# Retirement Income Analysis
## with scenario matrices

## William F. Sharpe

# 19. Ratchets

### *Habit Formation*

As discussed at length in Chapter 9, we generally take the view that the amount of welfare that retirees obtain from a set of possible future retirement incomes can be considered the sum of expectations of the desirabilities of incomes in each of a number of future years. More specifically, we assume explicitly or implicitly that preferences can be represented as a series of utility functions, one for each future year and personal state, with the overall desirability of a set of possible future incomes measured by the sum of the expected utilities in each of the future years and states. More succinctly, we assume that people have *time-separable utility functions*. This viewpoint motivates the graphs which plot the relationships between PPC and real income in each future year and relevant personal state or states as well as the derived measures of the associated cost-efficiencies of incomes.

We have allowed for the possibility that the curve representing the utility of income in a particular year and personal state may have a kink at some *reference level* of income, so that utility decreases at a greater rate just below that income than the rate at which it increases just above it. Formally, this means that an associated curve with marginal utility on the vertical axis and income on the horizontal axis will be discontinuous at that level of income, but this can be approximated, if needed, by including a very steep section in a continuous function.

Consider, for example, a couple that earned a real income of $X in the last year before retirement. Given savings in commuting costs, formal attire, etc., they may feel that an annual real income of $0.75X in their first year of retirement will provide a comfortable standard of living. But they also feel that $1 less than this amount will lower utility by considerably more than $1 more will increase it; formally, their utility curve has a kink at the reference level of $0.75X. If they will continue to feel this way in future years , with a real income of $0.75X remaining as an important reference point, no violence is done to our implicit assumption that retirees have time-separable utility functions. This would be true even if the couple could not contemplate life with an income of less than $0.75X, with each year's utility curve vertical (or almost vertical) at that point.

But what if the couple's reference point for income in a year depends in some manner on their income in the previous year? Perhaps their utility curve is kinked for a real income equal to the amount they obtained in the prior year. Or maybe the kink occurs at a point equal to the average of their last 3 years' income. Or some other recent period. In any such case, the income obtained in, say, year *t* affects utility functions for one or more later years. This greatly complicates the task of evaluating the desirability of a change in income in one year, since the direct impact on the utility of that income in the year in question and also the impact of that income on the utility functions for income in future years need to be taken into account.

The idea that previous consumption influences preferences for present consumption is not new. As early as 1949, J.S. Duesenberry incorporated such behavior in a macro-economic model deal with *Income, Saving and the Theory of Consumer Behavior*. In a 2005 note titled *The Mysterious Disappearance of James Duesenberry,* decrying the failure of the profession to adopt his approach, Robert H. Frank wrote:

> *To explain the short-run rigidity of consumption, Mr. Duesenberry argued that families look not only to the living standards of others, but also to their own past experience. The high standard enjoyed by a formerly prosperous family thus constitutes a frame of reference that makes cutbacks difficult, which helps explain why consumption levels change little during recessions.*

In a 1995 Review of Economics Studies paper titled *Duesenberry's Ratcheting of Consumption: Optimal Consumption and Investment Given Intolerance for any Decline in Standard of Living*, Philip Dybig noted:

> *A number of models with preferences exhibiting such types of habit formation have been proposed. And many institutional funds have default spending rules consistent with such preferences. It is not unusual for a non-profit organization to have a policy that calls for spending a given percentage of the average value of its endowment over some number of previous years or months (for example, 5% of the average of the past 36 month-end values). Such policies, designed to smooth the amounts spent from year to year, are sometimes adjusted following dramatic changes in endowment values, usually by altering the percentage spent. Nonetheless, the goal is to avoid radical changes in annual budgets.*

An extreme form of such an approach requires that income never declines. When conditions are favorable, annual income may increase, but it will never decrease. This is often termed a *ratchet policy*.

Miriam-Webster's Learner's Dictionary defines ratchet as:

> *a device made up of a wheel or bar with many teeth along its edge in between which a piece fits so that the wheel or bar can move only in one direction*

This chapter covers two approaches that provide income that can only increase – that is, "ratchet up". The first is guaranteed by an insurance company. The second has been suggested for use by individuals, possibly with guidance from an investment advisor. These ratchet strategies are illustrative, not exhaustive, but should provide insight into the broader class of income strategies designed for retirees with preferences for income that cannot be adequately served by approaches that are consistent with maximization of time-separable utility functions.

Now to the guaranteed approach.

## *Variable Annuities with Guaranteed Lifetime Withdrawal Benefits*

Wouldn't it be nice if a couple could (1) spend from accumulated savings without any need to cut spending if returns are poor, (2) be assured that such income would last as long as one or both are alive, and (3) be able to take additional money from savings in the event of an emergency if they are willing to compromise (1) and/or (2) to do so? At some cost, and possibly risk, they can. The vehicle is a financial product that combines mortality pooling and insurance against the possible effects of poor investment returns.

Several such products are available. Some use the term *Guaranteed Minimum Withdrawal Benefit* (GMWB), others *Guaranteed Lifetime Withdrawal Benefit* (GLWB)  We will focus on one of the latter that is representative of the genre and has relatively low costs.

Here is the description of the product, offered by Vanguard in late 2016:

> ***Guaranteed income for life through the Vanguard Variable Annuity***
>
> *If you withdraw your retirement assets without a clear plan, you increase your risk of running out of money. Secure Income™, the optional Guaranteed Lifetime Withdrawal Benefit (GLWB) rider available through the Vanguard Variable Annuity, offers you protection from market volatility and guaranteed payments for life.**

Of course there is the footnote:

> ** Product guarantees are subject to the claims-paying ability of the issuing insurance company.*

Here is how the product works.

First, you invest your savings in one of three Vanguard funds. Each has a combination of stocks and bonds. The Balanced Portfolio invests 60%-70% in stocks, the Moderate Allocation Portfolio 60% in stocks and the Conservative Allocation Portfolio 40% in stocks. Expense ratios for the three funds range from 0.44% to 0.73%. These costs include an *administrative fee* of 0.10% and a *mortality and expense risk fee* of 0.19% (although the reason for the latter is not stated). The description indicates that the average (total) expense ratio is 0.54%.

Second, you contract for a *Secure Income Rider* from a unit of Transamerica Life Insurance, a company that is now part of Aegon, a Dutch company. According to its web site, Aegon is ".. one of the top-10 largest insurance companies in the world ... one of the worlds' leading providers of life insurance, pensions and asset management .. (and has) operations in over 20 countries, including the USA, where we're known as Transamerica." The US roots go back to The Bank of Italy, created in 1904 by Amadeo Giannini in a converted San Francisco saloon.

The Vanguard site indicates "The guarantee is subject to the claims-paying ability of the issuer … highly rated for financial strength." In late 2016, the Transamerica web site provided the following ratings of "the relative financial strength and operating performance of the company" as of effective dates in 2011:

> *A.M. Best's A+ rating is the second highest of 16 ratings*
>
> *Standard & Poor's AA- rating is the fourth highest of 21 ratings*
>
> *Moody's A1 rating is the fifth highest of 21 ratings*
>
> *Fitch's AA- rating is the fourth highest of 19 ratings.*

Is *Secure Income*™ fully secure? Only time will tell. Let's see what the rider promises.

A key ingredient is the *Total Withdrawal Base* (TWB). The amount that can be withdrawn without compromising the guarantee is a fixed percentage of the TWB. This *withdrawal percentage* is fixed at the time the rider is purchased. Here is the 2016 table and footnote:

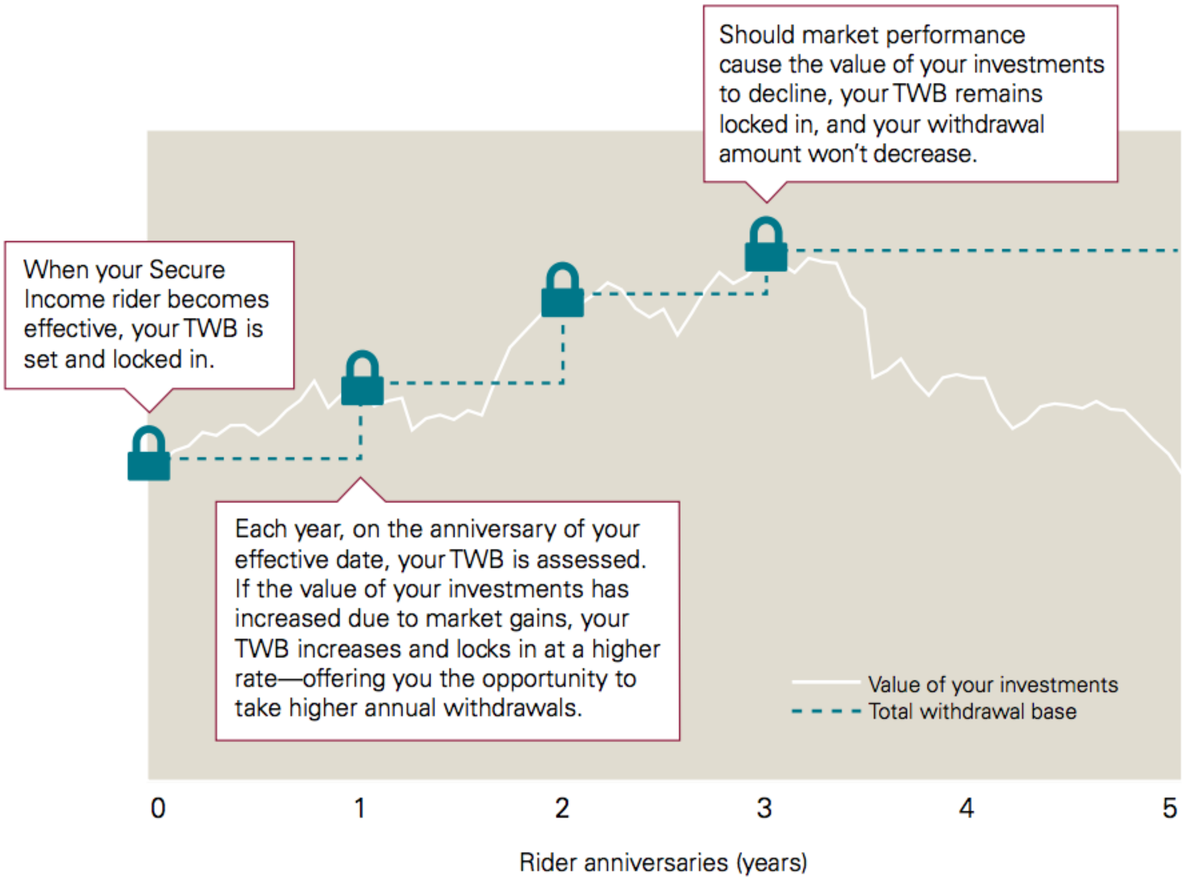
### Annual withdrawal percentages

| **Age at first withdrawal** | **Single life rider** | **Joint life rider** |
|---|---|---|
| 59–64 | 4.00% | 3.50% |
| 65–69 | 5.00% | 4.50% |
| 70–79 | 5.00% | 4.50% |
| 80+ | 6.00% | 5.50% |

*If you choose a joint life rider, the withdrawal percentages are based on the younger of the annuitant or the annuitant's spouse when withdrawals begin.*

This will undoubtedly seem familiar – remember the 4% rule? However, that approach attempted to keep *real* income constant; this approach is designed to keep *nominal* income constant or increasing – a significant difference, as we will see.

A graph from the Vanguard web site explains the way the TWB is determined each year:




Withdrawals are a fixed percentage of the TWB, which can never decrease. Thus withdrawals (income) can never decrease. If investment returns are sufficiently large to offset withdrawals and fees, then both the TWB and all future withdrawals may be greater – an *income ratchet* indeed. But it is a *nominal* income ratchet. As we will see, real income could well decrease in many or most years.

The possible benefits of the an approach are shown by an example on the Vanguard web site. It assumes that $250,000 was invested in an account with a GLWB rider at the beginning of 2003. Voila! No financial crisis for this lucky investor:


One must, of course, pay for the Secure Income rider. In late 2016, the annual cost was 1.20% of the *Total Withdrawal Base* each year (deducted quarterly) no matter what the actual value of the account might be. Another document on the Vanguard web site indicates that after the initial year "The rider fee for future premium payments … could be higher or lower, but not more than the maximum of 2.0%." The supplement to the prospectus makes clear that increased fees are applicable only to any additional money that might be added to the fund by the beneficiaries, so absent this, the fee at the creation of fund remains applicable for the initial investment.

Overall, this rather complicated arrangement provides the insurance company with fee income for some number of years, but requires it to provide income if the account value reaches zero before the beneficiaries are both dead. The insured investors pay fees for some number of years. If they live long enough, they may then receive payments (negative fees) from the insurance company until they die.

Unlike traditional annuities, the commitment to this arrangement by the retirees is not irrevocable. At any time, they may withdraw more than the designated percentage of the current TWB. However, any such withdrawal will reduce the TWB by an amount equal to or greater than the difference between the amount actually withdrawn and that provided by the formula. The summary to the prospectus provides a formula and an example in which the TWB is $100,000, the actual value of the fund is $90,000, the MAWA (maximum annual withdrawal amount) is 5.5% of the TWB or $5,500, but the total amount withdrawn is $7,000: $1,500 more than the formula would provide. In this case the TWB is reduced by $1,775.15 rather than the $1,500 *excess withdrawal*. Complexity aside, the moral is that excess withdrawals diminish the value of the guarantee, and withdrawing the entire fund value removes the guarantee completely.

Clearly, this is a complex financial product, but one with considerable appeal for some retirees. We will attempt to capture the essence of such Guaranteed Lifetime Withdrawal Benefit (GLWB) annuities using the Vanguard/Transamerica combination as a template. However, we will not attempt to incorporate possible default by the insurer, actions taken in the case of excess withdrawals, and other possibly important aspects of such an insured investment.

### *The iGLWB_Create Function*

For simplicity, we will call the data structure for a variable annuity with a guaranteed lifetime withdrawal benefit iGLWB. Here is the function that can create such a structure and provide default values for its elements:

```
function iGLWB = iGLWB_create( );
  % create a guaranteed lifetime withdrawal benefit data structure

  % initial amount invested
    iGLWB.initialValue = 100000;

  % single (s) or joint (j) life
    iGLWB.singleOrJoint = 'j';

  % single life withdrawal proportions of TWB (from-age to-age proportion)
   iGLWB.singleLifeWithdrawalRates = [ 59 64 0.040; 65 79 0.050 ; 80 120 0.060 ];

  % joint life withdrawal proportions of TWB (from-age to-age proportion)
  %   based on age of younger spouse
   iGLWB.jointLifeWithdrawalRates = [ 59 64 0.035; 65 79 0.045 ; 80 120 0.055 ];

  % expense ratio for insurance rider as proportion of TWB
   iGLWB.expenseRatioOfTWB = 0.0120;

  % expense ratio for fund management and other fees
  %   as proportion of account value
   iGLWB.expenseRatioOfFund = 0.0054;

 % save fee matrices with iGLWB data structure (y or n)
   iGLWB.saveFeeMatrices = 'n';

end
```

The variables and their initial values should not be surprising, since they are set to the 2016 parameters of the Vanguard/Transamerica product. We have extended the highest withdrawal rates to age 120, since this is the maximum allowed in the overall RISMAT system (although in 2016, to purchase a policy, both annuitants had to be under 91 years of age at the inception of the contract) . In the unlikely event that a beneficiary is younger than 59 we will apply the withdrawal rate for the initial age range, although the insurance provider would undoubtedly either reject the application or insist on a considerably lower payout rate.

Note that the terms for the Vanguard/Transamerica product (a) provide the same withdrawal rate for a wide range of ages (in particular, from 65 through 79), (b) are the same for both sexes and (c) do not take into account the age of the older partner in a joint policy. This may be convenient but seems rather crude for a number of reasons. First, we know that women have longer life expectancies than men of the same age. Moreover, the age of the older partner in a couple matters: a rider for a couple in which the younger partner is, say, 65 and the older 90 should be more profitable for the insurance company than one in which both partners are 65. Finally, a beneficiary at the top of an age range is less likely to outlive his or her investments than one who is at the bottom of the range. A desire for simplicity seems to have outweighed actuarial imperatives. We will return to this issue later in the chapter.

The function ends by setting a data element that will determine whether or not to add the matrices with fees for fund expenses and rider costs and contributions to the IGLWB data structure so that they may be analyzed later, if desired.

### *The iGLWB_Process Function*

Moving on, here is the first part of the *IGLWB_Process* function:

```
function [ client  iGLWB ]  = iGLWB_process ( client, market, iGLWB )

  % set parameters
    initialValue = iGLWB.initialValue;
    expPropTWB =  iGLWB.expenseRatioOfTWB;
    expPropFund = iGLWB.expenseRatioOfFund;

  % find proportion of TWB to withdraw
    minAge = min( client.p1Age, client.p2Age );
    if lower( iGLWB.singleOrJoint ) ==  'j' ;
       tbl = iGLWB.jointLifeWithdrawalRates;
     else
       tbl = iGLWB.singleLifeWithdrawalRates;
    end; % if find( lower(iGLWB.singleOrJoint ),'j') >0;
    rows = ( minAge >= tbl(:,1) )  &  ( minAge <= tbl(:,2) );
    withdrawPropTWB = sum( rows.*tbl(:,3) );
```

Note that we plan to modify both the client and iGLWB data structures, with the revised versions returned when the function is executed, as indicated by placing their names on the left side of the equal sign in the function header. The main program (e.g. *SmithCase.m*) would then include statements such as:

```
iGLWB = iGLWB_create( );
[client iGLWB] = iGLWB_process( client, market, iGLWB );
```

Returning to *IGLWB_process,* the first section shown above assigns values from the data structure to simpler and shorter variables for convenience. The second selects the relevant withdrawal rate table depending on whether the case involves a single or joint lives. The last two statements (employing an approach only a programmer could love) find the relevant proportion of the Total Withdrawal Base to be withdrawn each year. We assume (counterfactually) that this amount is withdrawn at the end of each year, rather than in quarterly or monthly installments.

The next statements prepare needed matrices and vectors:

```
% create matrix of nominal market returns
  nrmsM =  market.rmsM .* market.csM;

 % get matrix dimensions
  [ nscen nyrs ] = size( client.incomesM );

 % set initial portfolio value vector
  portvalV =  initialValue * ones(nscen,1);
 % set vector of total withdrawal bases
  twbV = portvalV;

 % create nominal incomes and nominal fees matrices
  incsM  =        zeros( nscen, nyrs );
  feesFundM = zeros( nscen, nyrs );
  feesRiderM = zeros( nscen, nyrs );
```

First, we create a matrix of nominal market returns so that all computations can be done using nominal values (with key results converted back to real terms later in the program). Next, we find the number of scenarios and years for the analyses and create two vectors. The first, for the values of the portfolio in each scenario, has the initial value of the fund in each row; the second, for the initial total withdrawal bases, is at this point identical.

The next section creates three matrices in which results will be placed. The first is for incomes provided to the beneficiaries while one or both are alive plus the values of the estate in different scenarios. The second matrix will contain fees paid to the fund manager (based on the value of the investments), while the third will contain fees paid to the insurance company (based on the total withdrawal bases). For scenarios and years in which the investments are depleted and the insurance company must make payments to the beneficiaries, the entries in the *feesRiderM* matrix will be negative.

The next statements deal with the initial payouts and adjustments:

```
% set initial year payouts
  incsM( :, 1 ) =  withdrawPropTWB * twbV;
 % adjust portfolio values
  portvalV = portvalV – incsM(  :, 1 );
 % set initial year fees to zero
  feesFundM( :, 1 )  =  zeros( nscen, 1 );
  feesRiderM( :,  1 ) =  zeros( nscen, 1 );
```

Incomes in the initial year in each scenario are determined by multiplying the withdrawal proportion times the TWB value. In this case all the TWB values are the same and thus the incomes will be as well.

Next the incomes paid out are subtracted from the prior portfolio values to obtain a new vector of the latter. Since this is the initial year, all the portfolio values are the same, as are all the incomes, so the resulting portfolio values will be as well.

Finally, we set the fees for both the fund and the rider in year 1 to zero for all scenarios, since we choose to deduct fees at the end of each calendar year.

Next we do the hard work. The outer loop is designed to process each scenario for one year, then the next, and so on until all the years have been covered:

```
for yr = 2:nyrs
   …....
end; % for yr = 2:nyrs
```

Within this loop there are two sections: the first for states in a scenario in which someone is alive, the second for states in a scenario in which an estate is to receive any remaining funds. Other states are not processed, since there are no incomes or fees involved.

Here is the first section:

```
% find scenarios in which one or two are alive
  ii = find( (client.pStatesM(:,yr) > 0) & ( client.pStatesM(:,yr) < 4) );
  if length(ii) > 0
    % increment nominal values of portolio
     portvalV(ii) =  portvalV(ii) .*  nrmsM(ii,yr-1);
    % compute fees for fund and subtract from portfolio value
     feesFundM(ii,yr)  = expPropFund * portvalV(ii);
     portvalV(ii) = portvalV(ii) - feesFundM(ii,yr);
    % compute guaranteed withdrawals and add to incomes
     incsM(ii,yr) = withdrawPropTWB * twbV(ii);
    % subtract withdrawals from portfolio values
     portvalV(ii) =  portvalV(ii) – incsM(ii,yr);
    % compute rider fees
     feesRiderM(ii,yr) = expPropTWB * twbV(ii);
    % subtract rider fees from portfolio values
     portvalV(ii) = portvalV(ii) – feesRiderM(ii,yr);
    % for negative portfolio values, adjust rider fees
     negvalV = zeros( nscen, 1 );
     negvalV(ii) = min(portvalV(ii), 0);
     feesRiderM(ii,yr) = feesRiderM(ii,yr) + negvalV(ii);
     portvalV(ii) = portvalV(ii) - negvalV(ii);
    % set TWB values to max of portfolio values and prior TWB
     twbV(ii) =  max( portvalV(ii) ,twbV(ii) );
  end % if length(ii) > 0
```

The initial statement creates a vector with the row numbers of scenarios in which the personal state is 1, 2 or 3 in the year in question. If the length of this vector is greater than zero, the remaining statements are executed; otherwise no processing is done. Most of the statements in the following section operate only on the entries in various vectors and matrices for the selected rows, hence the *(ii)* terms.

First, each of the relevant portfolio values is multiplied by the nominal return on the market in that scenario over the prior year, giving the values at the beginning of the year in question:

```
% increment nominal values of portolio
  portvalV(ii) =  portvalV(ii) .*  nrmsM(ii,yr-1);
```

Next the fees for the fund manager are computed, based on the fund expense ratio. These are posted to the appropriate rows and column of the fund fees matrix, then subtracted from the corresponding portfolio values:

```
% compute fees for fund and subtract from portfolio value
  feesFundM(ii,yr)  = expPropFund * portvalV(ii);
  portvalV(ii) = portvalV(ii) – feesFundM(ii,yr);
```

The withdrawals are determined by multiplying the current total withdrawal base for each scenario by the withdrawal proportion. The results are posted to the appropriate rows and column of the incomes matrix, then subtracted from the corresponding portfolio values:

```
% compute guaranteed withdrawals and add to incomes
  incsM(ii,yr) = withdrawPropTWB * twbV(ii);
% subtract withdrawals from portfolio values
  portvalV(ii) =  portvalV(ii) – incsM(ii,yr);
```

Similar procedures are performed for the guaranteed income rider fees:

```
% compute rider fees
  feesRiderM(ii,yr) = expPropTWB * twbV(ii);
% subtract rider fees from portfolio values
  portvalV(ii) = portvalV(ii) – feesRiderM(ii,yr);
```

At this point, deductions of incomes and rider fees may have created some negative portfolio values. This makes no economic sense. More importantly, the rider is designed to avoid such a situation from happening. To rectify the situation we create a vector for all the scenarios that initially have only zero values. Then we adjust those in the currently chosen scenarios so that each such entry will be equal to (a) zero if the portfolio value is positive or (b) the negative value if the portfolio value is negative. For the chosen scenarios, the resulting values are added to the rider fees, giving negative values for the scenarios in which the insurance company must provide income to the beneficiary. Finally the results are subtracted from the portfolio values. This will affect only the values that were negative, each of which will thus be reset to zero:

```
% for negative portfolio values, adjust rider fees
 negvalV = zeros(nscen,1);
 negvalV(ii) = min(portvalV(ii) , 0);
 feesRiderM(ii,yr) = feesRiderM (ii,yr)  +  negvalV(ii);
 portvalV(ii) = portvalV(ii)  –  negvalV(ii);
```

Finally, each of the relevant TWB (total withdrawal base) values is reset to equal the larger of the previous TWB or the current portfolio value, providing the desired *ratchet,* where applicable.

```
% set TWB values to max of portfolio values and prior TWB
 twbV(ii) =  max( portvalV(ii) , twbV(ii) );
```

The second section within the loop that processes each year in turn is somewhat simpler:

```
% scenarios in which estate is paid
     ii = find( client.pStatesM( : , yr )  ==  4 );
     if length(ii) > 0
        % increment nominal values of portolio
          portvalV(ii) =  portvalV(ii) .* nrmsM(ii, yr-1 );
       % compute fees for fund and subtract from portfolio value
          feesFundM(ii,yr)  = expPropFund * portvalV(ii);
          portvalV(ii) = portvalV(ii) - feesFundM(ii,yr);
      % pay remaining portfolio values to estate
          incsM(ii,yr) = portvalV(ii);
          portvalV(ii) = portvalV(ii)  -  incsM(ii);
     end % if length(ii) > 0
```

In this case only rows in which the personal state equals 4 are affected. Each of these occurs in the first year after the last beneficiary has died and thus must be preceded by state 1, 2 or 3. In each such case, the estate receives the entire remaining value of the portfolio. As before, we multiply the prior portfolio values by the nominal returns in the prior year, then compute and subtract any fees paid to the fund company. The remaining values are posted to the incomes matrix, then subtracted from the portfolio values for good measure.

The remainder of the function starts by converting the incomes matrix and the two fees matrices to real values,dividing by cumulative inflation in each scenario and year:

```
% convert nominal incomes matrix to real
  rincsM = incsM ./ market.cumCsM;
% convert nominal fees matrices to real fees
  rfeesRiderM = feesRiderM ./ market.cumCsM;
  rfeesFundM =  feesFundM  ./ market.cumCsM;
```

Then the real incomes and fees are added to the corresponding matrices for the client:

```
% add results to client income and fee matrices
  client.incomesM = client.incomesM + rincsM;
  client.feesM  = client.feesM + rfeesRiderM + rfeesFundM;
```

If desired, the fee matrices are then added to the iGLWB data structure:

```
% if desired add matrices of fees to iGLWB data structure
    if lower(iGLWB.saveFeeMatrices) == 'y'
      iGLWB.feesRiderM = rfeesRiderM;
      iGLWB.feesFundM  = rfeesFundM;
    end;
```

and the function ends:

```
end % iGLWB_process
```

### *GLWB Rider Costs and Benefits*

The GLWB rider requires that payments be made to the insurance company for some period of time, the length of which will depend on changes in the market value of the underlying portfolio and the length of the insured individuals' lifetimes. If the portfolio value falls to zero while one or both of the beneficiaries is/are alive, the insurance company will provide payments until the last one dies.

It is straightforward to compute the present value of the possible cash flows to the insurance company and the present value of cash flows from the company. Assuming that the rider fee matrix has been added to the IGLWB data structure, the former can be computed with one statement:

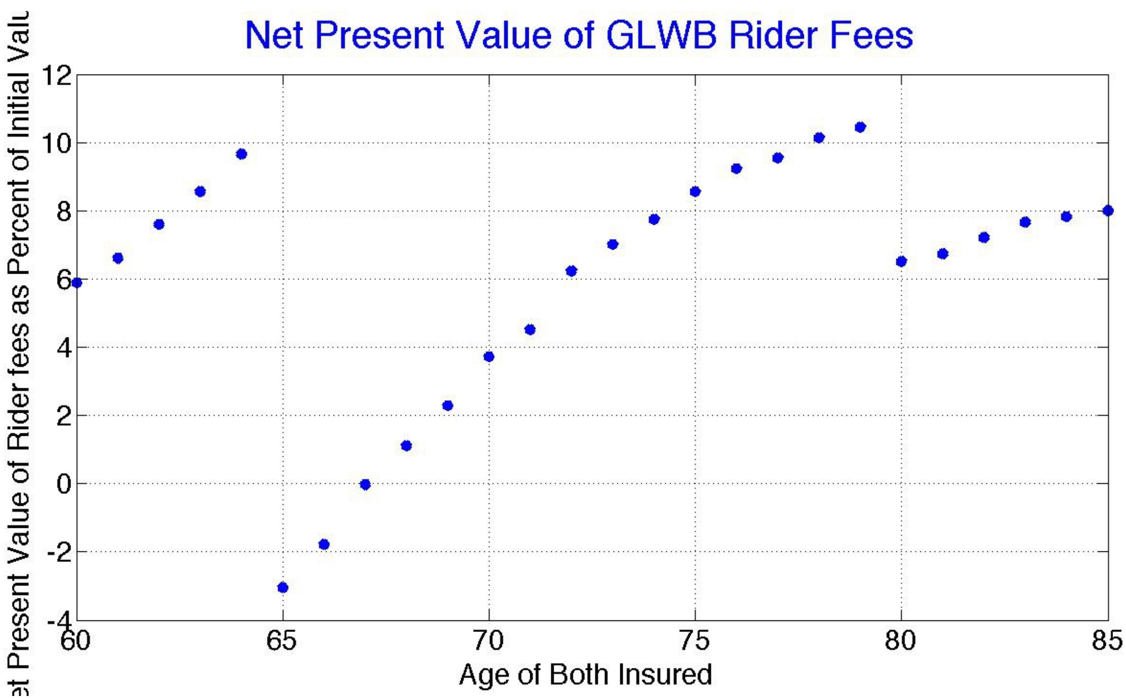
```
pvAmountsPaid = sum(sum(market.pvsM.*(iGLWB.feesRiderM.*(client.feesM>0))))
```

as can the latter:

```
pvAmountsReceived =sum(sum(market.pvsM.*(iGLWB.feesRiderM.*(client.feesM<0))))
```

The present value of the amounts paid minus that of the amounts received is the net cost of the rider in today's dollars. Dividing this by the initial investment ($100,000 in our example) gives the net cost as a percentage of the amount saved. Recall that the guaranteed withdrawal rates are the same for wide ranges of beneficiary age. Moreover, the terms depend only on the age of the younger partner even though in some scenarios the older partner may be the last to die. Thus the net cost of the rider will depend on the age of the younger partner and, within a range with the same withdrawal rate, should be greater for older beneficiaries.

The figure below shows the results of performing these computations for a number of different couples, with results expressed as percentages of the initial portfolio value. For each case, the two partners were the same age, but these ages spanned three different ranges, with withdrawal rates of 3.5%, 4.5% and 5.5% of GLWB based on the table reproduced earlier in this chapter.




The differences in expected cost to the client, and hence expected profit to the insurer, are very large indeed, suggesting that a potential buyer might want to take this information into account when deciding on the best age at which to purchase such a product (if at all).

It is somewhat startling to see that in some cases, the insurance company appears to subsidize the purchaser. This seems to be the case for the Smiths (and the present values would be even more favorable since Bob is 67 rather 65 as assumed in our figure). It is instructive to explore this issue a bit more.

First, there is the possibility that Bob and Sue will choose to withdraw more than 4.5% of the TWB at some future time. If they take everything, all payments to and from Transamerica will stop. And since payments to the company are positive up to some point and negative thereafter, this will be advantageous to the insurer. This is similar to cases in which people let long term care policies *lapse* before collecting any benefits. In Bob and Sue take more than 4.5% of TWB in any year but not the entire fund, benefits will be reduced: as the Vanguard site says: “excess withdrawals ... may reduce or eliminate the benefit provided by the Secure Income Rider.” In any event, excess withdrawals are likely to increase the profitability of the policy for the insurance company.

Second, it is important to understand that larger payments made to the beneficiaries occur when both the insured have enjoyed long lives and the market performance has been bad. In our economic model, the price (present value) of $1 in a future year is greater, the smaller is the cumulative return of the market up to that point. Thus the fact that the overall net present value of fees earned by the insurance company taking all possible scenarios into account may be negative, does not mean that the company will experience losses in the majority of possible scenarios. In fact, when Bob and Sue purchase the GLWB rider there is a substantial chance that the present value of the payments made to Transamerica will exceed that of any payments received from Transamerica in later years. In one analysis, for 88.3% of the scenarios the sum of the present values of the outflows from the beneficiaries exceeded that of the inflows to them. But many of the remaining scenarios involve payments from the insurance company in years with very bad market returns, and such payments are worth more today since they occur when money is scarce. In this arrangement, the odds are that the insurer will win, but if it loses, the losses are more painful because it will have to make payments in bad times.

Third, our analysis focuses on the scenarios that might happen for a single insured couple, but the insurance company undoubtedly has many such clients. Some will be similar to Bob and Sue, but many will be younger or older and possibly receiving different percentages of their TWB values. This provides the insurance company with some diversification, leading to a higher probability that it will profit by writing income riders of this type. That said, our analysis captures at least some of the effects of pooling mortality within a particular age cohort. Why? Because the 100,000 scenarios in an analysis include some with similar market return patterns but different mortality outcomes for Bob and Sue. To estimate the possible results for a large cohort of people like Bob and Sue, we repeated the GLWB analysis using a *client.pStatesM* matrix with the value 3 in each cell (so that both are alive in every year in each scenario). Then we multiplied the entries in each column by the probability that a similar couple would be alive in the year in question, giving the expected cash flows for a cohort of such clients. The results gave similar present values for the total amounts that might be paid and received and also the expected percentages of scenarios in which the insurance company could make money. But the standard deviation of the present values across scenarios was approximately 72% of that obtained for just Bob and Sue. Pooling the mortality of members of a particular age cohort thus can reduce the likelihood of extreme profit or loss for the insurance company. But at least for Bob and Sue's age group, the insurance company's prospects are not rosy. One presumes that Transamerica relies on other age cohorts (which, as we have previously seen, can provide significant expected profits), purchases of income riders at different times and policy lapses to make profit.

When considering all these issues, it is important to remember that security market risk in a given year can be reduced by diversifying portfolio holdings but diversifying claims on those holdings among numerous policy holders does not reduce market risk. This makes it imperative to analyze the financial condition of any party offering to guarantee income derived from security markets, whether a traditional insurance company or one of the other financial firms offering such products. *Caveat emptor*.

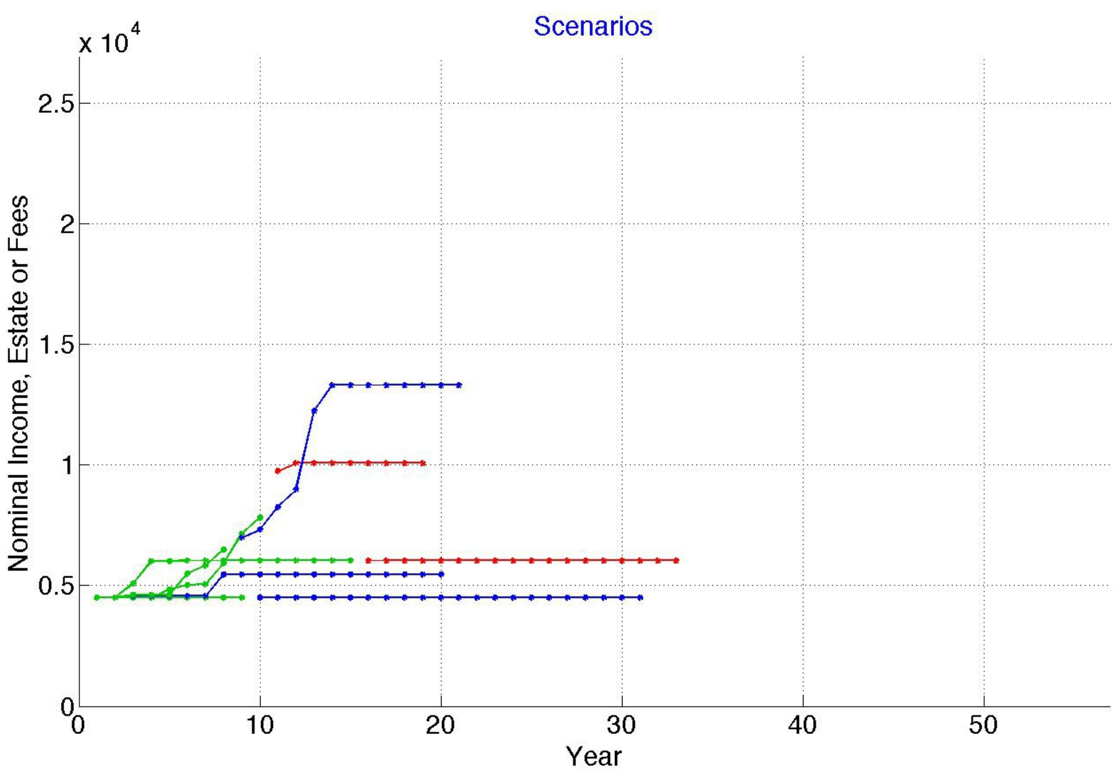
*GLWB Analyses*

It is time to find some of the properties of a GLWB approach. As before, we focus on Bob and Sue Smith and use the default elements of the client and iGLWB data structures for our example.

To begin, consider the results for five randomly selected scenarios. The figure below shows the nominal incomes, obtained by setting:

```
analysis.plotScenariosTypes = {'ni'};
```

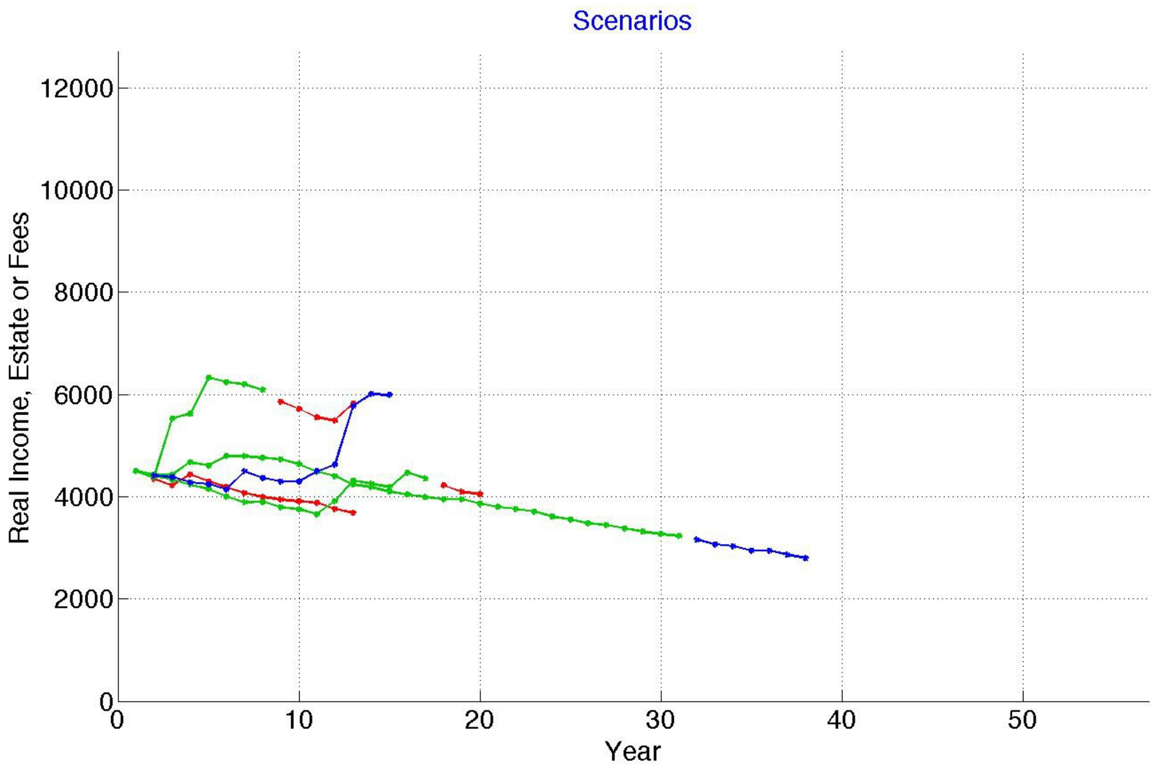
As intended, the first year's income is $4,500. In one scenario it never increases from that amount. But in others it does. There is in fact one case in which Bob and Sue both live long enough to enjoy a nominal income of over $13,000 per year.



But as we have emphasized over and over, it is *real income* that should matter to most people. To see real incomes, we set:

**analysis.plotScenariosTypes = {'ri'};**

Here is the result for five (other) randomly selected scenarios:




The picture is very different indeed. In years without a ratchet, real income tends to fall as inflation erodes the purchasing power of nominal income. In some scenarios, increases in the Total Withdrawal Base are sufficiently large and/or frequent to keep ahead of inflation, in others not. In real terms, the highest incomes are slightly more than $6,000, and in the scenario in which payments are made for 38 years, the value can fall to less than $3,000.

This highlights the key differences between a typical GLWB strategy and the type of constant spending policy policies covered in Chapter 17. GLWB strategies *maintain or increase nominal spending*, while constant spending strategies *attempt to maintain constant real spending*.

>>>

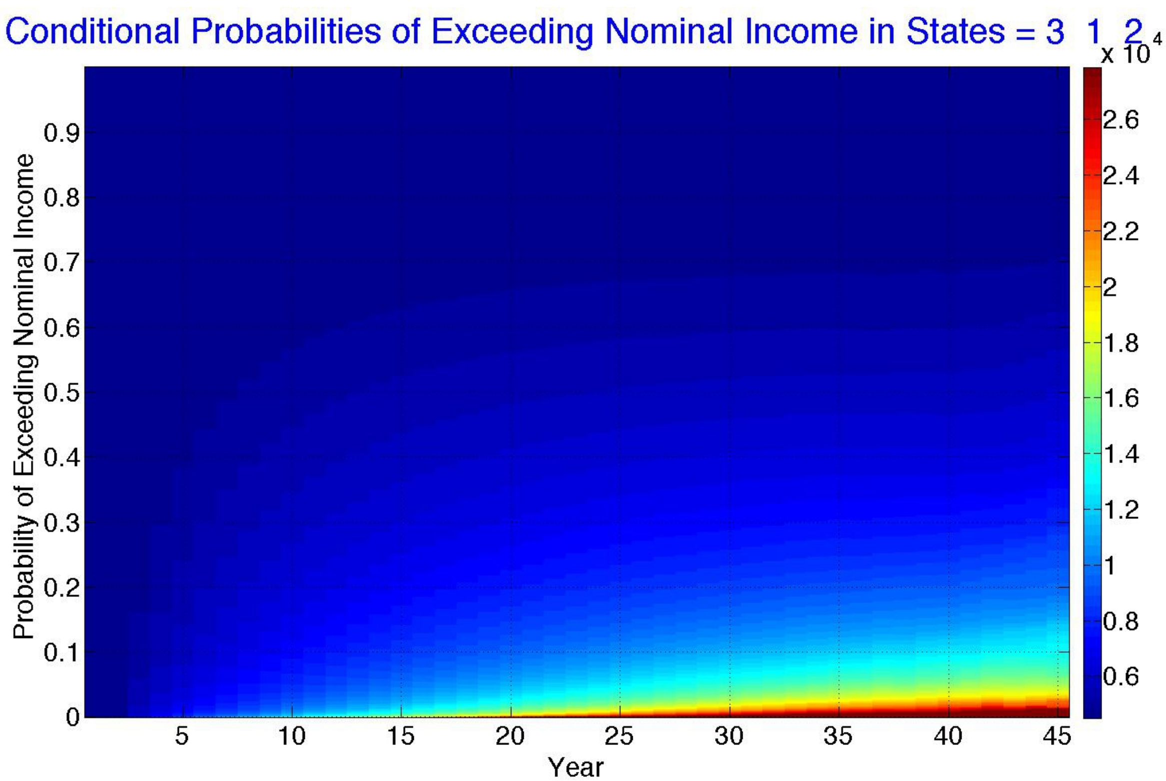
To get an idea of the ranges of incomes in different years, we produce an income map for incomes conditional on Bob and/or Sue being alive. To begin, we look at nominal incomes, setting the analysis element to:

**analysis.plotIncomeMapsTypes = {'nc'};**

And to provide a more dramatic set of colors we truncate all the incomes greater than 25% of the actual maximum income.

**analysis.plotIncomeMapsPctMaxIncome = 25;**

This provides the following when the analysis structure is processed:



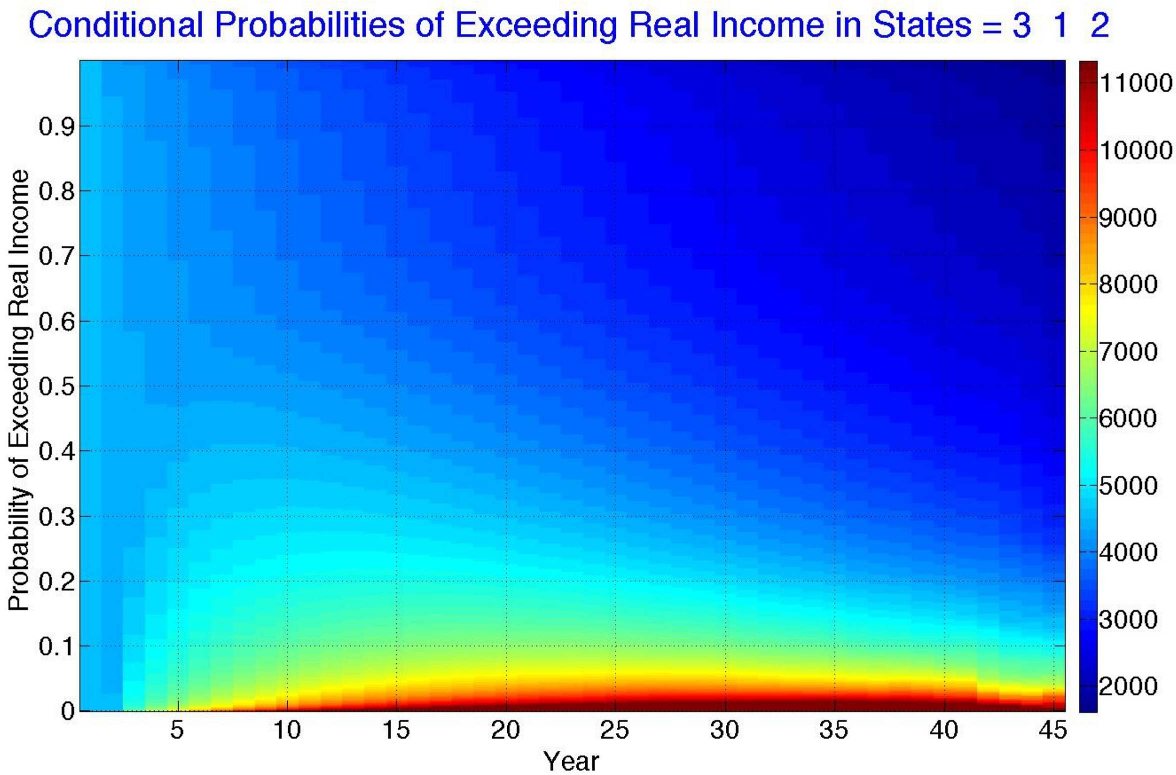

Recall that the colors for each year reflect the cumulative probabilities that income will exceed various levels. Here, the darker the color, the smaller is the associated income. In any year, larger incomes will plot lower in the graph, since there are lower probabilities of exceeding them. Here, in the early years the ranges of nominal incomes are smaller than in later years. Moreover, the ratcheting process leads to ranges in which there are greater chances of higher nominal incomes in later years.

But the map of real incomes tells a different story. To obtain it we set the data element to:

**analysis.plotIncomeMapsTypes = {'rc'};**

Again, we truncate all the incomes greater than 25% of the actual maximum income.

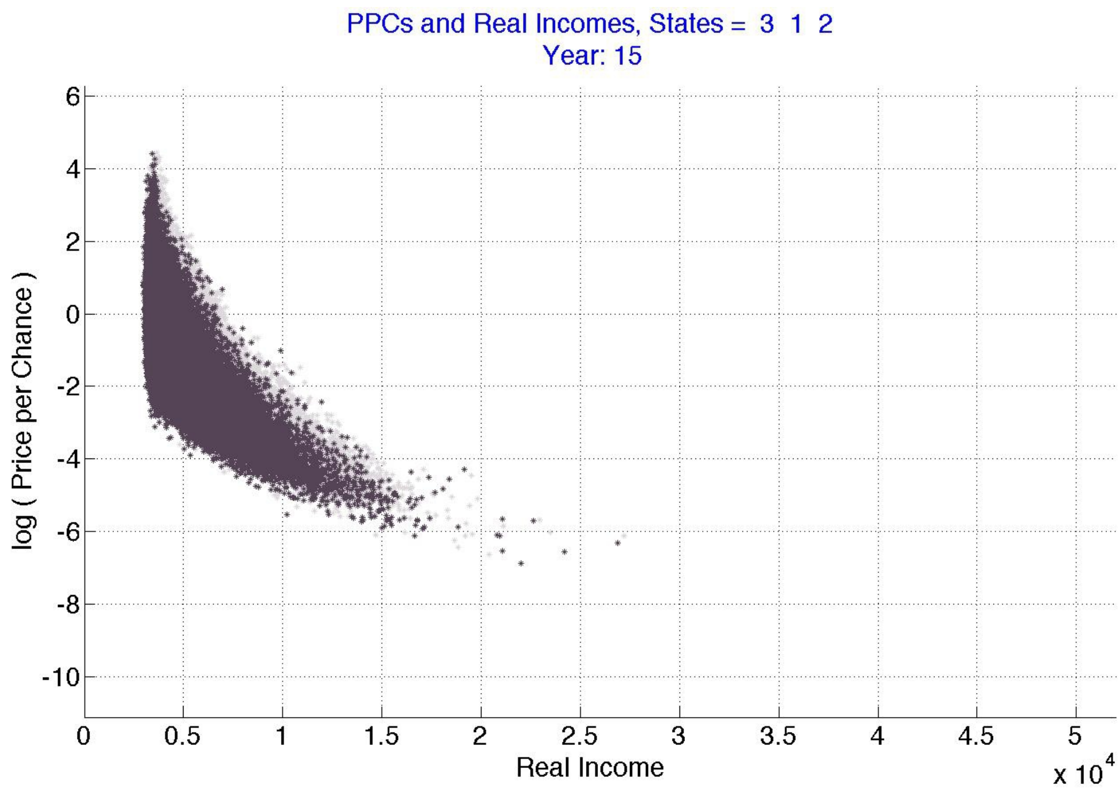
Here is the result:



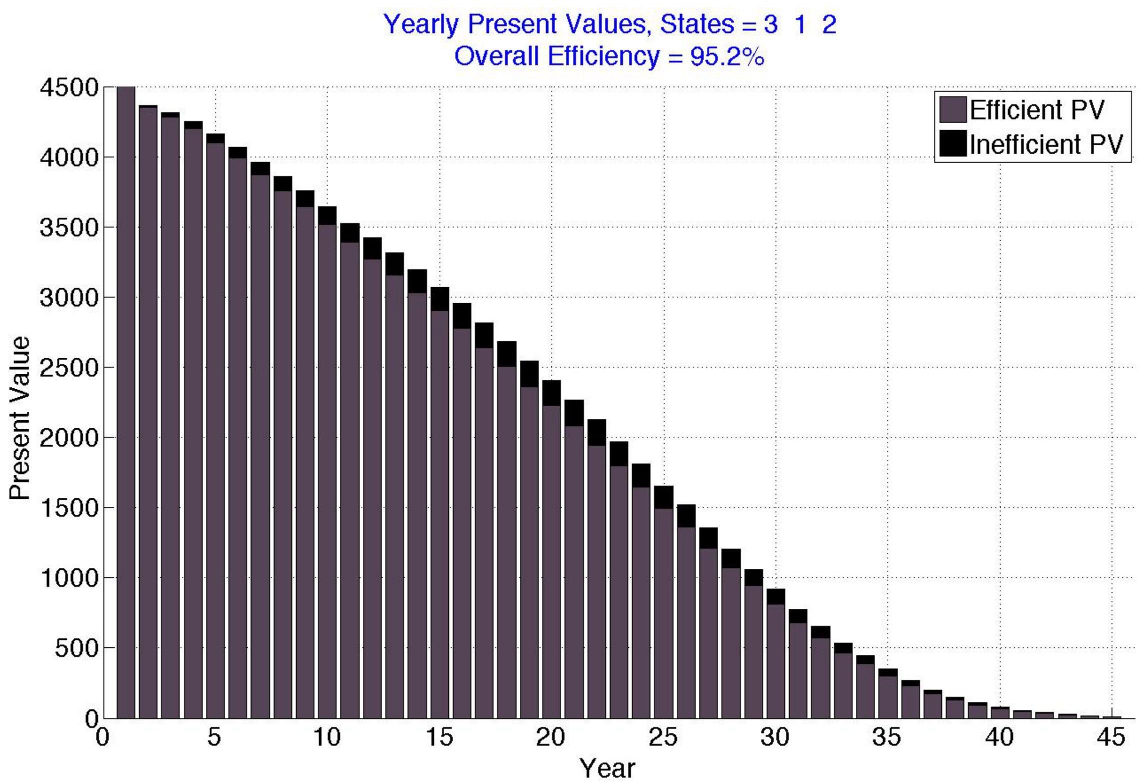

In the early years, the ranges of real incomes are smaller than in later years. And over time, the lowest real incomes tend to be considerably lower. This may be fine for the Smiths, but it is important that they focus, as here, on the likely purchasing power of their future income. For some investors, the appeal of approaches that provide ratcheted nominal income may be due in large part to *money illusion* – a term coined in the 1920's by the famous economist Irving Fisher. Here is the rather grand Wikipedia entry:

> *In economics, **money illusion**, or **price illusion**, refers to the tendency of people to think of currency in nominal, rather than real, terms. In other words, the numerical/face value (nominal value) of money is mistaken for its purchasing power(real value) at a previous point in the general price level (in the past). This is false, as modern fiat currencies have no intrinsic value and their real value is derived from all the underlying value systems in an economy, e.g., sound government, sound economics, sound education, sound legal system, sound defence, etc. The change in this real value over time is indicated by the change in the Consumer Price Index over time.*

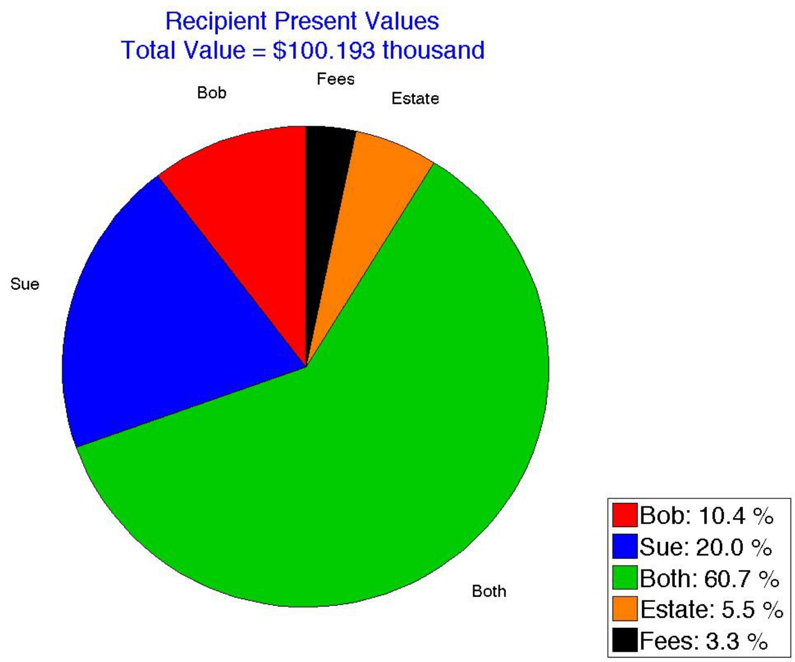
Not surprisingly, it is difficult to impute any clear set of marginal utility functions from the relationships between incomes and present values. Here is the graph showing the logarithms of Price Per Chance and Real Incomes up to and including year 15:



Despite the scatter of points in the PPC/Real Income graph, the cost-efficiencies of individual years' incomes are disappointing, but not abysmal. As the following figure shows, the overall efficiency is 95.2%, indicating that the same probability distributions of income could, in principle, be obtained by investing slightly more than 95% as much money in a cost-efficient manner.



Finally, we turn to the present values of the incomes going to the participants:




As usual, the total present value differs slightly from the amount invested due to sampling error, but it is very close to actual amount of  $100,000.

Given the tendency for real income to decline more often than it increases, it is not surprising that the present value of incomes provided when Bob and Sue are both alive is the largest. This is followed by the value of Sue's possible incomes while alone, then Bob's for the usual reason (Sue is younger and female and hence likely to outlive poor Bob).

Interestingly, the value of the estate is relatively small. This is also understandable. As we have seen, in a number of scenarios, the portfolio will become worthless while incomes are still required. In such cases there will be no estate left. Of course there are scenarios in which money will remain in the portfolio for the estate, but the net effect is to keep the present value to under 6% of the initial amount.

The rather remarkable aspect is the small amount of the original investment likely to go to the mutual fund provider and the writer of the income rider. As we showed earlier, in this case the present value of the income rider's possible cash flows is negative if the beneficiary does not take excess withdrawals, so here the present value of the possible cash flows to the fund and insurance company is roughly 3.3% of the initial investment. But recall that the income rider is most favorable for people like the Smiths. A separate analysis indicates that if they were both 70, over  8.75% of their savings would go to fees. And if they were both 75, fees would take over 13% of their hard-earned money.

Most of these graphs are included on a video available at:

http://www.stanford.edu/~wfsharpe/RISMAT/SmithCase_Chapter19.mp4

Due to the very small but positive possibility of a very large real income in some scenario and year, the script sets the analysis parameter for the maximum income to be shown to 50% in order to spread out the income distributions:

```
analysis.plotIncomeDistributionsPctMaxIncome = 50;
```

Otherwise, the analysis settings are standard.

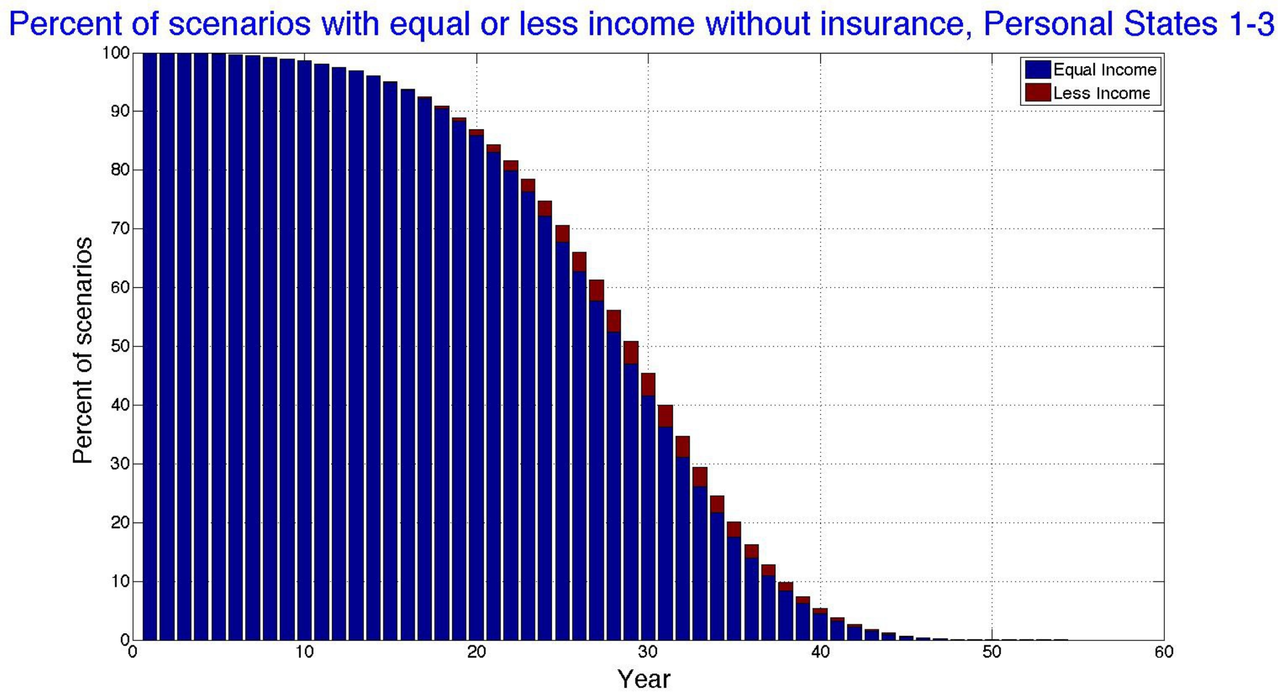
### *Homemade GLWBs*

Before leaving GLWBs, it is instructive to consider an alternative approach which attempts to provide incomes similar to those of a guaranteed benefit without purchasing a more expensive mutual fund plus a guaranteed lifetime withdrawal benefit insurance policy.

The goal of analyzing such a “homemade GLWB” is to calculate what would have happened if Bob and Sue had tried to obtain the same income in each scenario and year while alive as they would have received from an actual GLWB, doing so using a cheaper mutual fund and withdrawing only the income that the insured approach would have provided. In some scenarios, of course, the portfolio value would be exhausted and income would be smaller or zero. But in others, incomes would be the same and the amount left to the estate could be greater than would be provided by the insured approach.

For an example, we again assumed that all portfolio funds would be invested in our market portfolio. First we determined the incomes that would be provided using the Vanguard/Transamerica approach with a fund expense equal to 0.54% of the portfolio value each year and an insurance rider with an expense equal to 1.20% of the TWB value each year. We then analyzed the results that would be achieved if funds had been invested in the same portfolio, and the same incomes for the years in which Bob and/or Sue are alive deducted from portfolio values, until the portfolio values were zero or the estate received the remaining value. In effect, we assumed that each year Bob and Sue computed the results they would have obtained with the insured GLWB approach, then withdrew the resulting income or the remaining portfolio value, whichever was larger.

The figure below shows the results. In each year, the percent of scenarios with insured income is shown by the total height of the bar. The blue portion of each bar shows the percent of scenarios in a year in which the "homemade" approach would match the insured amount; the red portion shows the percent of scenarios in which the homemade approach would provide less income or none at all.




As can be seen, the danger of running out of money if the GLWB approach is simulated rather than adopted is relatively small. And the accompanying lower expenses lead to a significant possibility that Bob and Sue can leave a larger estate. After careful consideration, they might decide against purchasing a GLWB rider. And, of course, they might choose not to attempt to simulate a GLWB approach either.

Recall also that our previous graph showed that the costs of the Vanguard/Transamerica contracts are lowest for couples in which the younger is 65, as is Sue. Moreover, Bob is not too much older. For couples of other ages a similar analysis of results with actual versus simulated GLWB insurance could well make the latter appear even less attractive.

After careful consideration, Bob and Sue might decide against purchasing a GLWB rider on the grounds that a simulated approach would be preferable. But, it is entirely possible that they would prefer to obtain retirement income in some entirely different manner. In the remainder of this chapter we briefly discuss an alternative ratcheting approach. Then, in the next chapter, we consider a very different way to generate income.

## *Floor/Surplus Strategies*

Another approach to providing ratcheted income (that can increase but never decrease) was developed by Philip Dybvig for possible use by an endowment with a perpetual life. The first paper, published in the *Review of Economic Studies* in 1995, was titled "*Duesenberry's Ratcheting of Consumption: Optimal Dynamic Consumption and Investment in the Stock Market*." A second, in the *January/February 1999 Financial Analysts Journal,* "*Using Asset Allocation to Protect Spending*", is more pragmatic.

The basic idea is to divide an investment portfolio into two parts. The first, which Dybvig called the *committed account*, is invested in fixed income assets that will insure that the current level of spending can be continued forever. The second, the *discretionary account*, is invested in various assets, some or all of them risky. Depending on the performance of the latter, at some times in the future funds would be transferred from the discretionary account to the committed account, to be used to increase the annual amount of spending forever.

Dybvig's approach assumes an endowment with preferences characterized by time preference and constant relative risk aversion plus the additional requirement that spending never decrease. In his model, the optimal strategy depends on two preference parameters:  the pure rate of time discount, $\delta$, and the degree of relative risk aversion, *R*. He describes two possible ways to specify their values.

The first approach: *from simulation, "..* uses plots of annual spending from the endowment and its value ...(to give a) picture of the range of reasonable performance scenarios", since this "helps a user to internalize the implications of such a strategy." The user compares the ranges offered by alternative strategies, then picks the preferred one.

The second approach, *from underlying preferences*, is more theoretical. "The idea is to think about your preferences for different random and nonrandom outcomes, and to compute from these preferences what your values of $\delta$ and *R* must be. For example, you can figure out *R* by thinking about what increase for sure (3 percent? 4 percent) would be considered just as good as a 50/50 chance of increasing by 10 percent or getting no increase at all."

Dybvig recommends the first approach and is skeptical of the quality of choices coming from the second: "Most people do not find these parameters intuitively satisfying … The problem is that the sample questions used to elicit preferences are not similar to realistic questions in endowment management." Your author concurs.

In subsequent work, Jason Scott and John Watson adapted Dybvig's basic idea to obtain a similar approach for a retiree with a limited life. Their article in the September/October 2013 *Financial Analysts Journal* was titled "*The Floor-Leverage Rule for Retirement*." Like Dybvig, they propose two accounts – one that can provide a *floor* on the amount that could be spent in each future year (for an estimated life span, or until a deferred annuity begins to provide income), the other with *surplus* that can augment the floor account when appropriate, thus increasing spending in each subsequent year.

The primary example in Dybvig's 1999 paper for an endowment with an infinite life invests the discretionary account completely in stocks. Scott and Watson, considering beneficiaries with limited lives, propose that such an account be *levered (*for example, with the initial amount to be invested in stocks equal to three times the value of the discretionary account, financed by borrowing an amount equal to two times the value of the discretionary account ). Hence the title of their article.

We choose to call the two accounts in any such retirement income scheme the *floor* and the *surplus*. The floor is invested to provide an income that will remain constant for an intended number of years or life span. The surplus is invested in risky assets, possibly with leverage. According to a specified rule, funds may be transferred periodically from surplus account to the floor account, then used to augment all of the remaining guaranteed incomes.

A challenge for anyone wishing to adopt such a floor/leverage rule is finding a way to obtain significant leverage for the surplus account. Perhaps not surprisingly, financial engineering has provided a possible solution. In late 2016, there were two exchange traded funds (ETFs) designed to provide returns close to those of a 3X leveraged S&P500 fund: *ProShares UltraPro SP500* (UPRO) and *Direxion Daily SP500 Bull 3X shares* (SPXL). Each had an expense ratio of slightly less than 1.0% per year. At the time, the market value of UPRO was slightly over $700 million, while that of SPXL was under $500 million.

Here is an excerpt from UPRO's summary prospectus:

> *The Fund seeks daily investment results, before fees and expenses, that correspond to three times (3x) the daily performance of the Index. The Fund does not seek to achieve its stated investment objective over a period of time greater than a single day.*

The presumption is that the S&P500 will not fall more than 33.3% in a single day (which would wipe out the fund entirely). But this means that the *annual* return will not equal that of three times the index less borrowing costs and expenses, with the difference depending on daily variations in the index.

Both the UPRO and SPXL documents provide a table with some possibilities. The two tables are identical (except UPRO uses the term "One Year Volatility Rate" and SPXL simply "Volatility Rate" for the annualized standard deviation). Both appear to be derived from a theorem presented by R.A. Jarrow in "*Understanding the risk of leveraged ETFs*" published in *Finance Research Letters*, 7 (2010). Here is the SPXL version, which is more colorful:

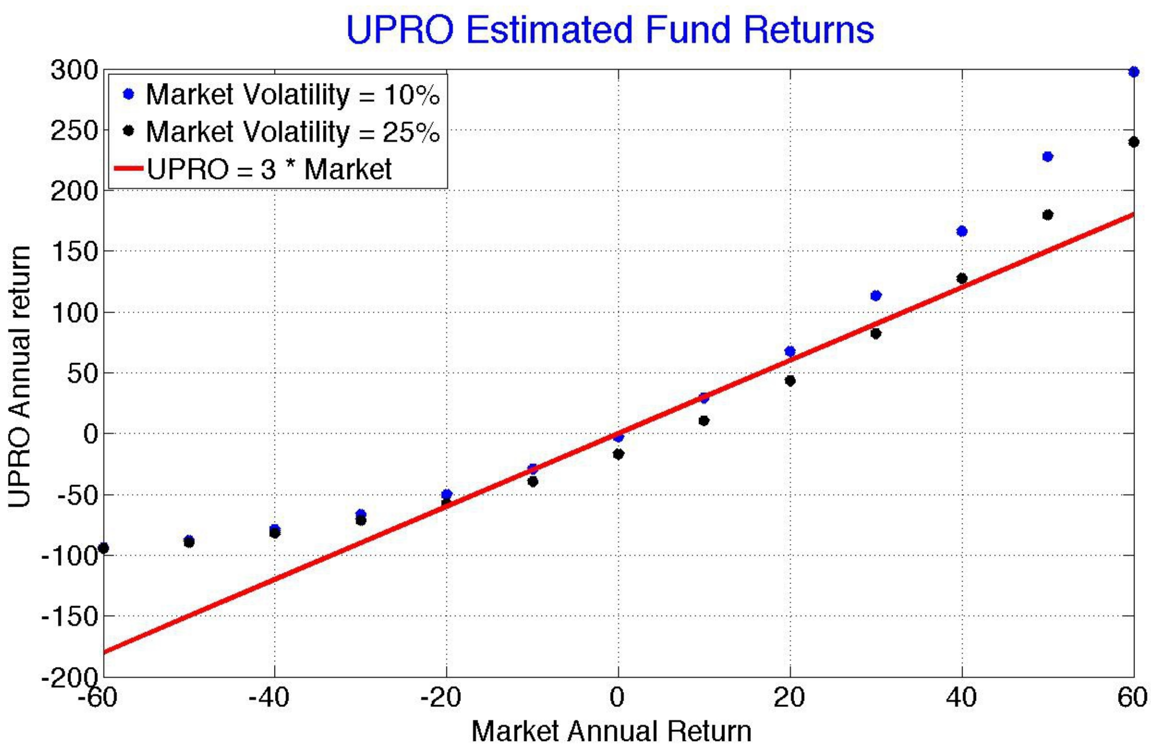
| One Year Index | 300% One Year Index | Volatility Rate | | | | |
|---|---|---|---|---|---|---|
| Return | Return | 10% | 25% | 50% | 75% | 100% |
| -60% | -180% | -93.8% | -94.7% | -97.0% | -98.8% | -99.7% |
| -50% | -150% | -87.9% | -89.6% | -94.1% | -97.7% | -99.4% |
| -40% | -120% | -79.0% | -82.1% | -89.8% | -96.0% | -98.9% |
| -30% | -90% | -66.7% | -71.6% | -83.8% | -93.7% | -98.3% |
| -20% | -60% | -50.3% | -57.6% | -75.8% | -90.5% | -97.5% |
| -10% | -30% | -29.3% | -39.6% | -65.6% | -86.5% | -96.4% |
| 0% | 0% | -3.0% | -17.1% | -52.8% | -81.5% | -95.0% |
| 10% | 30% | 29.2% | 10.3% | -37.1% | -75.4% | -93.4% |
| 20% | 60% | 67.7% | 43.3% | -18.4% | -68.0% | -91.4% |
| 30% | 90% | 113.2% | 82.1% | 3.8% | -59.4% | -89.1% |
| 40% | 120% | 166.3% | 127.5% | 29.6% | -49.2% | -86.3% |
| 50% | 150% | 227.5% | 179.8% | 59.4% | -37.6% | -83.2% |
| 60% | 180% | 297.5% | 239.6% | 93.5% | -24.2% | -79.6% |

As can be seen, the red boxes indicate possibilities in which the levered strategy does worse than 3 times the index return, and the green boxes those in which it does better.

UPRO provides the additional information that:

> *The Index's annualized historical volatility rate for the five-year period ended May 31, 2016 was 15.76%. The Index's highest May to May volatility rate during the five-year period was 23.25% (May 31, 2012).*

Here is a plot of the returns for the first two volatility levels in the chart, which may (or may not) bracket likely future risks:




Two aspects are relevant.

First, the relationship is clearly non-linear, with the fund returns likely to be greater than 3 times the market returns when the latter are very low or very high, and less than 3 times the market when market returns are within more normal ranges.

Second, there can be substantial differences in the annual fund return for any given level of annual market return, depending on whether or not variation in the daily returns is smaller or larger. Thus the annual performance of the UPRO shares will depend not only on the annual market return (the compounded value of the daily returns) but also on the variations along the path that index returns follow from day to day through the year. In our pricing model, this means that annual return and hence income will not be a function of only price per chance, resulting in some cost inefficiency (that is, the same distribution of income could be obtained at a lower cost).

The UPRO literature does not discuss borrowing money to obtain returns equal to 3 times those of the index. This raises the obvious question: how do they operate? The answer is given in the summary prospectus:

> *The Fund obtains investment exposure through derivatives. Investing in derivatives may be considered aggressive and may expose the Fund to greater risks than investing directly in the reference asset(s) underlying those derivatives. These risks include counterparty risk, liquidity risk and increased correlation risk*

At any given time the fund will have multiple agreements with other financial institutions (counterparties). Some of these counterparties may in turn may have agreements with yet other counterparties. And so on. With daily settlement, this may not be of great concern. But it is likely to incur additional implicit or explicit costs and may involve some further risks.

In addition to these issues and costs, there can be disparities between the net asset value of an exchange traded fund and the price at which its shares trade.

These issues covered, we return to the floor-leverage approach.

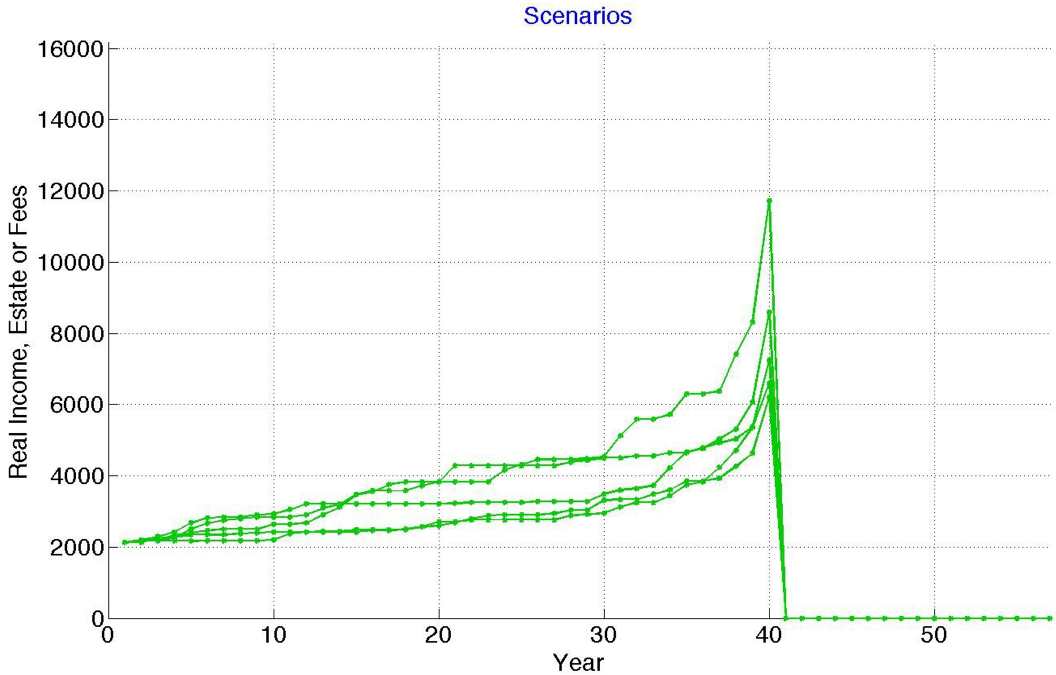
In their paper, Scott and Watson present a simple rule of thumb that closely approximates the theoretical optimal strategy for retirees with average risk tolerance. It involves three steps:

> 1. *85%—Build a riskless spending floor. At retirement, allocate 85% of available assets to purchase a sustainable lifetime spending floor. The floor type and resulting spending rate depend on the preferences of the retiree. Throughout retirement, money is withdrawn from the floor portfolio for spending.*
>
> 2. *15%—Invest in a 3× leveraged equity portfolio. The remaining 15% of assets, the surplus, is invested in a mutual fund or exchange-traded fund (ETF) that is rebalanced daily to maintain 3× leverage with the stock market. Together, the riskless floor (85%) and the surplus (15%), which is invested in a 3× leveraged ETF, provide downside protection and equity upside. Because all assets can be purchased and held between spending reviews, portfolio maintenance is minimal.*
>
> 3. *Annual spending review. Annually review the surplus portfolio. If it exceeds 15% of the total portfolio value, sell any surplus in excess of 15% and use the proceeds to purchase additional floor spending. Spending may increase but always remains sustainable; it ratchets.*

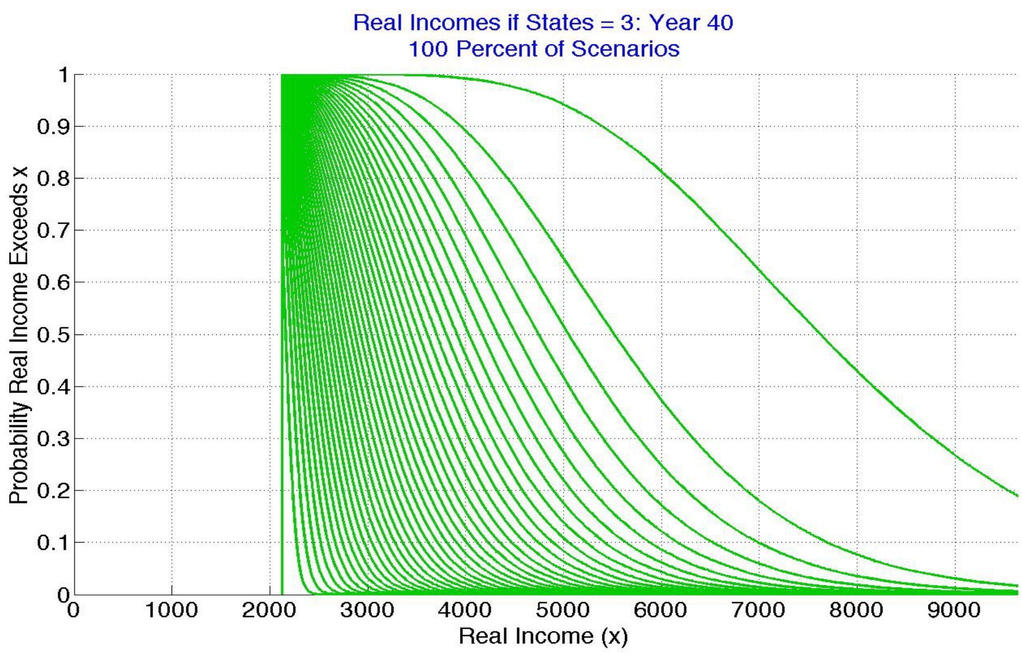
The paper then considers alternative implementations. Some focus on real incomes, others on nominal incomes. The initial floor and subsequent increments can be invested in a ladder of bonds designed to provide constant income until some future year (for example, 40 years from the initial date) or in annuities designed to generate constant income for remaining lives, and so on.

To illustrate some of the characteristics of the general approach, we consider a highly simplified case with a 40-year horizon. In each year, the floor is invested in a ladder of riskless bonds with zero real return after expenses. In the initial year, real income provided by the floor is equal to 1/40'th of its initial value. In the second year, real income from the floor is equal to 1/39'th of the value of the floor at the time, and so on until year 40 when the entire amount of the floor and surplus is spent. At the outset, 85% of the initial value is placed in the floor account and 15% in the surplus account. We assume that the surplus is invested to provide a return equal to 2 (rather than 3) times the market real return minus 1 times the riskless real rate (since with our standard assumptions it is almost impossible for the market portfolio to lose half its value in a year). At the end of each year, the value of the surplus account is compared with 15% of the total value of the two accounts, and any excess transferred to the floor account.

Here are five scenarios. As intended, real income can increase, but never decreases.

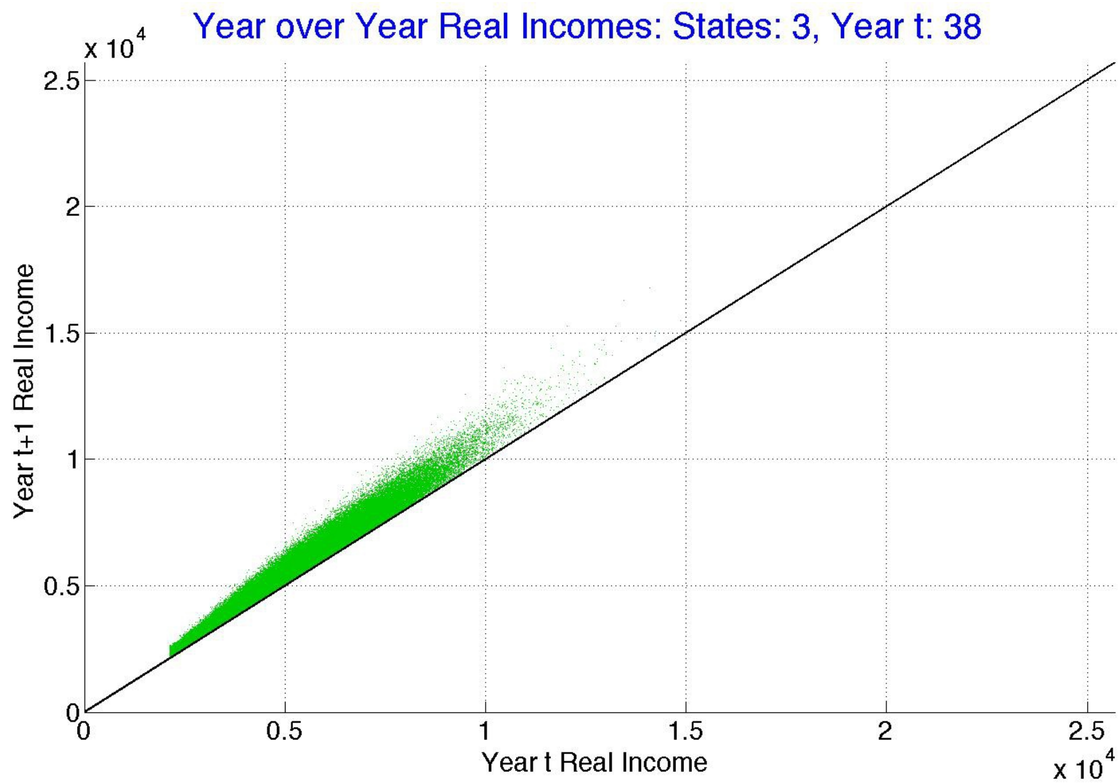



The probability distributions of real income reflect this attribute as well:

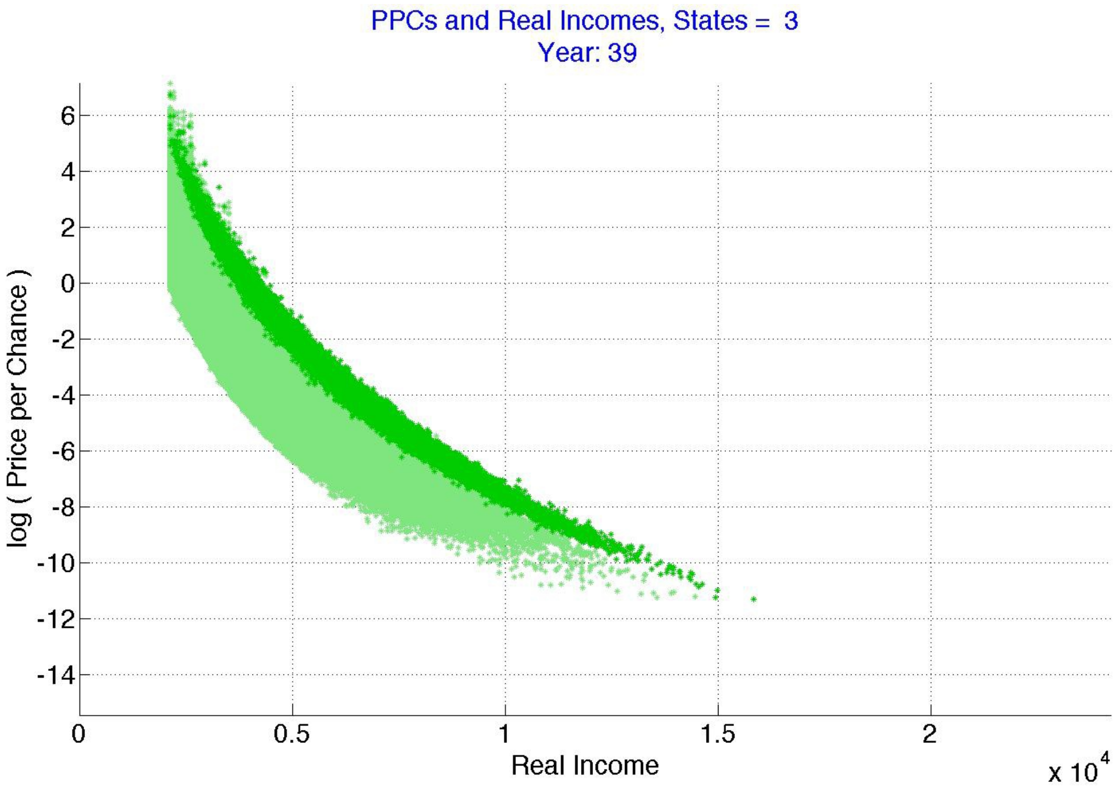



Each year's distribution plots on or to the right of that for the prior year. The last curve is considerably to the right of its predecessor, since we have assumed that the entire remaining surplus account is used to supplement income when the 40-year horizon is reached.

In the year-over-year income graph shown below, every point plots on or above the 45-degree line. Until the last year, the points lie quite close to the line. We stop at the data for years 38(t) and 39 (t+1) to not show the final disgorgement in year 40.

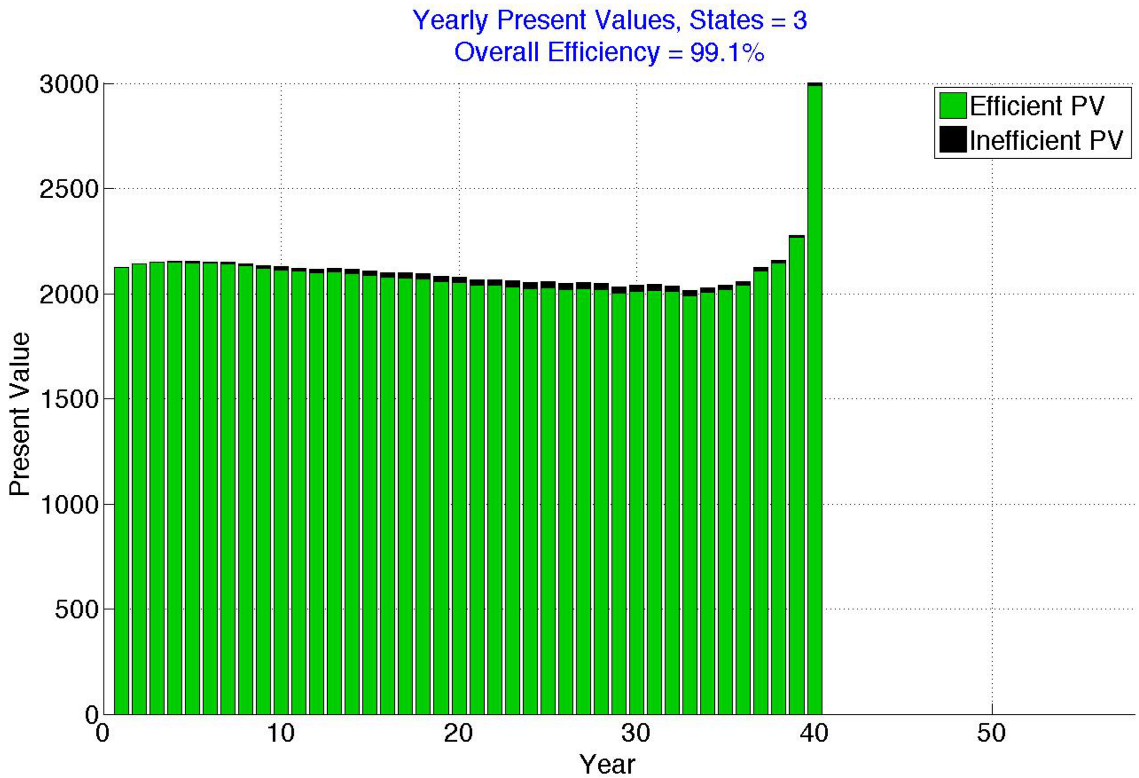


The plots of PPCs and Real incomes provide a sense of the implied marginal utility functions. Here are the results for the first 39 years, with those for year 39 in the darker shade:




Clearly, there is path-dependence. For example, the value of the portfolio after two full years will be comprised of two parts. The first will depend on the riskless return in each year., and the second on the market's total return in the first year times the riskless rate in second year. Thus there will not be a one-to-one correspondence between income and the cumulative market return over the two years. And since price per chance does have such a correspondence, real income will not be a one-to-one function of PPC. In future years, there will be multiple parts of total return, most of which will be the product of some years of riskless returns times the product of subsequent market returns. As indicated, the darker points in the figure show incomes for year 39. Despite the scatter, the overall relationship does reflect the degree of risk-aversion implicit in the choice to leverage the surplus.

While our floor/surplus strategy is not completely cost-efficient, the yearly present values graph below (based on the assumption that both Bob and Sue live for 40 years) shows that the degree of inefficiency is relatively small. The overall cost is 99.1% of the amount that could provide the same set of annual income probability distributions efficiently. But of course that strategy would not insure that income never falls from year to year.




Another aspect of the strategy is apparent in this graph. For an actual strategy, the present values of incomes would typically be smaller for later years than for earlier ones, reflecting the smaller probability that beneficiaries will be alive. Here no  account is taken of mortality and real incomes are likely to increase, to some extent offsetting the effects of lower present values for incomes farther in the future. And of course the final spike for the present value of incomes in year 40 reflects the expenditure of the entire surplus fund.

This example captures some of the aspects of floor/surplus approaches, but is far too simplified to represent practical applications. Our software does not have an all-stock portfolio, and valuations are annual rather than daily. Moreover, we do not attempt to forecast the terms on which immediate annuities might be offered in future years. We thus leave detailed analyses of actual floor/surplus strategies to others.

Perhaps not surprisingly, Financial Engines, where Jason Scott and John Watson did the initial research for such approaches, adopted some of the floor/surplus ideas (but apparently without leverage). This is from the company's 2015 annual report:

> *… the Income+ optimization approach divides the portfolio into three components. The first portion of the assets is used to structure a fixed income portfolio from the options in the plan that best match the duration of the income payments through age 85. A second portion of assets is set aside to enable the optional future purchase of an annuity outside of the plan that can maintain the income payments for life. Income+ allows participants to purchase such an annuity up to the age of 85. We do not provide any of these annuities or other financial products. Finally, a third portion of assets is invested in a diversified mix of equities to provide growth potential and to help the payouts keep up with inflation. Over time, the equities are gradually converted into additional fixed income assets to support a higher floor.*

Undoubtedly, other financial firms follow approaches incorporating some aspects of floor/surplus strategies. But as this is being written, details are not widely available.

## *Behavioral Economics*

Before concluding this chapter, it is useful to step back from the details of ratchet strategies to examine the underlying economics that drives them.

Wikipedia's entry for *behavioral economics* begins:

> ***Behavioral economics,*** *along with the related sub-field* ***behavioral finance****, studies the effects of psychological, social, cognitive, and emotional factors on the economic decisions of individuals and institutions and the consequences for market prices, returns, and resource allocation, ...*

Note that there are two aspects – how individuals behave and the impact of such behavior on aggregates such as market prices, returns, etc.. It is at least possible that some individuals make decisions that are inconsistent with traditional economists' view of rational behavior but that markets are influenced more by other individuals and institutions who behave more like the traditional *homo economicus*. And markets are not democratic – rich people have more influence than poor ones.

In 2013, Daniel Kahneman, a Nobel Prize winner and one of the giants of the field, published an encyclopedic book, *Thinking Fast and Slow* with experimental results, many documenting behavior not corresponding to traditional economic models of rational decision-making. Richard Thaler, another major figure in the field, is quoted by Wikipedia as saying "conventional economics assumes that people are highly-rational – super-rational – and unemotional. They can calculate like a computer and have no self-control problems." Needless to say, he chooses to disagree.

Our overall model of market behavior and the pricing of risky assets relies on a market dominated by individuals who behave in ways more or less consistent with "conventional economics". Moreover, most of the income strategies in this book are designed for retirees with preferences conforming with traditional notions of "rational behavior". The ratchet approaches in this chapter are exceptions, since they assume an extreme case of an "endowment effect" in which retirees will not accept any possibility of a decline in income, no matter what the cost in foregone opportunities might be.

We will not attempt a general discussion of this very important division within the economics and financial economics professions. But the following comments on the merits of floor/surplus approaches seem warranted.

The motivation for ratchets is the premise that it would be difficult or overly depressing for a retiree to reduce consumption. On the other hand, it would be nice to increase it, if possible. The proposed solution is to establish a floor of consumption per year that is less than could be obtained, holding back the remaining portion of wealth to invest in a surplus account. Then, from time to time, money can be taken from the surplus account to increase the floor income in every future year. After each such increase, the idea of a decline in income from the new floor is inconceivable. And so on...

This would seem to involve a certain amount of self-deception. The retiree presumably knows at the outset that there is a surplus and that in all likelihood,income will increase at some future time. But he or she is presumed to not be disappointed if such an increase doesn't happen for years, or is smaller than likely, etc.. More formally, utility functions are vertical to the left of each reference point, but the reference point increases when portfolio returns are sufficient, at which time the utility function becomes vertical at the newer, higher income.

This is not to say that there may well be a kink in one's utility function at the current level of income (as described in Chapter 9). If so, m-shares of the *up-flat-up* variety (as in Chapter 15) might appeal. But these would at best, relate only to a constant reference income level, not to a level that increases when surplus is utilized.

Another concern is that ratchet strategies may be based on nominal incomes rather than real incomes. As we have seen, this is the case for the type of GLWB approach covered earlier in the chapter. But this would seem to involve another type of self-delusion. Decreases in real income of varying magnitudes due to inflation are presumed to be acceptable, but any decrease due to poor investment returns is to be avoided at all costs.

Skeptics sometimes claim that behavioral economics is based on “choices made by undergraduates in psychology classes playing games for small or no rewards” (source unknown to this author). Perhaps. But there is no doubt that many humans often make choices that seem inconsistent with the kind of optimizing behavior assumed in many economic models. That said, it seems to this author that however elusive, the goal should be the design of income strategies for informed retirees based more on reason than emotion. The traditional admonition applies: more research is needed.

Having posed these important issues and examined some proposed solutions for a particular set of goals, we return in the next chapter to the subject of lockboxes, focusing on strategies that do not involve mortality pooling, at least for some initial number of years.