Financial Planning via Deep Reinforcement Learning AI

Gordon Irlam
(gordoni@gordoni.com)

June 23, 2018

Abstract

This paper introduces AIPlanner, a financial planner based upon deep reinforcement learning. AIPlanner provides an investment and consumption strategy intended to optimize lifetime well-being. The results of AIPlanner are very close to the precise analytical solution, as well as to the precise solution computed using stochastic dynamic programming. Deep reinforcement learning is additionally capable of delivering results for far more complicated and realistic financial models that other approaches can't handle. As an example of this capability, a bond model that includes both a yield curve and time varying interest rates is employed. Compared to other popular approaches, in one reasonable scenario, AIPlanner was found to effectively deliver approximately $1,000 to $8,000 of additional consumption per year.

1 Introduction

1.1 Machine learning

Machine learning is the branch of computer science that seeks to give computers the ability to learn, that is, to progressively improve, in their performance of a particular task.

1.2 Reinforcement learning

Reinforcement learning involves learning based on rewards received as a task is being performed. The goal of reinforcement learning is to maximize the total rewards received. Many reinforcement learning algorithms involve the iterative refinement of a policy function and a value function. The policy function specifies the action to employ on a given time step given the current observed state of the environment. The policy function can be refined if we know the value function, which specifies the cumulative expected future rewards from performing a particular action on a given time step when in a given state. Conversely the value function can be refined if we know the policy function specifying the action that will be performed on a subsequent time step.

1.3 Deep learning

Deep learning is machine learning that uses artificial neural networks containing two or more hidden layers. Neural networks are a very crude approximations to the role of neurons in the visual system. Neural networks typically comprise an input layer which receives the inputs, one or more hidden layers which use artificial neural units to process the data, and an output layer to deliver the results.

Neural networks are capable of approximating an arbitrary function that maps inputs to outputs. Typically large amounts of data are required to train neural networks. For visual processing tasks the depth of neural network may be anywhere from 20 to 200 layers, while for non-visual function approximation, the neural network is commonly far more modest, comprising perhaps only 2 hidden layers.

1.4 Deep reinforcement learning

Deep reinforcement learning uses neural networks for the policy and value functions. It has met with much success at playing Atari video games, Mnih et al. (2015). The action space of Atari video of games is discrete button presses. Financial planning primarily involves dealing with actions that involve continuous quantities, such as money, asset allocations, and bond durations. Deep reinforcement learning has also met with success in continuous action spaces as shown by its solving numerous multi-joint dynamics physics problems, Lillicrap et al. (2015).

At their heart all of these problems can be described by Markov Decision Processes (MDP): there is an environment which an agent is interacting with, the agent is able to fully or partially observe the environment, the agent receives rewards in response to its actions, and the agent seeks to maximize the expected total reward received.

1.5 The consumption and investment problem

An important problem in financial planning is to determine the consumption and investment strategy that optimizes lifetime consumption. This is capable of being formulated as a MDP. It thus seems plausible to try and use deep reinforcement learning to address this problem.

2 Description of AIPlanner

AIPlanner uses deep reinforcement learning to perform financial planning.

Calculations and results reported by AIPlanner are given in real (inflation-adjusted) terms rather than in nominal terms. Consumption and investment decisions are made once per year.

2.1 Interaction with financial environment

At its most general, the actions taken by AIPlanner at each time step comprise: the amount to consume, the amount of real (inflation-indexed) Single Premium Immediate Annuities (SPIAs) to purchase, the amount of nominal (non-inflation-indexed) SPIAs to purchase, the asset allocation to use, the duration to use for real bonds, and the duration to use for nominal bonds.

Observation of the current state of the financial environment by AIPlanner comprises the following variables: life expectancy, real guaranteed income (Social Security, inflation-indexed pensions and SPIAs), nominal guaranteed income (nominal pensions and SPIAs), portfolio size, the short term real interest rate, and the short term inflation rate.

2.2 Action mapping

By default the deep reinforcement learning algorithm I use generates actions in the range $(-\infty, \infty)$ with a roughly normal, $\mathcal{N}(0,1)$ distribution. For the amount to consume I originally performed a tanh mapping of the generated action onto the interval $(-1, 1)$ followed by an affine mapping onto the interval $(0, 1)$. I used the resulting value to represent the fraction of the available assets to consume. Unfortunately, under this scenario, the initial mean consumption fraction is 0.5, and so the portfolio is initially depleted in the first few years. Gradually the agent learns the appropriate consumption fraction is much lower, but training the agent to recognize the value of poor consumption decisions comes at a cost to the accuracy of more appropriate consumption decisions. As a result the performance of the trained agent is good, but not as good as possible. To overcome this problem, for each consumption action, I first come up with a rough estimate for a reasonable consumption fraction, f, which is computed as the the investment portfolio size divided by the life expectancy plus the current guaranteed income. I then map the original generated action onto the interval $(0, 2.f)$. This has the desirable property that on average the initial recommendation is f, and so the agent only learns more appropriate actions.

Similar scaling tricks are played with each of the other values to learn so that they are both within the appropriate range, and display a learning appropriate standard deviation.

For the fraction of wealth to annuitize what is important to learn is not how much to annuitize each year, but the fraction of total wealth that should be in the form of guaranteed income. Learning the latter means the model will give good results for an investor whose guaranteed income fraction is a long way from the more optimal scenarios generated during training.

2.3 Training

There are many deep reinforcement learning algorithms to choose from. AIPlanner is currently using the Proximal Policy Optimization (PPO) algorithm, Schulman et al. (2017). Both the policy and value functions are described by neural networks with 2 hidden layers. Each hidden layer containing 64 neural network units. The layers are fully connected, meaning that every unit in layer 1 is connected to every unit in layer 2. Standard hyper-parameter values for the training of the neural network are used, except for the temporal discount factor, which is set to 1, instead of 0.99. The temporal discount factor could have been used to impose a preference for receiving rewards earlier in life.

AIPlanner is trained by executing the PPO algorithm for 1 million simulated time-steps, that is, 1 million simulated years of investing. At the outset AIPlanner knows very little about investing, and makes all sorts of mistakes, such as consuming most of the portfolio in the first year. Through a process of trial and error, and iterative refinement of the policy and value functions, AIPlanner gradually becomes a master investor.

Each annual consumption amount is assigned a value, the utility, reflecting how desirable it is to consume that amount. Because an incremental dollar brings decreasing additional happiness as the level of consumption increases, the mapping from consumption to utility is not linear, but is closer to being logarithmic. Utility maps directly to the reward received by the PPO algorithm.

Life expectancy is potentially stochastic. One approach would be to train the model with different random life expectancies for each episode. This approach would be inefficient. To maximize the benefits of training, the financial model is simulated through to a fixed maximum possible age of 120 years. The utility of consumption for each age assuming the individual is alive is then weighted by the probability of the individual actually being alive at that age. In this way a single training episode contains the results of simulating all possible life expectancies.

AIPlanner was trained separately for each of the different results reported in this paper.

2.4 Evaluation

Monte Carlo simulation is used to assess the performance of AIPlanner. For each assessment 2 million simulated years of investing are performed against the trained model. Such a long assessment period is necessary on account of the large variability of investment returns. After 2 million simulated years, the standard error in the measured performance assessment of AIPlanner is around 0.1%. The performance assessment metric is certainty equivalent consumption. The certainty equivalent consumption is the constant consumption amount which has the same average utility as the consumption sequence actually experienced. Certainty equivalents are explained in detail in Appendix A.

The precise performance of AIPlanner depends upon the initial state of the neural network being trained. To get a handle on the variability in performance for each scenario 10 neural networks are trained, each starting from a different random initial state. Results are reported for the neural network that trained the best, that is, the neural network with the highest certainty equivalent score.

2.5 Computational performance

Both training and evaluation each take around 20 minutes on a current generation (2018) server. Apart from training and evaluating each of the 10 neural networks in parallel, there is little performance gain from using a multi-core machine. Somewhat unexpectedly, using a GPU was found to result in longer training and evaluation times. Perhaps GPUs only accelerate performance for many layers deep neural networks.

Once trained AIPlanner is able to retrieve the recommended strategy for a given situation very rapidly; in about 350 microseconds for a single recommendation, or 2 microseconds per recommendation for a large batch of recommendations. The deployment model than is to train AIPlanner offline, then to take the trained model and to make it available for online interaction.

When performing Monte Carlo simulation the performance bottleneck is not determination of the recommended strategy, but updating the financial model in response. It takes 50 microseconds to update the financial model in the absence of SPIAs or 240 microseconds if it is necessary to compute the prices of both real and nominal SPIAs.

3 Description of the financial model

The model used by AIPlanner is somewhat simplified. The effect of taxation is ignored. There is no emergency fund for contingent expenses. Investment expenses and transaction costs are ignored. There is no reason why these factors couldn't be incorporated into AIPlanner in the future.

Mortality, consumption, and investment decisions, including rebalancing the portfolio to the new target asset allocation, occur on a periodic annual basis.

Leverage, that is borrowing against one asset class to buy more of another, was not allowed.

No attempt was made to model the standard error associated with each of the returns estimates. This omission is intentional. AIPlanner is capable of taking the standard error into account. Its omission makes it easier to validate the results against other methods which fail to consider the effect of the standard error.

3.1 Mortality

The financial model used by AIPlanner describes a retired female. Their current age is 65. They were born in 1955. They have a mortality as specified by U.S. Social Security Administration cohort life tables Actuarial Study No. 120. Since potential financial planning clients are typically in better health than the average retiree, the q-values (probability of death) were multiplicatively adjusted by a constant factor. This was done to increase the life expectancy over that of the Social Security tables by 3 years. The client's total life expectancy at age 65 is thus 88.5 years.

3.2 Wealth

The client starts with $16,000 of real guaranteed income per year. This matches the average annual Social Security retirement benefit in 2018. The client has $500,000 in investable assets. This is a relatively high value, but it is not unreasonable for a financial planning client. The client has no other assets or liabilities.

3.3 Utility

Utility is determined using a constant relative risk aversion utility function, with a coefficient of relative risk aversion of 3. This is on the higher end of what strikes me as plausible. However, to justify observed asset allocations, some believe the coefficient of relative risk aversion should be significantly higher than this.

The client has an unlimited psychological ability to handle changes in nominal investment portfolio value. That is they have a very high risk tolerance. The client's goal is simply to maximize expected mean utility. Assuming a high risk tolerance makes it easier to compare the results to other approaches.

3.4 Stocks

Market parameters are derived from World and U.S. averages reported in the Credit Suisse yearbook over the period 1900-2016, Dimson et al. (2017). Returns are assumed to be log-normally distributed.

Stocks are modeled as having an arithmetic real return of 5.0%, and a volatility of 17.4%, resulting in a geometric real return of 3.6%. This annual return is 1.5% less than the global average annual return reported in the Credit Suisse yearbook. This was done for pedagogical reasons. With a higher real return, the recommended asset allocations are frequently close to 100% stocks making model validation less meaningful. Alternatively the chosen returns could be justified on the grounds that the stock market is currently somewhat over valued, and this over valuation portends lower future returns.

3.5 Bonds

I model several different types of bonds.

3.5.1 IID bonds

Independent Identically Distributed (IID) nominal bonds are the simplest. They are intended to model long term government bonds in the Credit Suisse yearbook. Arithmetic annual real return is 2.4%, and volatility is 11.2%.

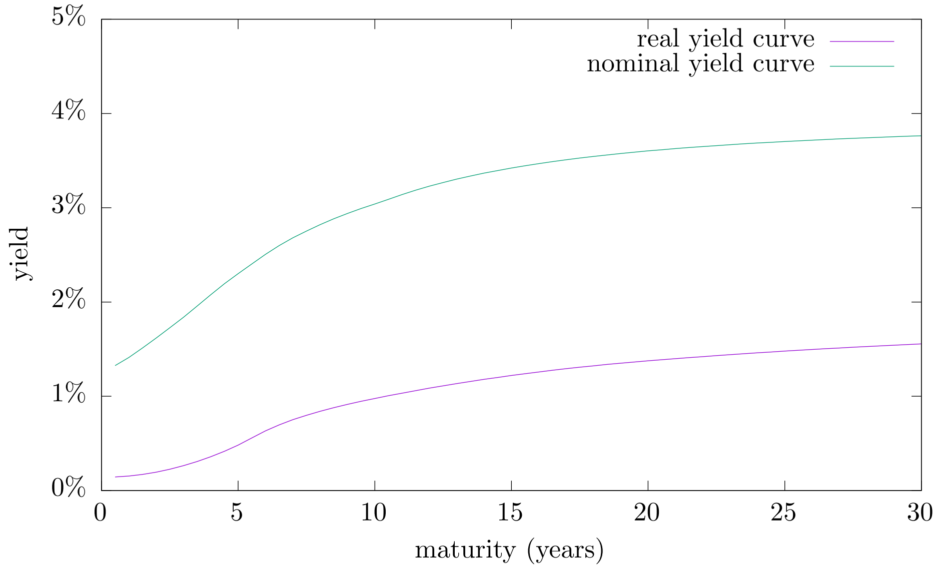

Figure 1: Typical modeled real and nominal yield curves.

3.5.2 Real bonds

For a more sophisticated bond model I use the single factor Hull-White model with constant parameters α and σ and time-varying parameter θ. This model provides a yield curve, and allows the price of bonds to vary over time, reflecting stochastic variability in the interest rate. Real bonds have a typical yield curve reflecting the average real yield curve reported by the Treasury over the period 2005-2017. The parameters of the Hull-White model were calibrated to produce a short term real yield volatility similar to what was observed over this period, and a real return volatility of about 7%. This later figure was arrived at by a simple scaling of the observed ratio of the volatility of real bonds to nominal bonds against the reported real nominal bond volatility in the Credit Suisse yearbook.

Typical yield curves for the modeled real and nominal bonds are shown in Figure 1. The shape of the actual experienced yield curves will differ, depending on both the typical yield curve and the short interest rate in accordance with the Hull-White model.

Bonds are modeled as zero coupon bonds having a single duration with a maximum permitted duration of 30 years. To first order such zero coupon bonds do a good job of matching more complicated bond funds composed of multiple bonds having different maturities, provided the duration of the bond fund matches the duration of the zero coupon bonds. When warranted AIPlanner was allowed to vary and select the best bond duration for each time step.

3.5.3 Nominal bonds

Nominal bonds were modeled by modeling inflation using the Hull-White model, and deflating the real bond price by the price of inflation to produce the nominal bond price. The nominal bond inflation risk premium was assumed to exactly offset the real bond ill-liquidity premium, so that expected inflation equals the difference between real and nominal yields. Inflation was tuned to produce a typical nominal yield curve matching the nominal Treasury yield curve, and nominal bond return volatility roughly matching the Credit Suisse yearbook.

	Duration	Mean yield	Real return	Volatility
Intermediate-term real bonds	5	0.5%	1.0%	3.8%
Long-term real bonds	15	1.2%	2.1%	7.1%
Expected long term inflation	-	3.1%	-	-
Intermediate-term nominal bonds	5	3.6%	1.5%	5.7%
Long-term nominal bonds	15	4.3%	2.4%	11.2%

Table 1: Measured bond and inflation parameters.

The modeled bond parameters are more fully described in Table 1. The arithmetic mean real return and volatility of long-term nominal bonds match the reported values for long-term government bonds in the Credit Suisse yearbook. That volatility matches is by design. That mean return matches is by happy coincidence.

As mentioned, the inflation model is tuned to match the nominal bond yield curve and returns to observed values. No attempt was made to prevent deflation, and the observed standard deviation in the inflation rate is high. The annual volatility in the expected short term inflation rate is 1.3%, matching the observed value over the period 2005-2017. Given this volatility, it is difficult to constrain the expected short term inflation rate to a standard deviation of just 1.1% as was observed during this period. Instead the observed standard deviation of the expected short term inflation rate is 2.5%, and the short term inflation rate is seen to meander between about -1% and 6%. Longer term average inflation rates are more constrained. This inflation volatility is more indicative of the data reported in the Credit Suisse yearbook, which reported average U.S. inflation over 1900-2016 as 3.0% with a standard deviation of 4.7%, than it is of 2005-2017.

3.6 Risk free asset

When required a risk free asset was modeled as having a real return of 0.8% and no volatility. This matches the geometric real return on U.S. Treasury bills reported by the Credit Suisse yearbook.

3.7 SPIAs

Real SPIAs are available for purchase at any age with a Money's Worth Ratio (MWR) relative to Treasury instruments of 94%. Nominal SPIAs relative to U.S. Treasuries have a MWR of 100%. These prices are intended to be reflective of the U.S. annuity market after factoring, say, the 2.35% California annuity guarantee association state tax. Prices for nominal annuities in the U.S. market probably exceed 100% when priced against Treasury bonds, but are below 100% when priced against the High Quality Markets corporate bond yield curve. There is no such thing as a free lunch; a MWR above 100% implies the issuers and thus the annuitants are probably taking on some default risk, but this is not the topic of this paper.

SPIAs in the real world only make sense for someone in good health, and reflecting this are priced using an annuitant life table with contract age adjustments applied. Rather than modeling this complication I used as the SPIA pricing mortality table the actual mortality table of the client. Results should carry over in the sense that if AIPlanner recommends purchasing a SPIA to the simulated client with the simulated life table, in the real world a SPIA would probably be recommended to an actual client providing their mortality matches or is below that of the annuitant life table.

In my model the yield curve is time varying. On account of this the price of SPIAs are recomputed for every

	Merton	Samuelson	AIPlanner
Initial consumption	\$19,887	\$19,767	\$18,374
Initial stocks fraction	50.2%	-	62.3%
Measured certainty equivalent	\$21,513	\$21,514	\$21,325

Table 2: Comparison of AIPlanner to the analytical results of Merton (continuous-time) and Samuelson (discrete-time). Discrete time values are given for a 1 year period.

period.

4 AIPlanner vs. other approaches

4.1 AIPlanner vs. The analytical solution

Given a risk free asset and a risky, or volatile asset, Merton's portfolio problem involves determining the asset allocation and consumption that will maximize lifetime utility. For log-normally distributed returns, constant relative risk aversion, no guaranteed income, and a fixed finite or infinite lifespan, Merton's portfolio problem has an analytical solution. This was shown by Merton (1969) in the continuous time case. Samuelson (1969) studied the discrete time case, and determined the optimal consumption for that case.

I consider the annual optimal consumption rate, asset allocation, and estimated certainty equivalent for a portfolio initially comprising \$500,000 of stocks and the risk free asset, without any guaranteed income, and a lifespan of precisely 30 years. I do this for the continuous time case, the discrete time case, and AIPlanner, as shown in Table 2. There is some discrepancy, but the important thing to note is that the certainty equivalent of AIPlanner is within 1% of the best theoretical possible value. So long as the certainty equivalent is close, it doesn't matter that the recommended consumption or asset allocation may differ.

Note that certainty equivalent amounts reported here are a good deal lower than reported elsewhere in this paper due to Merton and Samuelson's portfolio problem being unable to handle any guaranteed income. Also note that Merton's consumption amount is an instantaneous consumption rate, not consumption over a period, and Merton's asset allocation assumes assets are being continuously rebalanced rather than rebalanced once per year.

4.2 AIPlanner vs. Stochastic dynamic programming

Stochastic dynamic programming is a technique for solving problems involving uncertainty by working backwards in time and building up from the solution in the future to the solution in the present. Stochastic dynamic programming can be used to computationally precisely determine the optimal consumption and investment strategy. Unfortunately, stochastic dynamic programming suffers from the curse of dimensionality, making it computationally infeasible to solve models with more than a few variables.

I developed research software called Opal, that performs stochastic dynamic programming to solve the consumption and investment problem in simple cases, Irlam and Tomlinson (2014). Table 3 compares the results of AIPlanner to the results obtained using Opal with a stochastic life span, \$16,000 a year of real guaranteed income, and initially \$500,000 to invest between stocks and IID bonds. The number to focus on is the certainty equivalent, which for AIPlanner is 99.6% of the best possible value computed using stochastic dynamic programming. That the initial consumption and stocks fraction do not match exactly, but the certainty equivalent comes very close, is testimony to the insensitivity of the optimization problem to small

	SDP	AIPlanner
Initial consumption	\$38,299	\$39,189
Initial stocks fraction	69.8%	72.1%
Measured certainty equivalent	\$37,575	\$37,442

Table 3: Comparison of AIPlanner to the results of stochastic dynamic programming computed using Opal.

	SDP	AIPlanner
Initial consumption	\$41,450	\$40,853
Initial SPIA purchase	\$168,776	\$115,709
Initial stocks fraction	89.5%	80.1%
Measured certainty equivalent	\$42,980	\$42,834

Table 4: Comparison of AIPlanner to the results of stochastic dynamic programming computed using Opal in the presence of real SPIAs.

changes in strategy. The benefits of consuming slightly more in the first year are balanced out by the costs of consuming slightly less in subsequent years.

Later on I will examine the results for AIPlanner with real and nominal SPIAs. Opal contains a limited ability to model the optimal strategy with SPIAs, Irlam (2014). It is instructive to compare AIPlanner to Opal in the presence of SPIAs. Opal does not allow the SPIA lifetable to be adjusted to provide for additional life expectancy, so in this comparison the standard Social Security Administration cohort mortality was used for both the client's life expectancy and for constructing SPIA prices. Opal also requires the annuity pricing yield curve be static, and the bond returns be IID. These changes were made to AIPlanner. Table 4 compares AIPlanner to Opal using real SPIAs for this scenario.

Opal self validates by comparing the predicted certainty equivalent expected from SDP against the certainty equivalent computed using its recommended, supposedly optimal, strategy. In this case the self validation failed, producing a certainty equivalent of \$42,968 with a standard error of \$5, which with 95% confidence lies just outside the \$42,980 expected by Opal. This might be due to chance, although another possibility is that the grid which Opal used to perform the calculations wasn't fine enough. Unfortunately, attempting to use a finer grid failed due to computational instabilities.

When SPIAs are present, AIPlanner delivered 99.7% of the optimal performance.

4.3 AIPlanner vs. Common investing rules

Financial planners employ many different rules for determining consumption and investment strategies. Pfau (2015) contains a review of various consumption rules. The consumption rules I will consider are listed below.

- Perhaps the oldest withdrawal rule is the 4% rule, Bengen (1994). It involves consuming a constant inflation-adjusted amount each year, set at a fixed 4% percent of the initial investment assets. I adjust the 4% rule for the presence of guaranteed income by additionally consuming the full amount of guaranteed income received each year.
- The 4% rule generalizes to the optimal constant real amount strategy, in which the constant amount has been chosen to maximize the reported certainty equivalent.
- Guyton's Rule 2, Guyton (2004), specifies consuming the same nominal amount as the prior year if

the investment portfolio's nominal returns are negative, otherwise it specifies consuming the same real amount as the prior year. I apply Guyton's Rule 2 to the investment portfolio income, excluding guaranteed income. Guaranteed income gets added to the amount to consume later on. The initial consumption amount for Guyton's Rule 2 was chosen to maximize the reported certainty equivalent.

- Guyton-Klinger, Guyton and Klinger (2006), specifies increasing consumption by 10% if the investment consumption rate falls below 80% of the initial investment consumption rate; decreasing consumption by 10% if the investment consumption rate exceeds 120% of the initial investment consumption rate and there are more than 15 years to live; and not applying an inflation adjustment in years of negative nominal returns for which the investment consumption rate exceeds the initial investment consumption rate. Guyton-Klinger use a 40 year planning horizon. The initial consumption rate was chosen to maximize the certainty equivalent.

- Zolt (2013) describes Target Percentage Adjustment. This rule specifies not applying the inflation adjustment to investment portfolio withdrawals in years where the investment portfolio has fallen below an expected portfolio value. The expected portfolio value is computed assuming a constant real portfolio depletion rate using a maximum life expectancy and an expected portfolio rate of return. Zolt used 45 years as the maximum life expectancy. However, guaranteed income acts as a buffer preventing the worst effects of portfolio depletion. For this reason, I use a shorter maximum life expectancy of 30 years. For the return assumption I use the expected yield on nominal intermediate duration bonds, 0.8% real, or 3.8% nominal. Once again the initial consumption amount is chosen to maximize the certainty equivalent.

- Target Percentage Adjustment with Dynamic Life Expectancy describes using Target Percentage Adjustment with the actual initial life expectancy for the maximum life expectancy along with adjusting the expected portfolio value based on the then life expectancy.

- PMT describes using the then payout required to deplete the investment portfolio over a specified remaining duration assuming a specified rate of return. It is named after the PMT() function in Excel. On the Bogleheads discussion forum PMT is referred to as Variable Percentage Withdrawal. I use 30 years as the initial duration, and the intermediate duration nominal bond yield as the expected return.

- PMT with Dynamic Life Expectancy describes using the PMT formula with the current life expectancy as the remaining duration.

Asset allocation rules are less varied. The two most popular options appear to be using a fixed asset allocation, and using a stock allocation that declines with age. Based on observed asset allocations elsewhere in this paper I chose to use fixed 50%, 75%, and 100% stocks, with the remainder in intermediate duration nominal bonds. Declining asset allocation rules, such as age-in-bonds, were expected to perform poorly on account of their low stock allocations, and so were not evaluated.

Table 5 compares the performance of AIPlanner to the above strategies. The strategies in which consumption is fixed apart from possible inflation adjustments perform poorly. Such strategies fail to adjust consumption to reflect the current portfolio size. Guyton-Klinger, Target Percentage Adjustment, and PMT perform better, especially when PMT is combined with dynamic life expectancy. AIPlanner comfortably outperforms all the tested rules, including PMT. All the consumption rules are silent on the purchase of SPIAs, however AIPlanner is capable of making such a recommendation, so for reference the performance of AIPlanner is shown both with and without this feature enabled. Very roughly, for the chosen scenario, AIPlanner outperforms common investing rules by anywhere from \$1,000 to \$8,000 per year.

4.4 AIPlanner vs. Financial planning software

Comparing AIPlanner to existing financial planning software is difficult. To do a fair comparison it is necessary for AIPlanner and the financial planning software to share the same world model. This means

Asset allocation (stocks/bonds)	50/50	75/25	100/0
4% rule	$33,397	$32,766	$31,846
Constant real amount	$33,436	$32,835	$31,892
Guyton's Rule 2	$33,691	$32,968	$31,912
Guyton-Klinger	$34,443	$35,076	$35,041
Target Percentage Adjustment 30	$34,275	$33,740	$32,717
TPA with Dynamic Life Expectancy	$34,286	$33,745	$32,716
PMT 30	$33,758	$34,034	$33,868
PMT with Dynamic Life Expectancy	$35,931	$36,641	$36,650
Asset Allocation	dynamic		
AIPlanner without SPIAs	$37,532		
AIPlanner with nominal SPIAs	$41,139		

Table 5: Certainty equivalents for AIPlanner and various consumption and asset allocation rules using stocks and 5 year duration nominal bonds.

	Certainty equivalent
MaxiFi Planner's no SPIAs recommendation	< $35,973
MaxiFi Planner's 100% real SPIA recommendation	$39,171
AIPlanner with real and nominal SPIAs	$41,179

Table 6: Very tentative comparison of AIPlanner to financial planning software using the AIPlanner world model.

the mortality and return assumptions must match. But it means more than this. Apart from the optional bond and SPIA pricing models, AIPlanner currently has a very simple world view. On the other hand most financial planning software has a simple bond model, utilizing stochastic asset class returns, and may not consider the possibility of utilizing SPIAs, but have a more feature rich world view, with features such as estimating taxes, Medicare Part B premiums, and Social Security claiming strategies.

One piece of financial planning software that does include the possibility of utilizing SPIAs is MaxiFi Planner by Economic Security Planning. Whether MaxiFi Planner recommends purchasing SPIAs depends on the assumed nominal bond yield in their SPIA pricing model. By default MaxiFi Planner assumes a 5% nominal bond yield. This is a very favorable rate compared to what is available today. Setting the yield to 2.8% nominal to more closely match yields assumed by AIPlanner assuming 2% inflation in MaxiFi Planner, MaxiFi Planner did not recommend the purchase of any SPIAs, and recommended a time varying annual consumption of between $35,874 and $35,973. On the other hand using the default 5% nominal yield assumption MaxiFi Planner recommended an immediate 100% real SPIA purchase. These recommendations are for a 65 year old female with a fixed life expectancy of 95, $500,000 of assets and $16,000 of Social Security. The return assumptions used were a fixed 5% and 5.95% nominal total portfolio return. The payout of MaxiFi Planner's SPIA purchase strategy in the AIPlanner world model was computed. Assuming the world models are sufficiently similar it is thus possible to compare AIPlanner to MaxiFi Planner as shown in Table 6.

MaxiFi Planner's no SPIA recommendation results are particularly hard to compare to AIPlanner, so little should be read in to them. AIPlanner assumes a stochastic life expectancy with a total life expectancy of 88.5, but there is a chance of living all the way to age 119. MaxiFi Planner's is only able to handle a fixed lifespan, which was set to age 95. Equally important, AIPlanner assumes the full range of stochastic return possibilities, while MaxiFi Planner is assuming a fixed rate of return.

Comparing the SPIA recommendations is more meaningful, but even here there are issues. AIPlanner currently recommends a gradual SPIA purchase strategy. Due to transaction costs this is unrealistic. A few SPIA purchases in a life time is more likely. The extent to which this improves AIPlanner's results is not known, but it is assumed not to be substantial. If it were substantial this would justify making many SPIA purchases in a lifetime.

AIPlanner tentatively outperforms the recommendations of MaxiFi Planner by at least $2,000 per year. In the case of no annuitization recommendation, as is the case for most financial planning software, the difference appears significantly larger.

5 AIPlanner vs. human financial planners

Technically what a human financial planner needs to do is come up with the asset allocation and consumption amounts each year that give the highest expected utility over the remaining lifetime. They need to do this when future returns and future lifespans are uncertain. It's inconceivable that a planner's brain can properly weight all the future year-by-year probabilities to do that, so sub-optimal solutions become inevitable.

There are no magic tricks that human financial planners can employ. The best they can do is employ one of the well known rules I have already reviewed, or employ financial planning software, which may make use of the same rules.

I developed a platform that could be used to test financial planner's financial planning ability, by presenting them with the same information AIPlanner receives about the environment and asking them to input consumption and asset allocation recommendations. I have not deployed this platform, because I now feel that it is obvious that financial planners won't come close to matching the performance of AIPlanner. I could deploy it if this point is in dispute.

6 Results for more realistic models

An exciting thing about AIPlanner is it appears to largely overcome the curse of dimensionality. This allows the consumption and asset allocation problem to be addressed in increasing degrees of realism.

Unfortunately, because other approaches lack a solution in these scenarios, the results must necessarily be presented without a reference against which to assess the performance of AIPlanner. Instead in this section, I am interested in studying AIPlanner's performance, and learning from the strategies AIPlanner appears to employ.

6.1 Real bonds

The bond model is reasonably complex, displaying both a yield curve, and time varying interest rates. It hasn't previously been possible to address the consumption and investment problem with such a complicated and complete bond model.

Table 7 shows the results of AIPlanner using stocks and intermediate or long term real bonds (durations 5 or 15 years), as well as a scenario in which AIPlanner was allowed to dynamically determine the bond duration. Reflecting the yield curve, AIPlanner shows a preference for very long term duration real bonds.

	Initial duration	Initial bonds	Certainty equivalent
Intermediate-term real bonds	5	8.8%	$37,343
Long-term real bonds	15	29.4%	$38,059
Variable duration real bonds	25.9	42.6%	$38,198

Table 7: AIPlanner real bond results.

	Initial duration	Initial bonds	Certainty equivalent
Ultra-short-term nominal bonds	1	6.6%	$37,021
Short-term nominal bonds	2	7.6%	$37,090
Intermediate-term nominal bonds	5	12.7%	$37,532
7-year duration nominal bonds	7	14.1%	$37,766
10-year duration nominal bonds	10	16.3%	$37,789
Long-term nominal bonds	15	27.5%	$37,895
20-year duration nominal bonds	20	23.6%	$37,775
30-year duration nominal bonds	30	23.3%	$37,686
Variable duration nominal bonds	13.9	15.0%	$37,859

Table 8: AIPlanner nominal bond results.

6.2 Nominal bonds

Table 8 shows the results of AIPlanner using a variety of fixed duration or dynamically variable duration nominal bonds. The sweet spot appears to occur at around 15 years, and this is close to the initial duration chosen when AIPlanner is allowed to select the duration. The standard error of the certainty equivalent metric is around $40 for all of these scenarios, meaning that the results for long-term nominal bonds and variable duration nominal bonds are a tie. AIPlanner again shows a preference for long duration bonds, although not as long a duration as in the real case.

It is tempting to try to read something into the reported 27.5% initial long-term bond allocation and versus the 15.0% initial variable duration bond allocation, but the difference is probably not significant. The initial variable bond allocations spanned the range 4.0% to 40.0% in the 10 runs.

6.3 Real and nominal bonds

The certainty equivalents obtained using real bonds outperformed those of nominal bonds. This is to be expected given our bond model in which the inflation risk premium is exactly offset by the real liquidity premium. Real and nominal bonds thus have similar real yields, but nominal bonds carry additional volatility due to the volatility of inflation.

It would be hoped that by making available both real and nominal bonds at the same time, AIPlanner would be able to outperform the scenarios in which on a single type of bond is available. As shown in Table 9 this is not the case. Having two bond types to choose from produced results that outperformed nominal bonds alone, but not real bonds alone. The difference isn't great, but it is significant, and it points to one of the risks with machine learning. When given too many similar knobs to tune, spurious correlations may be identified by the agent which aren't borne out when the trained model is evaluated.

Initial real duration	29.1
Initial nominal duration	14.0
Initial real bonds	19.5%
Initial nominal bonds	19.1%
Certainty equivalent	$38,075

Table 9: AIPlanner variable duration real and nominal bond results.

	Certainty equivalent
Long-term real bonds	$38,059
Randomized long-term real bonds	$37,316
Long-term nominal bonds	$37,895
Randomized long-term nominal bonds	$37,382
Variable real and nominal bonds	$38,075
Variable real and nominal bonds real interest rate blind	$37,433
Variable real and nominal bonds inflation rate blind	$38,114
Variable real and nominal bonds fully blind	$37,484

Table 10: AIPlanner and the ability to observe the interest rate.

6.4 Taking advantage of interest rates

When interest rates are high the returns for short term bonds are obviously high. Moreover because returns are less likely to go much higher, it is more likely for yields to fall, and for the returns of existing long term bond to increase. Conversely when interest rates are low the returns of both short and long term bonds are likely to be low in the future. It is common to model the future return of bonds as independent variables, with the interest rate not being modeled at all, so the effect of this fact frequently isn't observed.

The AIPlanner bond model does however include dynamic real interest rates and inflation rates. Moreover the real interest rate and the inflation rate are two of the parameters made available to the agent when observing the financial environment. Thus it is technically possible for AIPlanner to take advantage of this information. The question is does it? It turns out, it does.

Table 10 shows the performance of AIPlanner with both the bond model, and a model in which the order of the bond returns have been randomized so that they are independent of the interest rate. As can be seen AIPlanner delivers worse results when there is no relationship between bond returns and interest rates. As further proof of what is going on the table also shows the results for variable duration real and nominal bonds in scenarios in which the reinforcement learning agent was not allowed to observe the real interest rate, the inflation rate, or both the real interest rate and the inflation rate. From this it is clear that the agent is taking advantage of the interest rate, but not the inflation rate in making bond investment decisions.

I never set out to teach AIPlanner to adjust its strategy depending on the interest rate, investing more heavily in bonds when the interest rate is high. It is gratifying to see it spotted and took advantage of this opportunity by itself.

6.5 SPIAs

Income annuities are a valuable tool for ensuring client don't outlive their money. Table 11 shows what happens in the variable real and nominal bonds scenario when it is also possible to purchase SPIAs. Even at age 65 the availability of SPIAs provide a substantial increase in expected certainty equivalent consumption.

	Initial SPIAs purchase	Certainty equivalent
No SPIAs	-	$38,075
Real SPIAs	$68,244	$40,762
Nominal SPIAs	$51,025	$41,363
Real and Nominal SPIAs	$31,610 and $15,339	$41,179

Table 11: AIPlanner and SPIAs.

Nominal SPIAs appear to have a small advantage over real SPIAs. This isn't too surprising given real SPIAs were assigned a 6% lower MWR. This result should be contrasted against that of Irlam (2015). There real SPIAs were found to have a slight advantage of nominal SPIAs. An explanation of this discrepancy is that with AIG, the former real SPIA price leader, having recently left the real SPIA market, Principal Financial Group is now the only U.S. provider of real SPIAs. In the present analysis I employ a lower real MWR than in my earlier analysis. This lower real MWR is intended to be indicative of SPIA prices in the present U.S. marketplace. As a result of this lower real MWR, nominal SPIAs now have a small advantage.

Slightly worse performance is observed when both real and nominal SPIAs are available, than when nominal SPIAs are available alone. This is ascribed to the agent chasing spurious correlations when it has too many similar knobs to tune.

7 Discussion

7.1 AIPlanner and financial planning today

7.1.1 Financial planners

Financial planning these days is mostly done by humans and most planners utilize financial planning software. Financial planning firms range from small independent one-person shops to large national firms like Ameriprise. There are also networks of independent firms such as the Garrett Planning Network and the National Association of Personal Financial Advisers. Popular software products utilized by these firms include MoneyGuidePro, eMoney, NaviPlan, and others.

Perhaps somewhat surprisingly most financial planning software today only projects, and does not plan. It is more of a calculator that makes projections of the future, than a tool that helps optimize the future. And even in projecting, by treating life expectancy as a fixed quantity rather than a stochastic variable, the quality and usefulness of its projections is questionable.

Financial planners tend to concentrate on the wealthiest segment of the population; who can afford the fees. Because wealth tends to accumulate with age (at least up to retirement), financial planning clients tend to be older than the general population. Most upscale clients do not face a serious risk of depleting savings during retirement so the focus of planning tends to be on the trajectory of wealth over time and projected bequests. Asset allocation and withdrawal strategies are typically tested on a trial-and-error basis, but the software does not perform optimization. Most software runs Monte Carlo projections of investment performance and the variability of results is typically shown year-by-year, but without consideration of mortality. Annuities are rarely incorporated in such software, which is designed to handle regular investments only.

The effectiveness of trial-and-error approaches to planning is limited. It is typically only possible to specify a simple consumption rule and a fixed asset allocation, rather than test approaches that are fully responsive to the evolving scenario. For example, a drop in equity prices may cause the total portfolio (which includes guaranteed income) to become more bond like, which then argues for a higher investment portfolio asset allocation to equities than was previously planned. This will restore the total portfolio asset allocation balance. It is however infeasible to incorporate such contingencies into trial-and-error planning.

7.1.2 Robo-advisers

A newer development is known as robo-advisers where computers automate many of the human functions and do most of the work. Betterment and Wealthfront are examples of robo-advisers. The development of rob-advisers has also given rise to hybrid models that combine computer-generated advice with the opportunity for clients to interact with human advisers. An example is Vanguard personal Advisory Services, which primarily offers asset allocation advice utilizing Vanguard funds and also provides the opportunity for limited interaction with financial planners employed by Vanguard.

The robo movement in particular has targeted younger investors who are at the early stages of accumulating wealth, and with a lower-cost business model that can address their needs at a reasonable cost. Robo-advisors so far have been focused on accumulation without getting into the complexities of the retirement stage. Services offered in addition to asset allocation and savings recommendations include periodic rebalancing and in some cases tax-loss harvesting. The robo-advisers also compete with each other in attempting to offer user-friendly platforms for gathering information about client assets from all sources and making it easy for clients to access up-to-date information about their investments. In general the robo movement has mostly attempted to automate what financial planners do rather than offering new services.

7.1.3 Do-it-yourself

Besides planners and robo-advisors, which attempt to make money on financial planning, there are also growing numbers of software packages offered for free for do-it-yourself financial planning. Examples include AARP which offers a simple financial planning package that can be utilized by both AARP members and the general public.

7.1.4 AIPlanner

A key distinction that differentiates planners, financial planning software, and robo-advisers from what AIPlanner offers is that AIPlanner is based on economics principles of life-cycle finance and consumption smoothing. It is ironic that while this branch of economics has been under development for close to 100 years and reflected the work of more than a dozen Nobel laureates including Paul Samuelson and Robert Merton, it has never crossed over to mainstream financial planning. Such planning has tended to rely on much more rudimentary strategies involving trial-and-error testing of retirement withdrawal strategies and asset allocations and measured results in terms of expected bequests and probabilities of plan failure. By contrast, life-cycle finance employs economic utility concepts that provide more robust measurement of financial plan performance than the typical planner approaches, and provides a basis for optimizing withdrawals and asset allocations instead of relying on trial-and-error approaches.

7.2 Investment strategy

7.2.1 Bond duration

AIPlanner includes a sophisticated bond model with both a yield curve, and time varying interest rates. Even in the presence of quite volatile inflation rates, AIPlanner demonstrated the desirability of long duration bonds.

The difference in outcomes between short duration and long duration bonds isn't huge, but is definitely there. A large difference in outcomes wasn't to be expected because the difference in yields is only 1-2%, bonds only made up 20-30% of the investment portfolio, and the investment portfolio only made up about 60% of the total portfolio once guaranteed income is factored in.

Long duration bonds are often viewed as being too volatile. This is the result of two illusions. Measurement illusion in which the return of bonds is constantly being measured, instead of the role the bonds are playing in the portfolio. They primarily buffer risk rather than deliver return. And what I will call IID illusion, in which bonds are seen as a source of IID returns. They are not a source of IID returns. Returns are time dependent, and for long term bonds low returns in one period increase the likelihood of high returns in a subsequent periods as interest rates meander. Remember, when held to maturity, a zero coupon bond delivers exactly its promised price, despite all the volatility in its value along the way.

7.2.2 Real bonds

Assuming the the inflation risk premium is offset by the real liquidity premium, AIPlanner also showed the desirability of inflation-indexed Treasury bonds, over nominal Treasury bonds. The history of real bonds is too short, and spans a period of financial crisis, to be able to make a solid call on this issue. But it currently seems reasonable to assume that the yield premium of real bonds on account of their relative ill-liquidity is now similar in magnitude to the yield premium of nominal bonds on account of their bearing inflation risk.

Inflation-indexed bonds are often seen as having too low a yield, and being too volatile. This is the result of money illusion, in which the yields on nominal bonds are viewed in nominal terms, and the volatility of inflation gets added to real bond returns, but hidden when viewing nominal bonds.

7.2.3 Asset allocation

AIPlanner also showed the benefits of tuning the bond allocation depending upon the current value of the real interest rate. We are presently in a low rate environment, which suggests reducing bond allocations. But stocks also presently appear somewhat overvalued, which suggests pulling back on the allocation to stocks. So, as far as being able to translate AIPlanner's asset allocation strategy into immediate financial advice, our ability is somewhat muted.

7.2.4 SPIAs

AIPlanner reaffirmed the desirability of income annuities for hedging longevity risk, and found a small preference for using nominal SPIAs over inflation indexed SPIAs in the U.S. market today. The difficulty with using nominal SPIAs is determining how much to consume, and how much to save. If the full payout is consumed, the client will find their consumption slowly getting eaten away by the effects of inflation. The client thus needs to set aside some money to buffer this.

7.3 Reinforcement learning

Weissensteiner (2009) applied a form of reinforcement learning known as Q-Learning to the consumption and investment problem. What differentiates the present work is the use of artificial neural networks, rather than self-organizing maps, the use of more advanced reinforcement learning algorithms, and a richer financial model.

In a 2017 review of machine learning and financial planning, Mulvey (2017) found financial planning had so far been relatively immune to machine learning technology. However, as a result of advances in deep reinforcement learning, including the results reported here, this appears to be changing. Deep reinforcement learning appears like a fundamental technique to use when performing financial planning in the future.

7.3.1 Reinforcement learning algorithms

AIPlanner currently uses the PPO deep reinforcement learning algorithm. I also tried the Deep Deterministic Policy Gradient (DDPG) algorithm, Lillicrap et al. (2015). Using DDPG I obtained substantially worse performance. This appears to be a result of initially poor outcomes continuing to be sampled form the replay buffer, making it more difficult to learn from the more likely outcomes. I have not had time to try out other deep reinforcement learning algorithms capable of performing continuous actions, such as TRPO as described by Schulman et al. (2015) and ACKTR as described by Wu et al. (2017). It is possible these alternative algorithms could offer better performance.

The field of deep reinforcement learning is evolving rapidly. As I write this, the PPO algorithm was proposed less than 1 year ago. Substantial effort is now required to keep up with new developments in the field of deep reinforcement learning.

Most deep reinforcement learning algorithms assume that it is only possible to learn through an online process of stepping the environment from one state to the next. This isn't true in the financial planning space, where it is possible to put the simulated financial environment into arbitrary states. This may provide new opportunities to learn, and new algorithms, that work by repeatedly placing the environment into states were the agent appears to be performing poorly, so that it can directly learn from them. The risk here is failing to ride the current momentum behind the online learning approach.

7.3.2 Acceptability of deep reinforcement learning

AIPlanner gives near optimal recommendations for simple scenarios, but it is impossible to be 100% certain of the quality of recommendations in more sophisticated scenarios. The social, legal, and regulatory acceptability of a financial planning using such a tool needs to be assessed. AIPlanner appears to perform very well in practice, but lacks an ironclad guarantee that it will do so in all circumstances. What we can say is AIPlanner has been tested for millions of simulated years, and it performs very well overall.

These issues of acceptability of the approach are not unique to financial planning. Similar issues exist in many other fields of deep learning, be they medical imaging diagnosis, or self driving cars.

7.4 AIPlanner

I intend to continue to refine AIPlanner, make it more featureful, and to explore licensing opportunities so that the benefits of AIPlanner reach as wide an audience as possible. Considerable effort will go into making AIPlanner production ready.

7.4.1 Extending the financial model

Taxes and transaction costs have not been considered. Unlike the case for stochastic dynamic programming where the curse of dimensionality rears its head, the impact of taxes and transaction costs could probably be readily included in the model.

The bond model is significantly more advanced than what is commonly used when addressing the consumption and investment strategy problem. It could however be further refined to reduce the standard deviation of inflation. This could possibly be by using a two-factor rather than a one-factor Hull-White model.

The SPIA model currently recommends partial gradual annuitization. This is a fine mathematical recommendation, but transaction costs mean that a more reasonable recommendation would be to allow the annuitization recommendation to build up, and then only purchase SPIAs a couple of times in a lifetime.

The financial landscape is far richer than currently considered by AIPlanner. Guaranteed income comes in many forms with many variants. Assets may exist in traditional IRA, 401(k), Roth, or taxable accounts. Available investment asset classes are far more varied. Home equity is an important part of the financial picture for many households. Life insurance and emergency funds are used to address unexpected contingencies. Utility does not follow a strictly constant relative risk aversion model, but more likely has breakpoints around the minimum required consumption level, and the maximum desired consumption level. There are also differing bequest motives. Clients have differing abilities to psychologically handle differing degrees of market risk. That is, they have different risk tolerances. Financial goals may involve funding college, purchasing a house, or funding retirement.

If clients present with different financial levels of what is fundamentally the same investment problem it seems it would be relatively straight forward to extend AIPlanner to handle these scenarios. But if the problems posed by different clients are fundamentally different, it might not be possible to shoehorn the clients into a one size fits all sort of model.

7.4.2 Extending AIPlanner

All of the scenarios considered here have been for a single individual. AIPlanner is also capable of generating financial plans for a couple. Several technical approaches to this were considered:

- First train a mode for an individual, and then use it to train a separate model for a couple. This is the approach taken by Sheikh et al. (2014) for stochastic dynamic programming. This approach would more than double the training time. A difficulty with this approach is deciding on appropriate model parameters to use for the individual. The values used should ideally match the values produced by the couple model when one party dies, but the couple model hasn't been generated yet. This approach is also dependent on the availability and accuracy of an estimate of the expected future rewards for a single individual. Since not all deep reinforcement learning algorithms may make such information readily available this approach was not considered further.

- Train neural network models for a couple and an individual in parallel. Initially the financial model would be gathering rewards for a couple, then when one party dies it would switch to training an individual, and finally sum the rewards for that individual and return them to the couple model. The techniques needed to implement this approach appear quite challenging.

- Maintain separate probabilities of both the couple and single individuals being alive in the financial model. This approach is incapable of handling death contingent guaranteed income or wealth, such as a change in SPIA payouts upon death, or a life insurance payout.

- Abandon computing the probability of being alive, and instead have death occur stochastically for both individuals. This approach is simple but reduces the amount of information obtained from each episode. This approach was observed to do as well for an initial couple as the next approach, but not as well for an initially single individual.

- Don't compute the probability of being alive for the couple, but do compute the probability of being alive for an individual once one party dies. The certain couple stochastically transitions into an individual with an alive probability. This is the approach I currently use.

The last 3 approaches all involve training a single model to be used by both the couple and a single individual. A couple requires less than twice the resources of a single individual for the same per person level of utility. Consequently, the aggregate utility for the couple is given by:

$$U_c\left(C\right) = 2 * U_s\left(\frac{C}{1+a}\right)$$

where U_c is the utility of the couple, U_s is the utility of a single individual, C is consumption, and a is a factor indicating the additional resources required by the second individual relative to the first. I use $a = 0.6$. Observation of the model includes the life expectancy of the couple, the life expectancy for a single individual once one member of the couple has died, and whether one or both individuals are alive. If SPIAs are present it is also necessary to track both the joint payout amount and the individual death contingent payout amounts.

The observed performance of the couple model in the presence of SPIAs appears sub-optimal. This may be the result of a 2 hidden layer deep neural network doing a poor job of computing logical operations; such as needing to use a different value function depending on whether one or two individuals are alive. For visual processing tasks much deeper convolutional neural networks do a good job of disambiguating such operations, but I am uncertain as to their performance on continuous control tasks. Consequently I am currently exploring implementing the second approach described above. In a few years deep reinforcement learning may have evolved to the point where such manual tuning isn't necessary.

Today AIPlanner only solves part of the consumption and investment problem: the strategy to use in retirement. The strategy to employ while working has not yet been considered.

Often the desired outputs of the model are not simply how much to consume and how to invest, but whether, when to retire, and what sort of retirement lifestyle to expect. Given a decent underlying financial model, determining these ancillary outputs using AIPlanner should be relatively straight forward. It simply requires repeated application of the strategy to the financial model, and assembling the results.

7.4.3 Training

All of the results presented for AIPlanner are based on training a model that starts out at a particular age and wealth level, the same age and wealth level as is then used during evaluation. To deploy AIPlanner it will be necessary to utilize models that were not trained using the exact same values as the client presents with. Preliminary results are that this might reduce the certainty equivalent performance of AIPlanner by about 0.3%.

AIPlanner is only trained to handle states that are similar to states it has previously seen. This means if someone approaches AIPlanner with a scenario that is far from the optimal strategy, AIPlanner may have little experience in handling it, and may make poor recommendations.

8 Conclusion

AIPlanner delivers results that appear very close to optimal. Comparing AIPlanner to the theoretical best possible results, AIPlanner delivered in excess of 99% of the theoretical best possible result in every case. Comparing AIPlanner to other popular approaches, in a reasonable scenario AIPlanner was found to effectively deliver approximately $1,000 to $8,000 in additional consumption per year, depending on the other approach.

AIPlanner appears to at least partially overcome the curse of dimensionality. It would have previously been infeasible to address a problem involving asset allocation and consumption, time varying real and nominal bond yields, variable bond durations, and real and nominal SPIAs. In short, AIPlanner is able to address far more sophisticated financial models than are possible using existing approaches to financial planning. We may no longer be constrained by what is computationally feasible, but primarily by the accuracy of the financial models we employ.

9 Acknowledgments

I am grateful for significant help and advice from Joseph Tomlinson on early versions of this manuscript. Joe wrote most of Section 7.1 looking at financial planning today and Appendix A explaining the certainty equivalent performance metric.

Appendix A Financial planning metrics

Researchers need to have ways to compare the prospective performance of various financial planning strategies. This is particularly important in retirement planning for evaluating various strategies for utilizing savings to provide for consumption spending during retirement. For example, one strategy might follow the classic 4% Rule; spend 4% of savings in the first year of retirement and increase spending by the inflation rate each year. Another strategy might involve taking withdrawals from savings based on the IRS Required Minimum Distribution (RMD) percentages that increase as a function of age. And there are many, many, more possible strategies, and an additional dimension gets added by the choice of an asset allocation strategy for savings.

So we face the questions of: How do we evaluate the expected performance of various strategies, and how to we determine which withdrawal strategy and asset allocation will work best? We can begin to sort this out by thinking conceptually about the appropriate criteria for evaluation. I'll offer the following suggestions:

- Higher consumption spending. If one strategy uniformly has higher consumption spending than another it is to be preferred. (We'll leave out bequest considerations to keep things simple, although bequests can be incorporated as end-of-life consumption.)
- Low probability of failure. An example of failure might be depleting savings during retirement and being forced to live on Social Security only. A strategy that minimizes the probability of failure is preferable to one that is riskier.
- Low magnitude of failure. Not only is the probability of failure important, but magnitude is also an important consideration. For example, a plan where spending drops from $40,000 to $16,000 of Social Security if funds are depleted is less desirable than one where the drop is only to $30,000 because a portion of savings has been used to purchase an income annuity to generate additional lifetime consumption.
- Low variability of consumption spending. Most individuals and couples would be wary of a disruptive strategy that required big year-to-year changes in retirement consumption, perhaps driven by volatile investment performance. If two strategies produce the same average consumption, but with different volatility, most people would choose the less volatile one.

One way to evaluate strategies would be to calculate separate measures for each of the above items. For example, one could run Monte Carlo projection models using existing or customized software and calculate measures for each of the four items and perhaps others as well. The problem with this approach is that evaluating multiple strategies involves attempting to compare a lot of numbers: the number of strategies times the number of measures. This naturally leads to the question of whether there might be a way to combine the above measures into some type of summary measure.

A.1 The utility function

It turns out that we can provide a useful summary measure by relying on basic economics that dates back centuries. The key is the utility function describing the desirability of different consumption amount. An example utility function is shown in Figure 2.

A.2 Properties of the utility function

Note the curvature of the utility function which mathematicians describe as concave. In terms of converting dollars into utility points this curvature means that, at low levels of income, an increase or decrease of, for

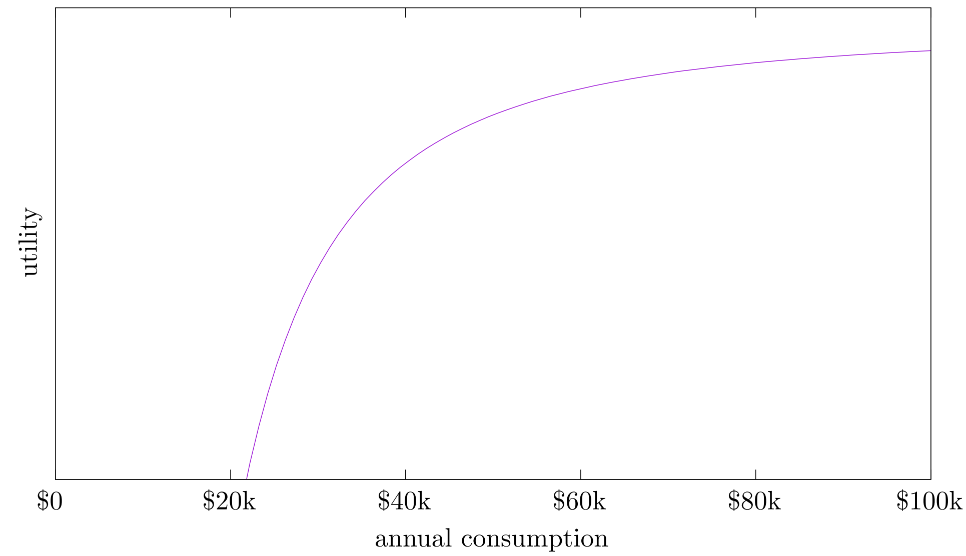

Figure 2: Example utility function, $\gamma = 3$.

example, $1,000 in consumption changes utility or well-being a lot. By contrast, at high levels of income, the change in utility for the same dollar change in consumption is significantly diminished. This phenomenon is what economists refer to as diminishing marginal utility.

The formula for this particular utility function is:

$$U = \frac{C^{1-\gamma}}{1-\gamma}$$

where U is utility, C is consumption, and γ is known as the risk aversion coefficient. It's a measure of tolerance for variability in year-to-year consumption. Surveys have shown typical γ values for consumers of about 2 or 3, with values ranging from 1 to about 10 for the range of highly tolerant of variability to very averse. A higher γ value means that the function is more curved or concave. This particular utility function is known as constant relative risk aversion (CRRA) and is the most popular function used by economists to convert consumption into utility.

For individuals, it would be feasible to design a short questionnaire that asks about trade-offs between level consumption and variable consumption, including trade-offs involving plan failure and big drops in consumption. The answers could be mapped to a risk aversion coefficient. Such a questionnaire would be similar to ones that have been used by economists in surveys about risk aversion.

I'll first briefly describe how the utility calculations work and then describe how utility analysis combines the above criteria into a single measure. Let's say we are modeling a 30-year retirement and we have year-by-year projected consumption spending, and we further assume that spending varies a bit from year to year. We use the above equation to translate each year's projected consumption spending into a utility measure. (On the graph we are determining the values on the y-axis that correspond to x-axis values for each of the 30 years of assumed retirement.) We then take an average of the 30 utility numbers and this can be thought of as the average annual utility over the course of retirement.

Unfortunately, utility numbers themselves lack intuitive meaning, although the higher the number the better, but we get around this by translating the average utility, $\bar{U}$ into what is known as certainty equivalent (CE) consumption. To do this we use the inverse of the utility function:

$$CE = ((1-\gamma)\bar{U})^{\frac{1}{1-\gamma}} \tag{1}$$

In terms of the utility function what is happening is that we go from an average utility on the y-axis of our graph to the corresponding consumption spending number on the x-axis. So we end up with our average utility measure translated into an annual retirement consumption dollar equivalent. Conceptually, the certainty equivalent is the level amount of annual consumption an individual would be willing to accept in exchange for variable year-to-year consumption.

A.3 Examples

These examples will show how the certainty equivalent combines the four criteria above into a single measure.

Variability of consumption. I'll start with the last of the four criteria. Let's say that over the course of retirement annual consumption will average \$40,000, but each year has a 50% probability of being either \$35,000 or \$45,000. If we assume a risk aversion coefficient of 3 and build a 30-year retirement model where half the years are \$35,000 and the other half are \$45,000, we can use the formulas to calculate a certainty equivalent of \$39,071. What this means is that the individual would be willing to trade average consumption of \$40,000 and that bounces around between \$35,000 and \$45,000 for the certainty of level consumption of \$39,071. We can also do the same calculation for a very risk averse individual ($\gamma = 10$) and determine that they would be willing to accept a level \$37,388 in exchange for the same volatile consumption.

Let's say that we go back to our γ of 3 and assume the same \$40,000 average consumption, but with more volatility where the \$35,000 or \$45,000 becomes \$25,000 or \$55,000. The CE then drops from \$39,071 to \$32,186. So CE as a performance measure takes consumption variability into account. Less variability, which means more certainty, is better than more variability, and gets a higher CE score.

Plan failure. We can show that the CE measure also takes into account the combination of probability and magnitude of failure: items two and three on the criteria list. We will make the assumption that, for a 30-year retirement, plan failure means that consumption goes along at a level \$40,000 per year and then drops to \$16,000 of Social Security benefits after 20 years when savings are depleted. And we will further assume that there is a 30% chance of plan failure. A little algebra shows that average consumption for the failure cases is \$32,000. However, the CE for the failure cases drops to \$24,121, well below the average consumption, and this illustrates the substantial utility penalty from running out of money and living 10 years on Social Security only.

The CE for the 70% of cases that do not fail (\$40,000 of level annual consumption for the full 30-year retirement) is indeed \$40,000 because there is no consumption variability. The weighted average CE that takes both probability and magnitude of failure into account works out to \$32,391.

Higher consumption spending. The effect of higher consumption spending on CEs is quite direct. For level consumption cases a change in the level of consumption spending changes the CE by exactly the same amount. For variable consumption, an equal change of all the amounts will not produce an exact change in the CE by the change amount, but it will be close.

A.4 Monte Carlo projections

Monte Carlo modeling may produce 10,000 or more retirement consumption paths, each with its own average consumption, plan failure (or not), and variability of consumption. It's straightforward to calculate the utility of each of the consumption paths. These utilities can then be averaged to calculate the average utility for the particular retirement strategy being tested. This average utility can be expressed as a CE using the inverse formula.

Programming is used to optimize the CE of retirement strategies by varying planned year-by-year asset allocations and withdrawals from savings.

References

Bengen, W. P. (1994). Determining withdrawal rates using historical data. *Journal of Financial planning*, 7(4):171.

Dimson, E., Marsh, P., and Staunton, M. (2017). Credit suisse global investment returns yearbook 2017. *Credit Suisse, Zurich*.

Guyton, J. T. (2004). Decision rules and portfolio management for retirees: Is the "safe" initial withdrawal rate too safe? *Journal of Financial Planning*, 17(10):54–62.

Guyton, J. T. and Klinger, W. J. (2006). Decision rules and maximum initial withdrawal rates. *Journal of Financial Planning*, 19(3):48.

Irlam, G. (2014). Floor and upside investing in retirement with nominal spias. *SSRN: 2531779*.

Irlam, G. (2015). Floor and upside investing in retirement with real instruments. *SSRN: 2554752*.

Irlam, G. and Tomlinson, J. (2014). Retirement income research: What can we learn from economics? *The Journal of Retirement*, 1(4):118–128.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529.

Mulvey, J. M. (2017). Machine learning and financial planning. *IEEE Potentials*, 36(6):8–13.

Pfau, W. D. (2015). Making sense out of variable spending strategies for retirees. *SSRN: 2579123*.

Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–246.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015). Trust region policy optimization. In *International Conference on Machine Learning*, pages 1889–1897.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Sheikh, A. Z., Roy, K. S., and Lester, A. (2014). Breaking the 4% rule. *J.P. Morgan Asset Mnagment*.

Weissensteiner, A. (2009). A q-learning approach to derive optimal consumption and investment strategies. *IEEE transactions on neural networks*, 20(8):1234–1243.

Wu, Y., Mansimov, E., Grosse, R. B., Liao, S., and Ba, J. (2017). Scalable trust-region method for deep reinforcement learning using kronecker-factored approximation. In *Advances in neural information processing systems*, pages 5285–5294.

Zolt, D. M. (2013). Achieving a higher safe withdrawal rate with the target percentage adjustment. *Journal of Financial Planning*, 26(1):51–59.