Figure 1 we note that in each case the mean cost decreases over time, but increases as the smoothing policy changes and the asset smoothing corridor widens. Moreover, we observe that the change from a 0 percent (market value) to a 20 percent corridor is greater than from 20 percent to 50 percent.

We usually think of actuarial smoothing of assets as a tool for preventing sharp increases in cost, but the method also prevents sharp decreases. Since the average cost has a downward trend due to the Exclusive Benefit Rule,[4,5] more asset smoothing tends—on average—to slow the decrease in cost and keep contributions a bit higher, for a bit longer. This means higher average costs. Nonetheless, we note that the difference in the mean cost is quite small: At 50 years the relative difference between the mean cost with no smoothing and with a 50 percent corridor is just 3 percent. Similar results are obtained when we look at the median cost.

We can evaluate multiple smoothing policies by plotting the average mean and median cost for years 20 through 100 versus the width of the asset smoothing corridor. This is shown below in Figure 2. As the smoothing corridor widens, both the mean and median cost for years 20 through 100 increase, though not by much. Again, there is only a 3 percent relative increase in the mean cost from a 0 percent corridor (market) to a 50 percent corridor. The relative increase in the median cost is higher, about 11 percent, due to the lower values of the median cost relative to the mean.

The small variation of mean and median actuarial cost among smoothing policies suggests that these cost-level

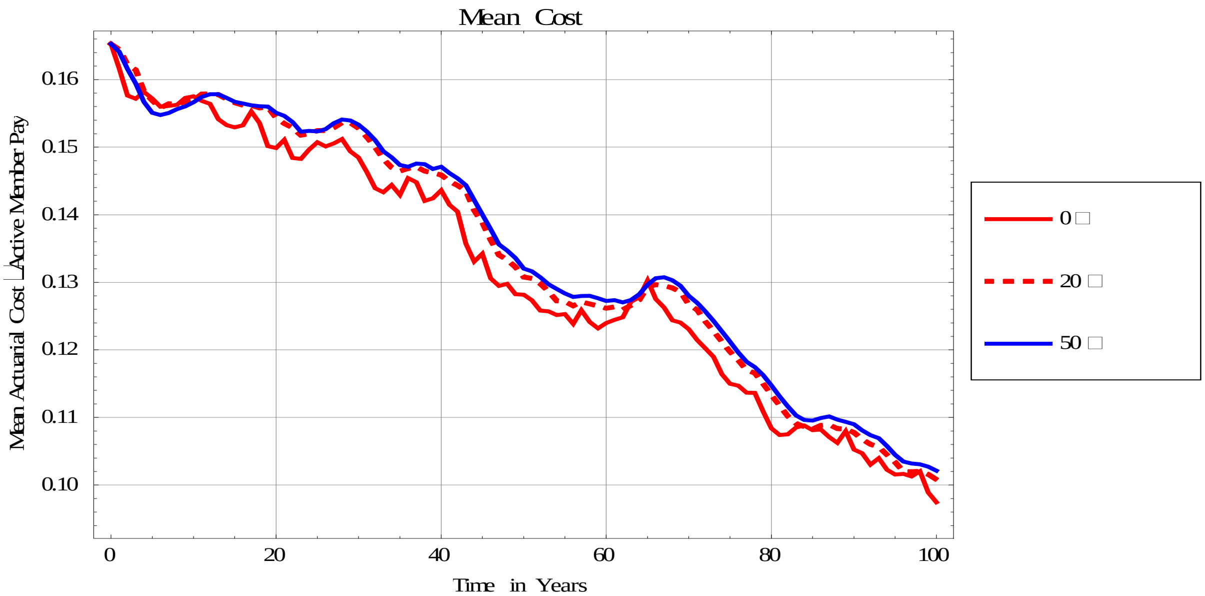
---

[3] Actuarial Standards Board. 2011. *Actuarial Standard of Practice 44: Selection and Use of Asset Valuation Methods*. Washington, D.C.: American Academy of Actuaries.

[4] McCrory, R. 2012. "Modeling Defined Benefit Pension Plans: Basic Dynamics," in *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*, Orlando, Fla. (submitted).

[5] Internal Revenue Code, *Section 401(a).*

metrics are poor policy discriminators for assessing asset smoothing approaches. However, in other cases, mean cost can be vitally important: For example, in any discussion on investment policy, mean cost will vary significantly by investment policy and will be an important measure.[6]



Year	0	25	50	75	100
0% Corridor	0.165275	0.150708	0.128161	0.114998	0.0974617
20% Corridor	0.165275	0.152456	0.130776	0.119707	0.100806
50% Corridor	0.165275	0.152329	0.132039	0.121233	0.102116

Figure 1: Mean Cost as a Fraction of Member Payroll by Year for the Model Plan with Three Smoothing Corridors (Each line represents the mean cost in each year of a 100-year simulation of 1,000 trial for each of three different smoothing policy corridors, with years on the horizontal axis and cost on the vertical axis.)

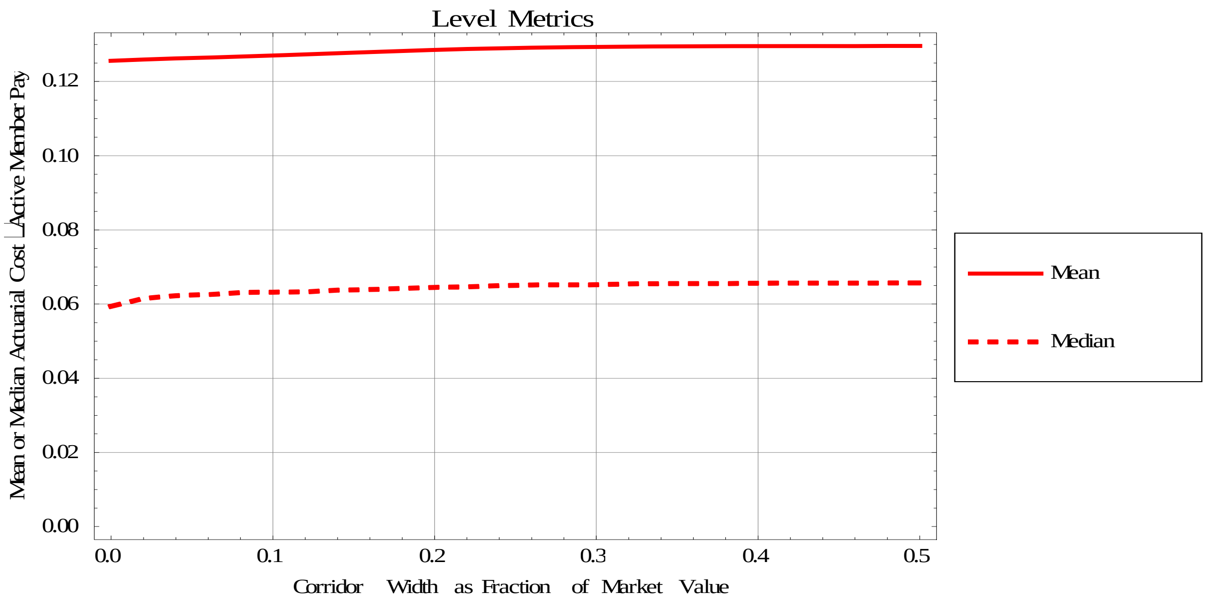
3.2 Predictability of Cost

We have investigated a number of traditional measures of global cost variability, including the standard deviation and interquartile distance of simulated Model Plan cost. While these measures provide some information about the overall variability of plan cost, they discriminate poorly among policy options. The internal operation of actuarial smoothing methods produces a significant reduction in the standard deviation of cost in the first few years after the method is adopted. However, after that initial period, the overall variability of cost is not significantly reduced.

An actuarial smoothing method adds a memory to the cost calculation system. Unexpected investment gains and losses are spread over a number of years, being recognized gradually over time [3]. Therefore, each year's actuarial cost depends not just on the current value of assets, but on asset values and investment gains and losses in past years. Given this historical information, it may be that cost *variation* is not affected much by actuarial smoothing, but cost *predictability* from year to year is substantially improved. Clearly, predictability is of great importance to plan sponsors who must include a pension cost in each year's budget. Are there metrics that measure cost predictability and that are useful for discriminating among asset smoothing polices?

---

[6] McCrory, R. 2012. "Modeling Defined Benefit Pension Plans: Basic Dynamics," in *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*, Orlando, Fla. (submitted).



Corridor Width	0%	12%	25%	38%	50%
Mean	0.125594	0.127305	0.128974	0.129536	0.129607
Median	0.059405	0.063283	0.064961	0.06555	0.065703

Figure 2: Cost Level Metrics by Smoothing Corridor for the Model Plan
(Width of the corridor on the horizontal axis is a fraction of market value; cost is expressed as a fraction of member payroll on the vertical axis.)

Figure 3 below plots the actuarial cost at time $t + 1$ against the actuarial cost at time t; no asset smoothing is in place. We note considerable variation between successive actuarial costs caused, of course, by changes in the value of plan assets. Note also that the cost variation decreases with increasing cost; at higher costs, asset levels are depressed and investment returns have a correspondingly lower impact on plan cost.[7]

When we compute a smoothed value of assets, at any time t we know the cost at t, but we also know a great deal more. In particular, we know the remaining state variables, including those investment gains and losses that have been deferred for future recognition through a smoothing method. Therefore, we can estimate the cost at $t + 1$ using this information and assuming a return on the market value of assets during the current year equal to the actuarial assumption. This estimated cost will be a better predictor of cost than the cost at time t, because the cost at time t does not take into account future recognition of deferred gains and losses. Accordingly, it would be more instructive to measure the correlation between the *actual* cost at time $t + 1$ and the *expected* cost at $t + 1$ measured at time t. This measures the unpredictability of cost. This is the content of Figure 4 below. Note in comparing Figure 4 with Figure 3 that the dispersion is smaller and the R-squared value is higher: 87.7 percent in Figure 3 (autocorrelation) vs. 98.7 percent in Figure 4 (actual/expected correlation). This is what we would expect: Knowledge of the internal state of the plan should give us additional insight into the cost next year.

In comparing Figures 3 and 4, we note that the dispersion of the array of points seems to change more than the R-squared value would suggest. Does this suggest another measurement? The Standard Error of Estimate (SEE) is the square root of the sum of the squared deviations of data points from the linear regression line. In our case the regression line is actual cost = expected cost, so the deviations from this line are significant. In Figure 3 the SEE is 4.94 percent, while in Figure 4 it is 1.58 percent, a reduction of almost two-thirds. Therefore, the actual cost/expected cost (A/E) SEE seems to be a better discriminator of cost unpredictability than the R-squared statistic.

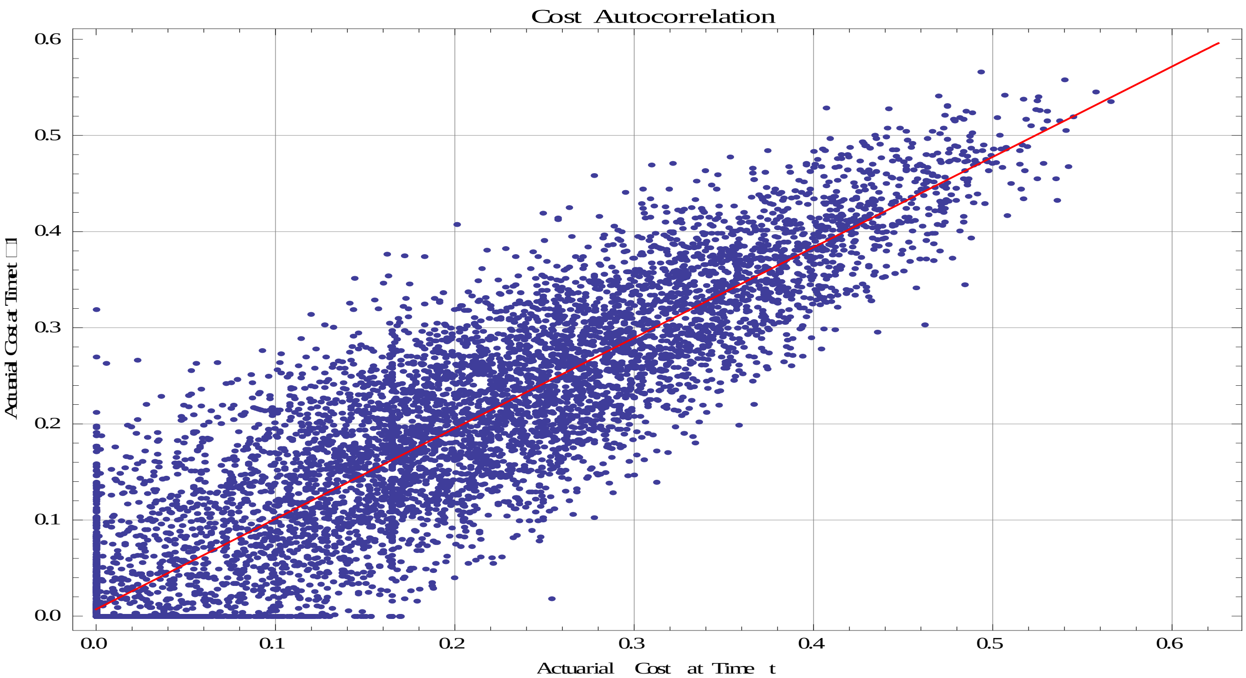
[7] McCrory, R. 2012. "Modeling Defined Benefit Pension Plans: Basic Dynamics," in *Proceedings of the 2012 Industrial and Systems Engineering Research Conference*, Orlando, Fla. (submitted).

We can now look at the A/E SEE for all smoothing corridors in Figure 5 below. Here we see that the A/E SEE decreases by a factor of four as we widen the smoothing corridor. By the time we get to a 20 percent width we have achieved about 88 percent of the possible improvement; to get to 95 percent of the maximum improvement we have to widen our corridor to 30 percent.

The A/E SEE seems to be a sensitive measurement of the right thing: How much will actual plan cost vary from our annual projections, using everything we know about the plan?

4. Equity Metrics

One of the goals of actuarial funding is generational equity. The concept of generational equity is that each generation of people who use the products of the employer—customers in the private sector and taxpayers in the public sector—should pay the cost of pension benefits for those employees who serve them. For example, each generation of taxpayers should pay for the pensions earned by the police who protect them; the cost of pensions for retired police officers should have been paid by prior generations of taxpayers.



Regression Line	R^2	Standard Error
0.00711682 +0.940955 x	0.877384	0.0494381

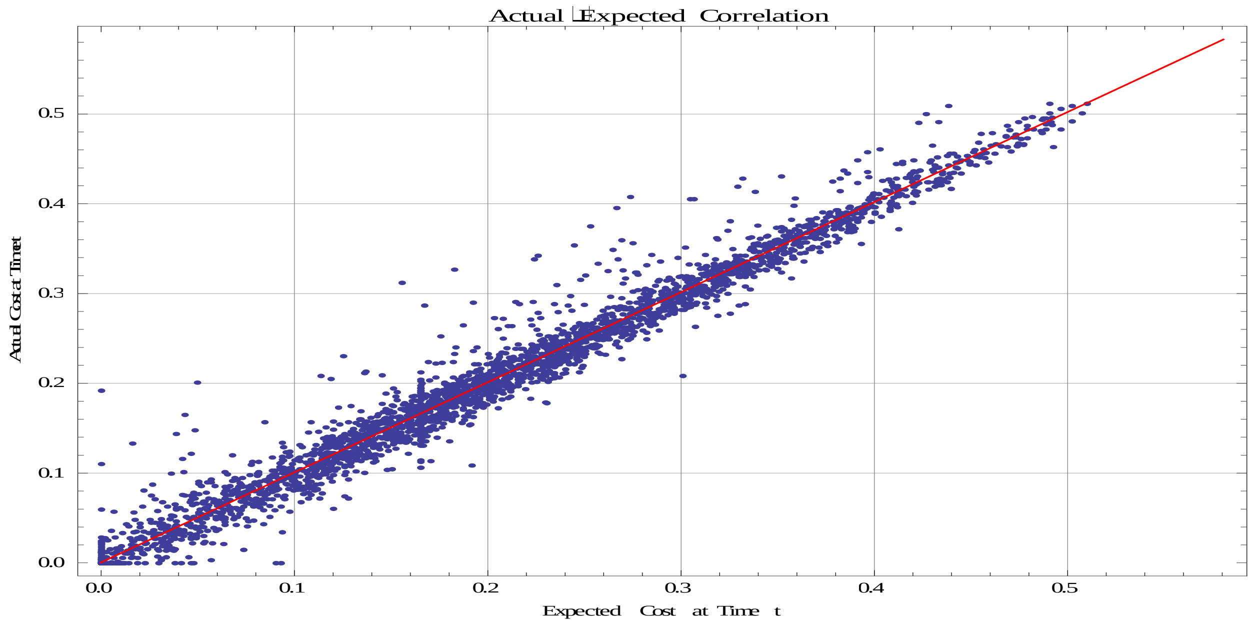
Figure 3: Plot of Cost at Time *t* + 1 vs. Time *t* for the Model Plan with No Asset Smoothing
(Each trial and year is shown along with a regression line and its parameters; 10,000 points are shown.)

How do we measure generational equity? Here we present several approaches. One method compares the actuarial cost of the plan—under whatever actuarial funding method we are using—with a standard of generational equity represented by the Individual Aggregate Cost Method.[8] This funding method is designed to enforce a high standard of equity across generations. A second method compares plan asset levels, measured by the funded ratio—the ratio of the market value of assets to the asset target, or actuarial accrued liability—with standards of over- and underfunding.

We start with the Operating Region of the plan. The Operating Region is a graph of actuarial cost (vertical axis) versus the funded ratio using market value of assets (horizontal axis). On this graph a point can be plotted at any

[8] Schneider, P. and B. Freedman. 2003. ERISA: A Comprehensive Guide, New York, N.Y.: Aspen Publishers, Inc.

time *t* that represents the status of the plan at that time. Where on this graph do we want our plan points to lie if we want to achieve generational equity? The region of the graph in which we want our plan's cost and funded ratio to lie could be called the Equity Region of the plan.



Regression Line	R^2	Standard Error
0.000447048+1.00347 x	0.987139	0.0157783

Figure 4: Plot of Actual Cost at Time *t* vs. Expected Cost at Time *t* Measured at *t* - 1 for the Model Plan with Five-Year Asset Smoothing Using a 20% Corridor (Each trial and year is shown along with a regression line.)

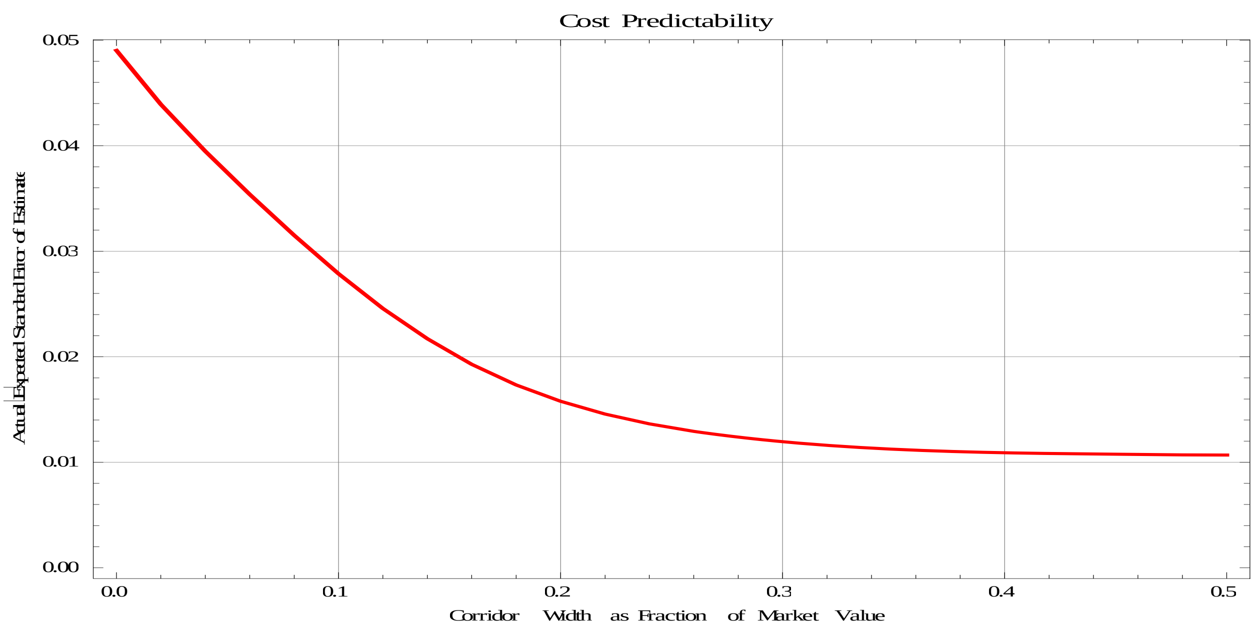


Figure 5: Plot of Cost A/E SEE for the Model Plan vs. Asset Smoothing by Corridor

4.1 Minimum Funded Level

When is a plan's funded ratio too low? For our purposes, we will look at the Inactive Funded Ratio, the level of funding necessary to pay all benefits for currently retired members. If the plan's funded ratio drops below this level, then insufficient assets have been accumulated to pay for current retirees, and no assets at all have been accumulated to fund future retirement benefits for current active members. This is a clear violation of any concept of generational equity. In the Model Plan, the funded ratio must be 60.0 percent or higher just to cover current inactive members, which is similar to the levels of inactive liabilities we see among actual plans. Therefore, our funded ratio should be in the region in Figure 6 above 60 percent.

4.2 Maximum Funded Level

When is a funded ratio too high? High funded ratios have historically been regarded as an opportunity to increase benefits or to take a contribution holiday. Either course of action has had unfortunate results. We suggest a policy decision that the funded ratio should not exceed the point at which the employer cost is zero. For our Model Plan, this occurs when the funded ratio reaches 125 percent. From the standpoint of generational equity, when this level of funding is achieved it indicates that prior generations of taxpayers have contributed too much, as things turned out, and that for some years current taxpayers will be free of contribution responsibilities. Accordingly, the area in Figure 6 includes funded ratios from 60 percent (the inactive-only funding level) to 125 percent (the zero-cost funding level).

4.3 Cost Equity

To set a policy for generational equity, we decide that, as a matter of policy, we will stay within a range of the Individual Aggregate Cost; the range will be plus or minus 50 percent of the normal cost of the Model Plan. This tolerance is arbitrary; another tolerance level may be more appropriate in a particular situation. The Cost Equity Region is shown in green shading in Figures 6 and 7. The shape of this region is interesting: When funding is below the inactive-only funded ratio of 60 percent, very high contributions—perhaps unrealistically high—are required to restore generational equity. The dashed red line in Figures 6 and 7 shows the Individual Aggregate Cost resulting from each funded ratio. The solid blue line shows the cost using usual actuarial cost method.

4.4 Equity Measured

Now let’s revisit the measurement of the effect of asset smoothing corridors, this time on Generational Equity as measured by the Equity Region of the Model Plan. To this end, Figure 6 shows the Operating Region of the Model Plan with no asset smoothing, and Figure 7 shows the Operating Region with smoothing using a 20 percent corridor. In each graph we show 10,000 simulated funded ratio/cost points. In Figure 6 we see that without smoothing 5,979 of the 10,000 points—about 60 percent—fall within the Equity Region. The points that fail this test are those with funded ratios below 60 percent, because the Individual Aggregate Method used as an equity standard requires much higher contributions when funding drops below the Inactive Funded Ratio.

In Figure 7 we observe that cost is no longer a simple function of the market value of assets. For each market funded ratio there is now a range of possible costs. The plan cost exhibits memory: The current year’s cost depends not only on current market value of assets, but also on the history of past investment returns. Moreover, note that the dispersion of simulated points increases with the funded ratio. Now only 4,252 funding/cost points fall within the Equity Region; while the remaining 5,748 points (57.5 percent) lie outside the Equity Region, indicating that our definition of generational equity is violated in those cases.

4.5 Equity Metrics by Policy

Figure 8 below is the plot of the equity measures for years 20 to 100 against the width of an asset smoothing corridor with five-year smoothing. For the most part, the chances of an equity failure increase gradually as the smoothing corridor widens, something we would expect. However, the chance of actuarial cost significantly different than Individual Aggregate Cost increases dramatically with a corridor of 10 to 15 percent or so. The sharp increase in cost failure occurs because the Individual Aggregate Cost is based on market value. When the actuarial value of assets drifts away from market value, actuarial cost differs from Individual Aggregate Cost, and we have the results shown. Indeed, after the corridor reaches 15 percent or so, all of the impact on generational equity has been realized and equity is not further compromised by a wider smoothing corridor.

The corridor width at which the cost equity metric jumps will depend on the tolerance that is allowed around the Individual Aggregate Cost standard. A wider tolerance will cause cost equity failure to increase sharply at higher asset corridor widths. Therefore, selection of the amount of deviation permitted is important in our performance measurements.

5. Specific Metrics

In any policy discussion, there are metrics that are unique to the policy in question. In the case of asset smoothing, one such metric could be the average absolute value of the distance between smoothed asset value and market value. Figure 9 below shows an example of such a metric: The mean absolute value of the difference between actuarial and market value of assets, as a percentage of market value, for the 100-year projection of the Model Plan with cost computed using four different smoothing methods: assets at market, five-year smoothing with a 20 percent corridor, smoothing by recognizing 1/15th of the difference between expected and actual asset values with a 20 percent corridor, and 20-year smoothing with a 50 percent corridor.

This simulation begins with smoothed assets equal to market value, and the smoothing policy is instituted at time 0 and phased in over time. Therefore, it is no surprise in Figure 9 that very soon after the smoothing policy is adopted, the average distance between actuarial and market value rises to an ultimate value and stabilizes.

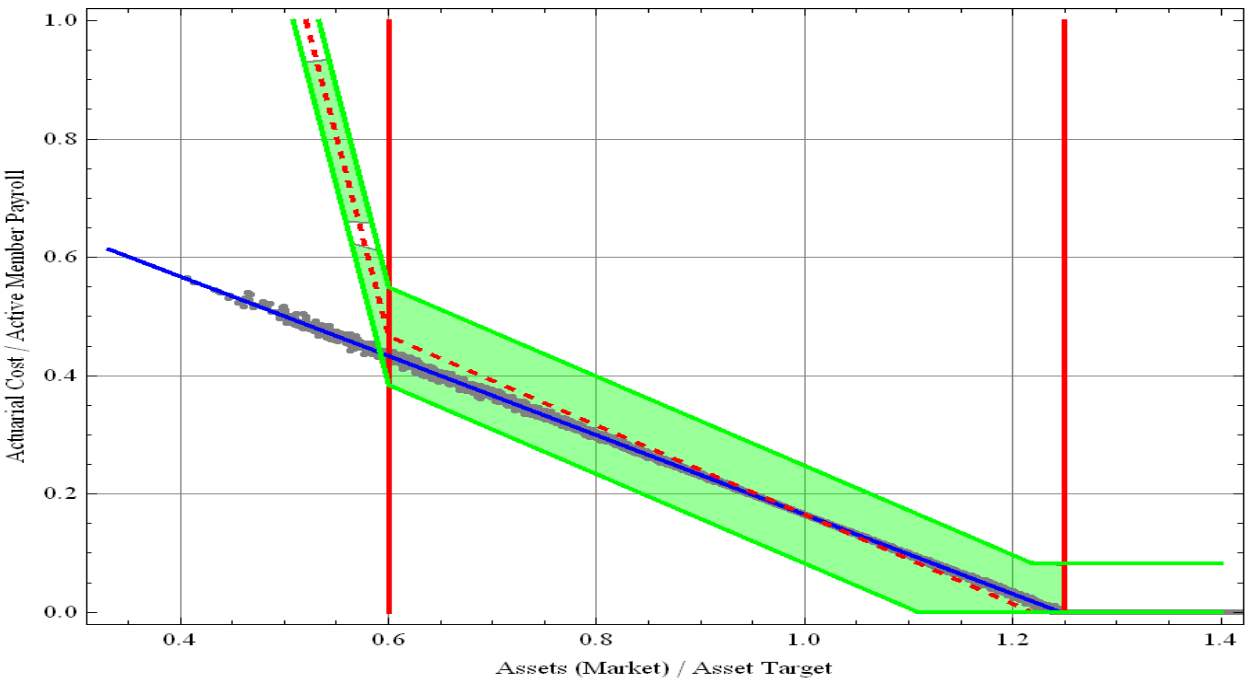


Figure 6: Equity Region for the Model Plan without Asset Smoothing
(The horizontal axis is the funded ratio using market value of assets, while the vertical axis is the actuarial cost. Simulated funding-equity points are shown in gray; 10,000 points are shown, of which 5,979 are within the Equity Region. Shown also are the Individual Aggregate Cost by funded ratio (dashed red line), actuarial cost by funded ratio (solid blue line), and an Equity Region within 50% of the normal cost of the Individual Aggregate Cost (green shading).)

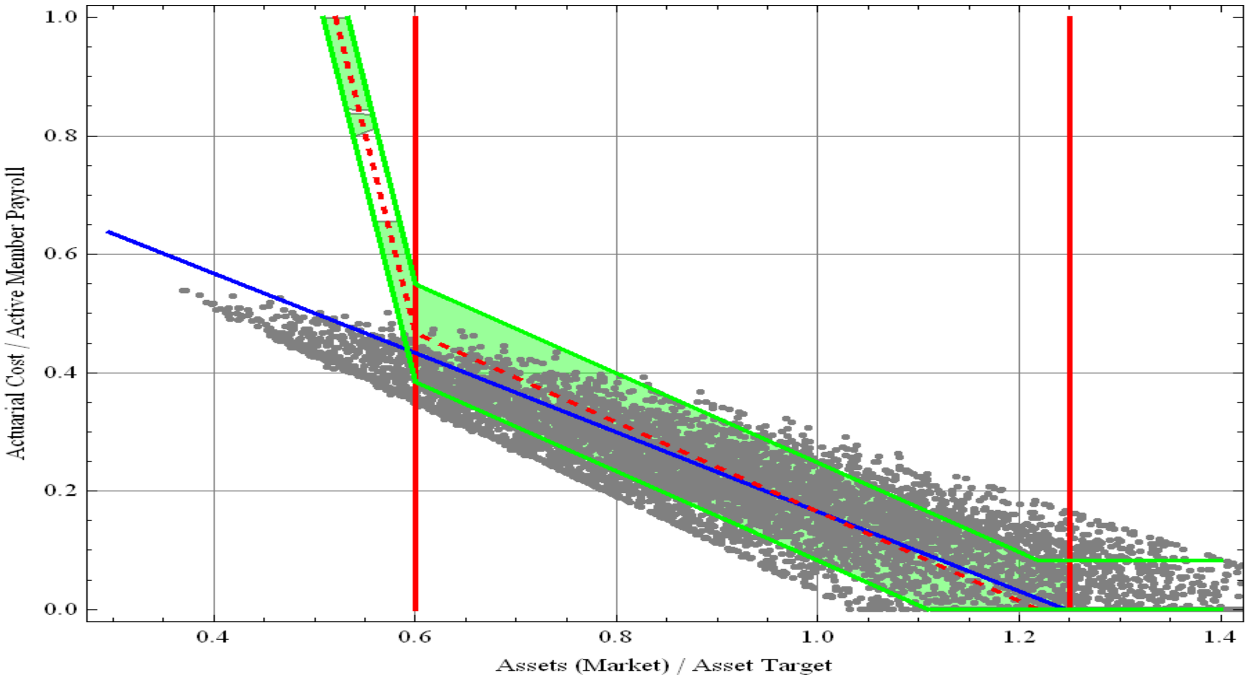
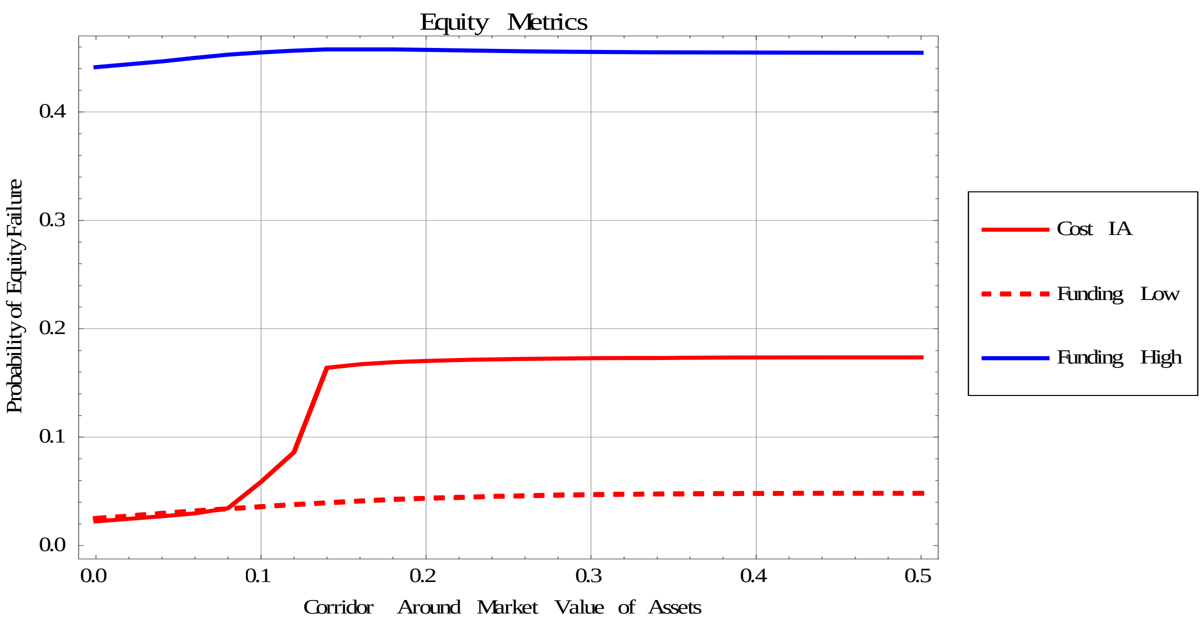
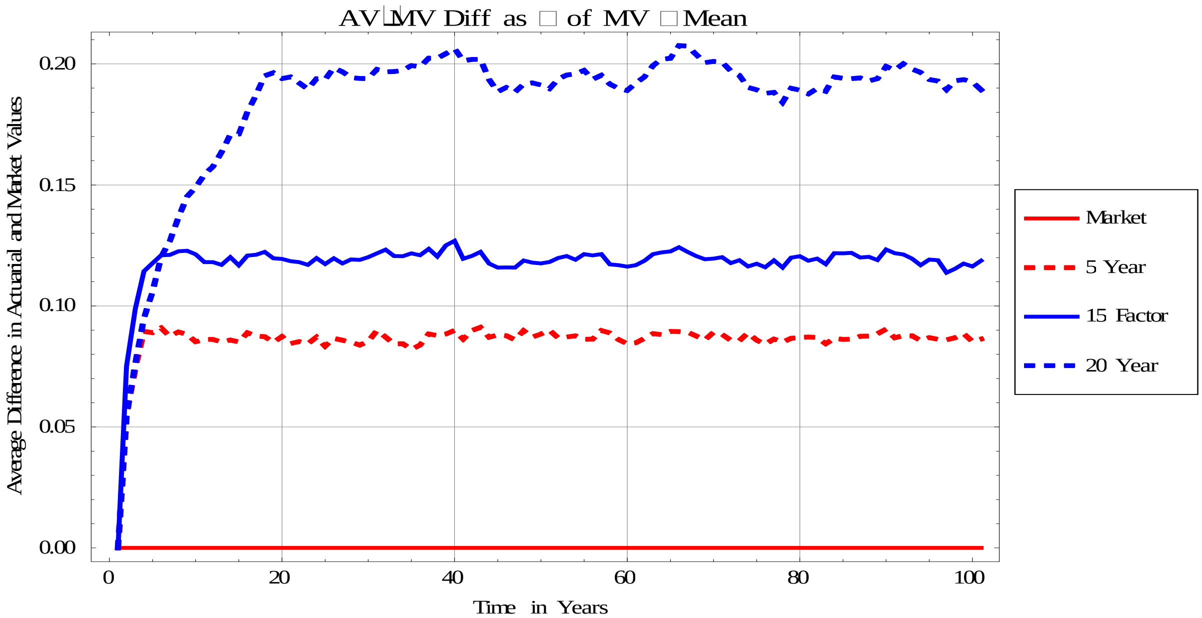


Figure 7: Equity Region for the Model Plan with Asset Smoothing
(Shown are simulated smoothed costs vs. the funded ratio at market value superimposed on the Equity Region. A 20% asset smoothing corridor around market is used. Of the 10,000 funded ratio/smoothed cost point, 4,252 are within the Equity Region.)



Corridor Width	0%	12%	25%	38%	50%
Cost Equity Failure	0.022284	0.086136	0.171827	0.17342	0.173704
Underfunding	0.025062	0.037852	0.045296	0.047963	0.04821
Overfunding	0.441469	0.456679	0.456457	0.454938	0.454691

Figure 8: Probability of Equity Failure by Asset Smoothing Corridor



Year	0	25	50	75	100
Market Value	0.	0.	0.	0.	0.
5 Year/20% Corridor	0.	0.086612	0.089974	0.08423	0.086335
Factor of 15/20% Corridor	0.	0.119661	0.118173	0.11597	0.118721
20-Year/50% Corridor	0.	0.198355	0.189761	0.187835	0.189332

Figure 9: Mean Absolute Value of the Difference Between Actuarial and Market Value of Assets, Over Time, by Smoothing Policy

After 20 years, the more smoothed the assets are, the farther apart are actuarial and market values, which is also no surprise. The asset corridor keeps the actuarial/market value distance in check.

This statistic—the average difference between smoothed and market value of assets—would be an important part of a suite of metrics used to evaluate the selection of a cost smoothing method.

6. Summary and Conclusions

There is measure in all things. — Horace

Horace, over 2,000 years ago, had never heard of defined benefit-pension plans but, if he had, he would have known that there are measures that can be applied to them, just as to anything else in the world.

However, finding the right metrics and applying them intelligently is not a trivial task. Each metric must be evaluated to determine if it is appropriate and discriminating for the task at hand: Appropriate for judging competing polices, and yielding a useful range of values to discriminate among them.

For example, we noted that mean actuarial cost is not affected materially by the actuarial cost smoothing method adopted, and so is a poor metric for that application. However, if the policy in question were investment policy, mean cost would be a critical decision factor. In every case, proposed metrics must be tested against proposed policies to determine if they are useful.

In this paper we have reviewed a number of measures of cost: Level, variability and predictability. We have also considered the generational equity of pension funding, and we have touched on specific measures that may be appropriate for evaluating certain policies. We have seen that many of the usual suspects—mean and standard deviation in particular—may not be particularly useful due to instability over time or relative insensitivity to changes in policy. On the other hand, we have found some less common measures that may be very useful, at least in some circumstances.

Throughout there are trade-offs in setting policy: Cost level versus cost variability, cost level versus generational equity, predictable cost versus equitable cost, and on and on and on. All these tensions and trade-offs are inherent in setting policy in general, and defined-benefit plan funding policy in particular. How much generational equity do we sacrifice when we invest in high-return but risky assets? What are the effects of asset smoothing on level and predictability of cost and on generational equity? All such questions remain unanswered unless the various funding goals are identified, measured and compared with an appropriate set of metrics.