Optimal Asset Allocation for Passive Investing with Capital Loss Harvesting *

Running title: Optimal Allocation with Capital Loss Harvesting
Keywords and phrases: Taxes, Capital Losses, Portfolio Optimization, Expected Utility, Passive Investing

Daniel N. Ostrov[†] and Thomas G. Wong[‡]

July 27, 2010

Abstract

This paper examines how to quantify and optimally utilize the beneficial effect that capital loss harvesting generates in a taxable portfolio. We explicitly determine the optimal initial asset allocation for an investor who follows the continuous time dynamic trading strategy of Constantinides (1983). This strategy sells and re-buys all stocks with losses, but is otherwise passive. Our model allows the use of the stock position's full purchase history for computing the cost basis. The method can also be used to rebalance at later times. For portfolios with one stock position and cash, the probability density function for the portfolio's state corresponds to the solution of a 3 space + 1 time dimensional PDE with an oblique reflecting boundary condition. Extensions of this PDE, including to the case of multiple correlated stocks, are also presented.

*The authors thank Sanjiv Das for many helpful comments during the writing of this paper.

†Address: Department of Mathematics and Computer Science, Santa Clara University; 500 El Camino Real, Santa Clara, CA 95053, USA. Email: dostrov@scu.edu. Telephone: 408-554-4551. Fax: 408-554-2370.

‡Address: Department of Physics, 0354 University of California, San Diego; 9500 Gilman Dr., La Jolla, CA 92093, USA

We detail a numerical algorithm for the PDE in the single stock case. The algorithm shows the significant effect capital loss harvesting can have on the optimal stock allocation, and it also allows us to compute the expected additional return that capital loss harvesting generates. Our PDE based algorithm, compared to Monte Carlo methods, is shown to generate much more precise results in a fraction of the time. Finally, we employ Monte Carlo methods to approximate the impact of many of our model's assumptions.

1 Introduction

In this paper we consider the question of determining which asset allocation maximizes the expected utility of a taxable portfolio at a future liquidation date. The portfolio contains a risky asset, or assets, subject to capital gain and loss tax rules and a tax-exempt, risk-free asset or the tax-exempt equivalent of a taxable, risk-free asset. Hereafter, we will refer to the risky asset as stock and the risk-free asset as cash. We focus on passive investing strategies since passive investing helps minimize tax loss in the presence of capital gains. Further, we augment the strategy by using the tax advantage active trading generates in the presence of capital losses.

The literature on the subject of optimal allocation in the presence of taxes is relatively sparse despite its obvious importance. This is largely due to the computational difficulties presented by even significantly simplified versions of the problem, as we discuss in the next subsection.

1.1 History of Optimization with Taxes

In his classic 1983 paper on portfolio optimization in the presence of taxes, Constantinides [5] advocates what we will call the *augmented passive strategy* just outlined above, namely, "(T)he optimal liquidation policy is to realize losses immediately and defer gains until the event of a forced liquidation." Deferring gains allows taxable accounts to benefit from the same tax deferral that retirement plans like 401(k) accounts guarantee. The second half of the augmented passive strategy, to immediately sell and repurchase assets with capital losses, is based on the same principle: just as it is always desirable to defer paying tax as long as possible, it is also always desirable to collect tax credit as soon as possible. More attention is generally given to the tax deferral half of the augmented passive strategy. This paper, however, quantifies the effects of immediately realizing losses, showing that it can have a significant impact on optimal asset allocation and portfolio growth.

The full optimal allocation problem is, of course, more complicated than just employing the augmented passive strategy. As Dybvig and Koo [10] point out, "The rule of...deferring capital gains indefinitely and realizing losses immediately...is in conflict with another very important goal of investment, namely finding the optimal balance of risk and return. For example, if the market has been bullish for a while, failure to realize capital gains may imply taking on more risk than is tolerable." Models that directly consider future rebalancing while retaining consideration of the full cost basis history contained in past purchases are extremely computationally intensive. Chang [3], [4] has formulated the general problem using continuous time optimal rebalancing, and shown that the value function is a viscosity solution to an infinite dimensional Hamilton-Jacobi-Bellman relation, which, of course, is well beyond the realm of computation. But even worse, models that allow for even a single rebalancing at some given future time are equally problematic. They also have an infinite number of dimensions if the model for the evolution of the stock price allows a continuum of future prices at the time of the rebalance (such as geometric Brownian motion models).

Given this immutable difficulty, significant simplifications to the model must be made for computation to be feasible. In [10], Dybvig and Koo retain the use of the full cost basis history, but restrict the evolutionary model to a single stock with growth governed by at most four periods of binomial jumps. By exploiting the nature of the final expected utility and applying an efficient nonlinear programming algorithm developed by Gill et al. in [13], DeMiguel and Uppal [8] were able to extend Dybvig and Koo's analysis to ten binomial periods for a single stock or four binomial periods for four risky stocks.

The advantages of deferring gains, however, generally require a significant number of years to become apparent, so simulations of passive portfolios that are accurate over long time intervals are heavily desired. As a simple example of the importance of the time horizon, we compare two portfolios, each containing stocks that grow at 7% per year without volatility and each subject to a capital gains tax rate of 30%. The first portfolio is run passively so capital gains are only paid at the portfolio's liquidation time, T. The second portfolio is run actively with continuous turnover and therefore is subject to immediate capital gains taxation from time 0 to time T. The ratio of the worth of the passive portfolio to the active portfolio at the final time T after all taxes have been paid is

$$e^{(.30)(.07)T}\left[1-.30\left(1-e^{-.07T}\right)\right].$$

If T equals 10 years, then the passive portfolio is worth 5% more than the active portfolio; if $T = 20$ years, the passive portfolio is worth 18% more; by $T = 40$ years, the passive portfolio is worth 66% more.

Substituting the average cost basis of all purchased stock in place of the full cost basis history allows for computation with more complicated evolutionary pricing models that can remain accurate over longer time horizons. The average cost basis is used for computing capital gains in Canada; it is also an option for computing capital gains with American mutual funds. The computational advantage of this model comes from the fact that a single variable now replaces what was previously an evolving function representing the full history of purchase prices. This average cost basis method has been developed by Dammon, Spatt, and Zhang [6], [7] for multi-period binomial evolution of a single stock. It has also been applied to binomial evolution of many stocks (see, for example, Gallmeyer, Kaniel, and Tompaidis [11]) and geometric Brownian motion models (see, for example, Ben Tahar, Soner, and Touzi [1], [2]). Of course, for the case of American mutual funds, it is preferable not to use the average cost basis, since it is always better to sell mutual fund holdings with the highest purchase price first while retaining the low purchase price holdings as long as possible; in other words, using the full purchasing history provides for a superior, though computationally more challenging, strategy.

1.2 Contribution of the Paper

In this paper, we fully quantify Constantinides' augmented passive strategy, thereby allowing us to explicitly compute the optimal portfolio for this strategy and the expected premium obtained from reaping losses. Specifically, we show how to locate the optimal initial portfolio allocation that maximizes the expected utility of the portfolio at the time of liquidation, given that trading between the initial and liquidation times only occurs to reap tax credits from any capital losses. Our PDE based model allows for the use of the full purchasing history of the stock position while also allowing the stock price to evolve in continuous time via geometric Brownian motion.

Because the model allows for the full purchasing history of the stock position, it can and should be reapplied on occasion with the new state of the market and portfolio to allow return and risk to continue to be balanced as suggested by Dybvig and Koo. For example, if an investor has a 40 year time horizon, the method of this paper can initially be used to give the optimal current rebalance, assuming the investor intends to only trade to reap capital losses. However, every, say, five years, the investor may wish to reapply the method with a 35 year, then 30 year, etc. time horizon so that current stock price and portfolio information can be considered in a new optimal rebalancing. Of course, each rebalancing again assumes the desired goal that no rebalancing will be needed in the future, only trades to cull capital losses.

The model would be better if it could assume even a single optimal future rebalancing at, say, the halfway-to-liquidation time, $T/2$, but, as noted before, this is not computationally possible, since it yields an infinite dimensional PDE. However, by restricting this rebalancing so that it always moves the portfolio back to the stock fraction chosen at $t = 0$, we can use Monte Carlo methods to approximate its effect. We will show that under some circumstances this rebalancing scheme is useful, but under others it is undesirable due to the early capital gains it can generate.

We also use Monte Carlo methods to approximate the effect of many factors that are not directly included in our PDE based model, nor in most of the previous models for optimization in the presence of taxes. These factors include wash sale restrictions, transaction costs, and the annual limit on undeferred losses that can be claimed under the American tax system. This allows us to extend our PDE model to approximate solutions that include these factors.

For passive investors, the assumption of no trading other than for tax arbitrage is philosophically close to their intent, making our PDE based model attractive. This is especially true in light of the model also being able to specify an optimal rebalance, including the possibility that there should be none, when an investor wants to occasionally consider rebalancing. If the rebalance requires purchasing stock, then the importance of culling losses increases since, as we will demonstrate, most useable losses occur within the first few years after a new stock purchase. For active investors, the generally passive assumptions of our model will be heavily violated, of course, so we believe active investors are better served by a model like Dammon, Spatt, and Zhang's, which allows for active continuous trading, though at the cost of needing to replace the full purchase history with the average purchase history, and thereby generally requiring an imperfect tax strategy.

Our model also allows for a number of useful alterations and extensions. We explore three of these here. First, we can extend the PDE and boundary condition analysis to multiple correlated stocks. Second, because our model computes probabilities instead of directly determining the value function for the portfolio's expected utility, we can easily accommodate probabilistic distributions for the time of liquidation. This is particularly useful if the investor is interested in optimizing the amount they bequest at death, since we can incorporate the probabilistic distribution for the time of the investor's death when all capital gains in the account are forgiven. Finally, we can also alter the model to use all cash culled from tax arbitrage to buy more stock. This may be preferable to many investors, since the presence of losses generally implies the portfolio is weighted more towards cash than is optimally desired.

By necessity, the PDE method presented here generates an approximate solution to a question that involves complex tax law. Since it corresponds to an algorithm that can be computed quickly, our method can be used to generate a starting point for future optimization models that can incorporate more details of the tax code. We make the assumptions and limitations of our model more explicit in the next section.

1.3 Financial Assumptions of the Model

Models for trading in a taxable portfolio require three sets of assumptions: assumptions for the stock and cash dynamics, assumptions for trading, and assumptions for the tax model. We next detail these assumptions for our PDE model. We explicitly indicate where our model can and cannot easily be altered to accommodate our trading and tax model assumptions, since these are less studied than our stock and cash dynamics assumptions.

For the stock and cash dynamics, we make the following common assumptions:

- We assume the stock evolves by geometric Brownian motion with a constant expected return, μ, and a constant volatility, σ.
- The tax-free interest rate for the cash position, r, is assumed to be constant.
- For simplicity, we do not consider dividends for the stock, although the model can easily be altered to include continuous dividends if desired.

We make the following assumptions for trading:

- We allow trades to rebalance the portfolio at the initial time $t = 0$ and look to maximize the expected utility of the liquidated portfolio at the final time $t = T$. The model can accommodate any utility function.

- Between $t = 0$ and $t = T$, we assume the augmented passive trading strategy, which only allows immediate selling and repurchasing of all stock with any capital loss, thereby generating tax credit. The money from the tax credit is added to the portfolio's cash position; otherwise money is neither added nor removed from the portfolio before liquidation.

  As discussed earlier, our PDE model cannot be altered to take into account the effect of future rebalancing, although it can be reapplied to perform rebalances. To get some feel for the effect of this restriction, the effect of a single future rebalance, as discussed above, is approximated through computationally intensive Monte Carlo methods in Subsection 5.2.2.

- We assume that stock and cash can be bought and sold with negligible transaction costs in any quantity, including non-integer amounts.

  As with many other models involving continuous dynamics and trading, such as the Black–Scholes model for valuing derivatives, literal use of our trading model would correspond to infinite transaction costs in the continuum limit. However, in more realistic scenarios, trading cannot occur more than once in a given time interval due to wash sale constraints, which are discussed below. In Subsections 5.1.1 and 5.1.2, we show for realistic parameters how the transaction costs quickly decay as this time interval increases to a realistic value. Given that the effects of transaction costs are small, Subsection 5.1.2 shows a method of approximating their effect by slightly decreasing the beneficial capital loss tax rate in the PDE model. In other words, the PDE cannot be directly altered to accommodate exact transactions costs, but their real world effect can be approximated.

- We prohibit being both long and short in a single stock.

  Prior to the 1997 revision of the American tax code, this would have been a significant assumption for American portfolios, since there was an important arbitrage opportunity by applying "shorting against the box." This was the practice of shorting, instead of selling, a held stock position, which deferred paying capital gains taxes on the held stock until death when all capital gains are forgiven. In 1997, this loophole was closed, removing the reason to hold both long and short positions in the same stock. For any tax code where this loophole has not been removed, shorting against the box should be considered.

- For simplicity, our model is described in detail only for long stock positions.

  In the single stock case, short positions are never advantageous, so this is not a relevant restriction. For multiple stocks, of course, short positions may be desirable. The model for multiple stocks in Subsection 3.1 is initially developed assuming all stocks are long, since this allows us to see how the single stock case generalizes. At the end of this subsection, however, we also show how to alter the model to accommodate short positions.

Finally, we make assumptions concerning the model's tax structure that are common to, or less restrictive than, those in the papers discussed in Subsection 1.1 above:

- We do not make a distinction between short term and long term tax rates. Also, for most of the paper, we do not make a distinction between the tax rates for capital gains and losses. However, we do distinguish between R, this tax rate at times $t \in [0, T)$, and R^{liq}, this tax rate at the liquidation time, $t = T$. This distinction is made since time T might correspond to the death of an investor, in which case $R^{liq} = 0$ instead of $R^{liq} = R$, as noted above.

  Our model for the augmented passive strategy is easily altered to accommodate different rates for capital gains and capital losses, as explained in Subsection 5.1.4. Note, however, that when the capital gains rate is smaller than the capital loss rate, our strategy can potentially be further enhanced by realizing gains as well as losses. This occurs because realizing gains raises the cost basis, which improves the chance of collecting the more highly compensated losses. Our model does not cover this. For a discussion of this phenomena under different tax model assumptions, see the working paper by Marekwica [14].

- We assume no wash sale constraints.

  For a loss to qualify for tax credit, wash sale rules in America require that the investor wait at least 31 days before repurchasing a stock that was sold at a loss. Our PDE model cannot be directly extended to take into account wash sale rules. However, the effect of wash sale rules can be significantly reduced through buying and selling similar stocks as detailed in Subsection 5.1.1. This strategy allows us to approximate, but not replicate, continuous trading without wash sale constraints. In Subsection 5.1.1, we show a method of approximating the effect of not trading continuously by slightly decreasing the beneficial capital loss tax rate used in the PDE model.

- We assume that tax credits are collected immediately, and there is no limit on the amount of realized losses.

  In reality, tax credits are collected at the end of the year, which has a minor detrimental effect. Of much more importance is the fact that annual net losses are limited to $3000 in the American tax system, and any losses above this limit must be deferred to later years (so, for example, a loss of $15,000 requires 5 years to be fully claimed). Neither of these effects are incorporated into our PDE model. Therefore, our model's results will become progressively unrealistic when the size of the stock position considered becomes large. However, as we discuss at the beginning of Section 5, the median size of a taxable stock position is approximately $50,000. For stock positions of this size (and up to around $100,000), we will see in Subsection 5.1.3 that the effect of the $3000 limit is not that large over long time horizons. Subsection 5.1.3 presents a method that approximates the effect of this $3000 limit (and the effect of waiting until the end of the year to collect losses) by, again, decreasing the beneficial capital loss tax rate used in the PDE model. In our view, this approximation becomes less trustworthy as we consider stock positions that are higher than the $100,000 range. For stock positions larger than this, the PDE model must be more fundamentally altered to better describe the effect of the $3000 limit. Fortunately, this can be accomplished by adding an additional state variable for accumulated losses to alter our model in the same manner that Marekwica's working paper [14] adds this variable to alter the model by Dammon, Spatt and Zhang [7]. Specifically, losses, after being multiplied by R, are immediately transferred to this loss variable instead of the cash variable. Then, as long as the loss variable remains positive, money is transferred at a constant rate of \$3000 times R per year from the loss variable to the cash variable. Of course, by the curse of dimensionality, the addition of the loss variable to the model will slow down the computation of the PDE solution.

1.4 Organization of the Paper

In Section 2 we start our analysis for a portfolio with a single stock and cash. We begin by discussing the model's mathematical features and defining financial variables in a way that allow us to keep track of the full portfolio history by keeping track of a single variable. We then determine the partial differential equation for V, the value function which represents the expected utility at the liquidation time T. This backward Kolmogorov PDE contains a first order oblique reflecting boundary condition corresponding to the realization of capital losses. Using this equation to compute V directly is difficult due to the problem of determining the remaining boundary conditions and the need for a much larger state space than is necessary, since we are only interested in the solution corresponding to the known initial state.

Since our PDE has 4 dimensions, we must be careful about computational time. Therefore, it makes much more sense to use the adjoint equation to the PDE for V, which is the forward Kolmogorov PDE for p, the probability density of the portfolio's state. Using this adjoint PDE allows considerable computational savings since the state space can be dynamically tailored as we move forward in time to only consider locations in the state space with non-trivial probabilistic mass. Also, as is well known, by using a scaled version of p, which we denote by u, the PDE can be reduced to a simple heat equation with a single spatial direction of diffusion, which generates additional computational time savings.

The boundary condition where trading occurs is particularly interesting. When the portfolio hits the trading boundary, we sell and repurchase immediately, which changes the state of the portfolio. It is therefore intuitive to think that the boundary condition for u (or p) just corresponds to pushing probabilistic mass in the direction going from the state before trading to the state after trading. However, this is *not* correct. The direction of the push is more complicated, as is common with oblique reflecting boundary conditions. The intuitive direction *is* correct at the boundary of the PDE for V, however, this direction changes when we convert to the adjoint formulation for u (or p).[1]

In Section 3, we detail extensions of our model. We begin by extending our model to the case of m correlated stocks. We will see that the corresponding PDE for u has $(2m+1)$ spatial dimensions and 1 time dimension, so we are computationally limited to, at most, two stocks by the so-called "curse of dimensionality", which currently prevents computing the solution of a PDE with more than five or six dimensions. We then extend the model in Section 2 to the case where the liquidation time has a probabilistic distribution. Finally, we specify how the model changes if the money culled from tax arbitrage is used to immediately buy more stock instead of being left as cash.

In Section 4, we detail numerical simulations for the single stock model developed in Section 2. In particular, we develop an algorithm that computes the PDE solution using numerical methods from first order PDE theory to accommodate the first order partial derivative boundary condition. This quantifies the change in the optimal asset allocation created by the inclusion of accruing tax credit from losses. We will discuss the results generated by our algorithm for a realistic base case and for cases where each parameter of this base case is altered.

In Section 5, we use Monte Carlo methods to approximate the effect of a number of the assumptions in our model. We begin by considering the effect of allowing only periodic, as opposed to continuous, trading on losses and avoiding triggering wash sale rules. We then consider the magnitude of the loss from transactions costs. This is followed by looking at the effect of the \$3000 annual limit on undeferred losses. These factors reduce the effectiveness of harvesting capital losses, but we also see how these factors may be partially or completely compensated by considering the fact that the tax rate for losses is bigger than the tax rate for gains. Finally, although we will show that the PDE based algorithm from Section 4 is much faster and more precise than Monte Carlo methods for determining the optimal strategy when there is no rebalancing, we will use Monte Carlo methods to approximate the effect of rebalancing the portfolio at the halfway time, $T/2$, to its initial stock fraction, since PDE based methods become impractical for this case.

[1] See, for example, [17], for a nice discussion of this phenomenon.

2 Mathematical Formulation of the Model for a Single Stock

In this section we develop our model for a single stock and cash. This will lead to a partial differential equation, specifically, the heat equation, and an oblique reflecting boundary condition that govern the evolution of the probability density function.

2.1 Notation and Model Structure

We begin by defining all the necessary mathematical quantities of the model.

First, we consider the state of the portfolio just prior to the rebalance at $t = 0$. Let c^{old} be the worth in $t = T$ dollars of the cash position just prior to the rebalance. (Note: *all* cash position variables in this paper will be in $t = T$ dollars, which will permit a more simple form of the PDE.) Let $S(0)$ be the stock price at $t = 0$. The purchasing history of the stock just prior to the rebalance is given by our introducing the function $N^{old}(b)$ — See Figure 1 — which we define to be the number of stock shares in the portfolio with a purchase price strictly greater than b. (The letter b is chosen for *b*asis.) Like cumulative probability distribution functions, this function is "càdlàg", the French abbreviation for being continuous from the right and having finite limits from the left. Unlike cumulative probability distribution functions, this function is monotone *decreasing* in b. Since we assume any amount of stock with a purchase price greater than $S(0)$ is sold and then rebought at $S(0)$ to generate tax credit, we have that $N^{old}(S(0)) = 0$. Also, since all purchase prices are positive, we have that $N^{old}(0)$ represents the total number of stock in the portfolio prior to the $t = 0$ rebalance.

Next, we define quantities relevant to the rebalance at $t = 0$. The control variable n represents the number of stock bought in the rebalance at $t = 0$. If $n < 0$, then stock was sold in the rebalance. We define $B(0)$ to be the highest purchase price of the portfolio's stock shares just after the rebalance. Specifically, if $n > 0$ then, since stock is bought at the current purchase price, we have $B(0) = S(0)$, whereas if $n \leq 0$ then $B(0) = \min\left\{b : N^{old}(b) \leq -n\right\}$. Next we define c^{init} to be the cash in $t = T$ dollars generated by this initial

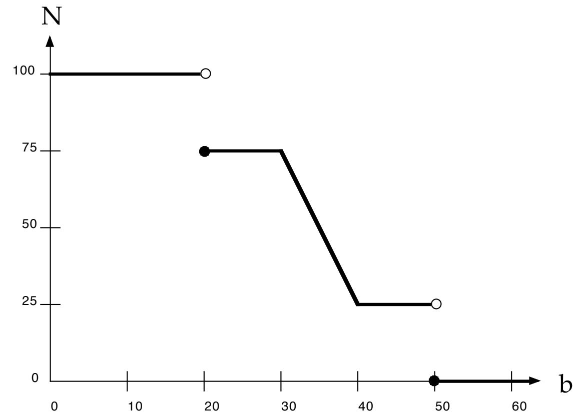

Figure 1: *Example of $N^{old}(b)$ or $N(b)$ functions. These two functions give the portfolio's full stock purchase history before and after the portfolio is rebalanced at $t = 0$. In the above figure we have a portfolio where 25 shares of stock were purchased at \$20 per share, another 25 shares of stock were purchased at \$50 per share, and the purchase prices of the final 50 shares were distributed uniformly between \$30 and \$40 per share.*

time rebalance (so $c^{init} < 0$ means cash was spent). The formula for c^{init} is

$$c^{init} = e^{rT} \left[\begin{array}{c} -\max\left[0, n\right] S(0) + \int_{S(0)}^{B(0)} \left[S(0) - R\left(S(0) - \beta\right)\right] dN^{old}(\beta) \\ +\left[S(0) - R\left(S(0) - B(0)\right)\right] \Big(\min\left[n, 0\right] - N^{old}\left(B(0)\right)\Big) \end{array} \right]. \tag{1}$$

Only the first of these three terms is non-zero if stock is bought. If stock is sold, the first term is zero, the second term (the Stieltjes integral[2]) represents most of the cash generated after capital gains are deducted, while the third term represents the remainder of the cash generated due to the fact that there might be only partial liquidation of the position at $B(0)$, the highest remaining purchase price. Finally, we define $N(b)$, the purchasing history for the stock just *after* the rebalance, which, as before, represents the number of shares of stock with an original purchase price $> b$. The formula for $N(b)$

[2]Note that even though $B(0) \leq S(0)$, the Stieltjes integral is not negative since N^{old} is monotonically decreasing.

is simply

$$N(b) = \begin{cases} 0 & \text{if } b \geq B(0) \\ N^{old}(b) + n & \text{if } b < B(0). \end{cases}$$

This incorporates the fact that any stock sold in the rebalance will have the highest possible purchase price to maximize the cost basis and thereby lower capital gains.

Now we consider the three relevant variables for times, t, between 0 and T. These are the state variables for the PDE. The first, $S(t)$, is the stock price. The second, $C(t)$, is the cash, expressed in T dollars, accumulated between (but not during) the rebalance and time t from tax credits generated by selling and repurchasing stock with capital losses. Finally, we introduce $B(t)$, the highest purchase price of the stock shares in the portfolio at time t. Since we sell and then repurchase any stock with a loss, we have that $B(t)$ equals the lowest stock price between time 0 and time t, unless the stock price never dips below $B(0)$, in which case $B(t) = B(0)$. Note that $N(b)$, the purchasing history just after the rebalance at $t = 0$, along with our trading strategy immediately imply the purchasing history of the portfolio's stock at any later time t; specifically, the purchasing history at time t equals $N(b)$ if $b < B(t)$, and it equals 0 if $b \geq B(t)$. This key fact, which motivated our introducing the function $N(b)$ and the variable $B(t)$, implies our model can specify the entire purchase history at any time $t \in (0, T)$ by only keeping track of a *single* variable, $B(t)$. This will make our model computationally tractable, and compares, in this computational sense, to Dammon, Spatt, and Zhang making their model tractable by only keeping track of a single variable, the average purchase price, instead of the entire purchase history.

Finally, we consider the time of liquidation, T. The liquidated cash worth of the portfolio at $t = T$ is

$$c^{old} + c^{init} + C(T) + \int_{B(T)}^{0} \left[S(T) - R^{liq} \left(S(T) - \beta \right) \right] dN(\beta).$$

Note again how the purchase history at $t = T$ is expressed strictly through $B(T)$ and the purchase history just after rebalancing. As a limit of integration, $B(T)$ serves as a cutoff to the purchase history just after rebalancing, thereby giving the $t = T$ purchase history. For the assumed known utility function, U, of an investor, our goal is to find the value of n that maximizes the expected utility of the liquidated cash worth of the portfolio at $t = T$. In light of this, for any specific choice of n, we define the value function

$\tilde{V}(s, b, c, t, n)$ to denote the expected utility of the liquidated portfolio at time T given that $S(t) = s$, $B(t) = b$, and $C(t) = c$, so our goal, restated, is to find which n maximizes $\tilde{V}\left(S(0), B(0), 0, 0, n\right)$. Defining $\tilde{p}(s, b, c, t, n)$ to be the probability density function at $S(t) = s$, $B(t) = b$, and $C(t) = c$ given n and its corresponding initial conditions, we can express $\tilde{V}\left(S(0), B(0), 0, 0, n\right)$ in terms of $\tilde{p}$:

$$
\begin{aligned}
& \tilde{V}\left(S(0), B(0), 0, 0, n\right) \\
= & \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{b}^{\infty} U\left(c^{old} + c^{init} + c + \int_{b}^{0} \left[s - R^{liq}\left(s - \beta\right)\right] dN(\beta)\right) \\
& \times \tilde{p}(s, b, c, T, n) ds db dc.
\end{aligned}
$$

Note here that s is only integrated from b to ∞ since, if $b > s$, the stock bought at price b would immediately be sold and then rebought, which reduces b so that it equals s again.

2.2 Determination of the Governing PDE

We divide the (s, b, c) state space into two regions: $b < s$, the no-trading region, and $b \geq s$, the trading region. In the no trading region, the portfolio is left alone. Since there are no trades, b, the highest purchase price in the portfolio, does not change, and, since c is expressed in time T dollars, it remains unchanged as well. Since we assume the stock price, S, changes by geometric Brownian motion

$$dS = \mu S dt + \sigma S dw,$$

where the expected return, μ, and the volatility, σ, are assumed constant, we can apply Ito's lemma to derive the backward Kolmogorov equation for $\tilde{V}$ in the no-trading region:

$$\tilde{V}_t + \mu S \tilde{V}_s + \frac{1}{2}\sigma^2 S^2 \tilde{V}_{ss} = 0. \tag{2}$$

Here, and in the remainder of the paper, we denote partial derivatives with subscripts (e.g., $\tilde{V}_{ss} = \frac{\partial^2 \tilde{V}}{\partial s^2}$).

In the trading region, $b \geq s$, we immediately take advantage of the tax arbitrage opportunity presented by realizing losses. Specifically, if the highest purchase price in the portfolio at time t is $b = s + \Delta b$ for some $\Delta b > 0$, we sell all the stock with purchase prices between s and $s + \Delta b$ and then repurchase them at the current stock price, s. This will generate tax credits in time T dollars worth $e^{r(T-t)} R \int_s^{s+\Delta b} N(b)db$. The expected utility of the portfolio before and after this trade is the same, so we have that

$$\tilde{V}(s, s + \Delta b, c, t, n) = \tilde{V}\left(s, s, c + e^{r(T-t)} R \int_s^{s+\Delta b} N(b)db, t, n\right). \tag{3}$$

Now we take the limit as $\Delta b \to 0$ and use that N is right continuous to conclude that on the $s = b$ boundary in the trading region, we have the following first order differential boundary condition:

$$\tilde{V}_b = \tilde{V}_c R N(b) e^{r(T-t)}. \tag{4}$$

This quantifies the fact that $\tilde{V}$ stays constant as we sell and then repurchase stocks with capital losses. This boundary condition holds as we approach the $s = b$ border from either "side"; that is, as we approach the border from within the no-trading region as well as when we approach it from within the trading region (see, for example, [9] and [12]).

As discussed in the introduction, it is computationally much more desirable to work with the adjoint equation for the probability density, $\tilde{p}$, than with the equation for the expected utility, $\tilde{V}$, just determined. This means finding the forward Kolmogorov equation, a.k.a. the Fokker-Planck equation, that corresponds to this backward Kolmogorov equation. To speed computations, we seek as simple a form as possible for this adjoint equation. This will be aided by replacing the (s, b, c, t) variables with the following (x, y, z, τ) variables

$$\begin{aligned} x &= \ln s \\ y &= \ln b \\ z &= c \\ \tau &= \frac{\sigma^2}{2} t, \end{aligned}$$

then defining the expected utility, V, and the probability density, p, in these new variables:

$$\begin{aligned} V(x, y, z, \tau, n) &= \tilde{V}(s, b, c, t, n) \\ p(x, y, z, \tau, n) &= \tilde{p}(s, b, c, t, n). \end{aligned}$$

The new variables simplify the backward Kolmogorov equation (2) to the constant coefficient equation

$$V_\tau + aV_x + V_{xx} = 0, \tag{5}$$

where

$$a = \frac{\mu - \frac{\sigma^2}{2}}{\frac{\sigma^2}{2}}.$$

The domain of this PDE is $x > y$, $y \in (-\infty, \infty)$, $z \in (-\infty, \infty)$, $\tau \in \left[0, \frac{\sigma^2}{2}T\right]$. The boundary condition (4) at $x = y$ becomes

$$V_y = G(y, \tau)V_z, \tag{6}$$

where $G(y, \tau)$ is defined by

$$G(y, \tau) = N\left(e^y\right) Re^{\left[y + r\left(T - \frac{2\tau}{\sigma^2}\right)\right]}.$$

To compute the adjoint equation we begin with the basic fact that

$$\frac{d}{dt} E\left[\tilde{V}\right] = 0, \tag{7}$$

where the expected value is conditioned on the choice of n and its corresponding initial conditions. Since

$$E\left[\tilde{V}\right] = \int_{-\infty}^{\infty} \int_{0}^{\infty} \int_{b}^{\infty} \tilde{V}(s, b, c, t, n)\tilde{p}(s, b, c, t, n) ds db dc,$$

transforming (7) into our new variables yields

$$\frac{d}{d\tau} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{y}^{\infty} V(x, y, z, \tau, n)p(x, y, z, \tau, n)e^x e^y dx dy dz = 0$$

where $e^x e^y$ is the Jacobian of the transformation. We can bring the (scaled) time differential operator inside the integral (which is justified since the resulting integrand is continuous) and then apply the product rule to obtain

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{y}^{\infty} \left(V_\tau p + V p_\tau\right) e^x e^y dx dy dz = 0,$$

or, substituting the backward Kolmogorov equation (5),

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{y}^{\infty} \left(-aV_x p e^x - V_{xx} p e^x + V p_\tau e^x\right) e^y dx dy dz = 0.$$

Now we perform the key step of applying integration by parts to transfer each x derivative from V to p. In the resulting expression, the boundary terms as x approaches infinity vanish, but, I, the boundary term on the $x = y$ trading boundary remains, so we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{y}^{\infty}\left[(a-1)\,p + (a-2)\,p_x - p_{xx} + p_\tau\right]Ve^x e^y dxdydz + I = 0, \quad (8)$$

where

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^y e^y \begin{bmatrix} (a-1)V(y,y,z,\tau,n)p(y,y,z,\tau,n) \\ +V_x(y,y,z,\tau,n)p(y,y,z,\tau,n) \\ -V(y,y,z,\tau,n)p_x(y,y,z,\tau,n) \end{bmatrix} dydz. \quad (9)$$

The formation of the adjoint equation requires a form of I where all the derivatives on V are transferred to p. Clearly, the V_x term in the second line of (9) contradicts this, but direct integration by parts is not possible without some work. We begin by defining the functions V^{bdry} and p^{bdry} :

$$\begin{aligned} V^{bdry}(y,z,\tau,n) &= V(y,y,z,\tau,n) \\ p^{bdry}(y,z,\tau,n) &= p(y,y,z,\tau,n). \end{aligned} \quad (10)$$

We can't apply integration by parts to V_x or V_y since they involve differentiation in directions off of the $x = y$ trading boundary, but V_y^{bdry} involves differentiation in a direction strictly within the trading boundary, so we can apply integration by parts to it. Application of the chain rule to (10) gives us that

$$\begin{aligned} V_y^{bdry} &= V_x + V_y \\ p_y^{bdry} &= p_x + p_y, \end{aligned} \quad (11)$$

so J, the term of I in (9) with V_x, can be rewritten

$$\begin{aligned} J &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^y e^y V_x(y,y,z,\tau,n)p(y,y,z,\tau,n)dydz \quad (12) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{2y}\left[V_y^{bdry}(y,z,\tau,n) - V_y(y,y,z,\tau,n)\right]p^{bdry}(y,z,\tau,n)dydz. \end{aligned}$$

Since integration by parts with V_y is not possible, but it is with V_z, we apply the boundary condition (6) to obtain

$$J = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{2y}\left[V_y^{bdry}(y,z,\tau,n) - G(y,\tau)V_z(y,y,z,\tau,n)\right]p^{bdry}(y,z,\tau,n)dydz.$$

Now we can perform the desired integration by parts in y and in z yielding:

$$J = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -V^{bdry}\left[2e^{2y}p^{bdry} + e^{2y}p_y^{bdry}\right] + GV^{bdry}e^{2y}p_z^{bdry}dydz,$$

and then applying the chain rule (11) to p_y^{bdry} gives

$$\begin{aligned} J &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} -V^{bdry}\left[2e^{2y}p^{bdry} + e^{2y}\left(p_x + p_y\right)\right] + GV^{bdry}e^{2y}p_z^{bdry}dydz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[-2p - p_x - p_y + Gp_z\right]e^{2y}V\,dydz. \end{aligned}$$

This is our desired form with V factored so, reinserting J back into I, we have

$$\begin{aligned} I &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \begin{bmatrix} (a-1)p \\ -2p - p_x - p_y + Gp_z \\ -p_x \end{bmatrix} Ve^{2y}dydz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(a-3)p - 2p_x - p_y + Gp_z\right]Ve^{2y}dydz, \end{aligned}$$

and substituting I back into (8) gives

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{y}^{\infty} \left[(a-1)\,p + (a-2)\,p_x - p_{xx} + p_\tau\right]Ve^xe^ydxdydz \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(a-3)p - 2p_x - p_y + Gp_z\right]Ve^{2y}dydz. \end{aligned} \tag{13}$$

Since V depends upon U, which is an arbitrary utility function, we must have from (13) that

$$(a-1)\,p + (a-2)\,p_x - p_{xx} + p_\tau = 0 \tag{14}$$

in the no trading region subject to the boundary condition

$$(a-3)p - 2p_x - p_y + Gp_z = 0 \tag{15}$$

on the trading boundary $x = y$.

Finally, we apply the standard transformation

$$u(x,y,z,\tau,n) = e^{\frac{a^2}{4}\tau + \left(1-\frac{a}{2}\right)x}p(x,y,z,\tau,n)$$

to reduce the PDE in (14) into the simple heat equation

$$u_\tau = u_{xx}. \tag{16}$$

This transforms the boundary condition (15) at $x = y$ into

$$2u_x + u_y - G(y,\tau)u_z = -u. \tag{17}$$

Equations (16) and (17) are our desired simplified adjoint equations for the probability density.

We note that because the boundary condition (17) is a first order differential condition, information about u at the boundary is pushed strictly along characteristic curves, which, for (17), means information is pushed along the direction $2\hat{x} + \hat{y} - G(y,\tau)\hat{z}$ where $\hat{x}, \hat{y}$, and $\hat{z}$ represent unit vectors in the x, y, and z directions. We contrast this with the characteristic curves for (6), which push information about V along the direction $\hat{y} - G(y,\tau)\hat{z}$. From the derivation of (6), we know that $\hat{y} - G(y,\tau)\hat{z}$ is the direction of movement for trading, so it is clear from the boundary condition (17), though perhaps surprising, that we are *not* transferring the probability mass u directly along the trading direction when we trade.

Combining all of the above transformations, we can now state our method to determine the best passive strategy, n, formulated in the (x, y, z, τ, n) coordinate system: We use standard one dimensional optimization techniques to find the n that maximizes

$$\begin{aligned} & V\left(\ln(S(0)), \ln(B(0)), 0, 0, n\right) \qquad (18) \\ = & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{y}^{\infty} U\left(\begin{array}{c} c^{old} + c^{init} + z \\ + \int_{\eta=y}^{\eta=-\infty}\left[e^x - R^{liq}\left(e^x - e^\eta\right)\right] dN(e^\eta) \end{array}\right) \\ & \times \left(e^{-\frac{a^2}{4}\frac{\sigma^2}{2}T + \frac{a}{2}x + y}\right) u\left(x, y, z, \frac{\sigma^2}{2}T, n\right) dx dy dz, \end{aligned}$$

where the probability density $u\left(x, y, z, \frac{\sigma^2}{2}T, n\right)$ is determined by solving the heat equation (16) up to $\tau = \frac{\sigma^2}{2}T$ subject to the trading boundary condition (17) and the initial condition

$$u(x, y, z, 0, n) = \frac{1}{S(0)^{\left(\frac{a}{2}\right)}B(0)}\delta\left(x - \ln(S(0))\right)\delta\left(y - \ln(B(0))\right)\delta\left(z\right), \tag{19}$$

where $\delta(\cdot)$ is the Dirac delta. This form of the initial condition follows from the original form of the initial condition: $\tilde{p}(s, b, c, 0, n) = \delta(s - S(0))\delta(b - B(0))\delta(c)$.

3 Extensions of the Model

3.1 Model for Multiple Stocks

Much of the logic for the single stock case applies to the case of m stocks, therefore, we only focus on where the analysis differs and on the main equations. Obviously, s, b, n, N, etc. are now vectors each with m components which we denote by subscripts (e.g., s_i where $i = 1, 2, ..., m$). Therefore, c^{init}, the cash in $t = T$ dollars generated by the rebalance, and $N_i(b_i)$, the purchasing history of stock i just after the rebalance, are given by

$$c^{init} = e^{rT} \sum_i \left[\begin{array}{c} -\max[0, n_i] S_i(0) + \int_{S_i(0)}^{B_i(0)} [S_i(0) - R(S_i(0) - \beta_i)] \, dN_i^{old}(\beta_i) \\ +[S_i(0) - R(S_i(0) - B_i(0))] \left(\min[n_i, 0] - N_i^{old}(B_i(0)) \right) \end{array} \right]$$

and

$$N_i(b_i) = \left\{ \begin{array}{cc} 0 & \text{if } b_i \geq B_i(0) \\ N_i^{old}(b_i) + n_i & \text{if } b_i < B_i(0). \end{array} \right.$$

(Note that all summations in this section — such as the sum over i in the c^{init} formula — are from 1 to m.)

We assume the standard correlated geometric Brownian motion model for the evolution of the stock prices,

$$dS_i = \mu_i S_i dt + \sum_j \sigma_{ij} S_i dw_j,$$

where σ_{ij} are the constant components of the volatility matrix and w_j are independent Brownian motions. For $m = 1$, scaling time removes the coefficient in front of the second derivative in the backwards Kolmogorov equation, but for $m > 1$, there are $\frac{m(m+1)}{2}$ coefficients to consider so time scaling cannot have the same effect. Therefore, in our scaled variables, we no longer scale time:

$$\begin{aligned} x_i &= \ln s_i \\ y_i &= \ln b_i \\ z &= c \\ \tau &= t, \end{aligned}$$

which leads to the backwards Kolmogorov equation

$$0 = V_\tau + \sum_i a_i V_{x_i} + \sum_{i,j} \gamma_{ij} V_{x_i x_j},$$

where

$$\begin{aligned} a_i &= \mu_i - \frac{1}{2}\sum_j \sigma_{ij}\sigma_{ij} \qquad \text{and} \\ \gamma_{ij} &= \frac{1}{2}\sum_k \sigma_{ik}\sigma_{jk}. \end{aligned}$$

Note that γ_{ij} are the components of a symmetric, positive semi-definite matrix. We will make the small additional assumption that this matrix is positive definite so, in particular, it is invertible.

The boundary condition on the $x_i = y_i$ hyperplane is

$$V_{y_i} = G_i(y_i, \tau)V_z,$$

where

$$G_i(y_i, \tau) = N_i\left(e^{y_i}\right) Re^{y_i + r(T-\tau)}.$$

This holds as we approach the hyperplane from $x_i < y_i$. More importantly, it holds as we approach the hyperplane from $x_i > y_i$ provided we are in the no trading region (defined by the set where $x_j > y_j$ for $j = 1, 2, ..., m$), so it can be used as a boundary condition for the backwards Kolmogorov equation on the boundaries of the no trading region. Although it has no effect on our analysis, it is worth noting that the condition does not hold as we approach a point on the hyperplane from $x_i > y_i$ if that point is in the interior of the trading region (so, for the point, we have that $x_j < y_j$ for at least one $j \neq i$). While, at these points, the derivative of V is discontinuous in directions transverse to the $x_i = y_i$ hyperplane, we note they are continuous in directions that remain within the hyperplane.

This continuity becomes important in the next part of the analysis. As before, we use integration by parts to transform the backwards Kolmogorov equation into its adjoint, the forward Kolmogorov equation. The term on the boundary $x_i = y_i$ that corresponds in the single stock case to J, defined in (12), now contains derivatives in all the x_j directions ($j = 1, 2, ..., m$). Since the derivative of V is continuous in directions within the hyperplane, integration by parts can be applied to all the derivatives in x_j where $j \neq i$. In the direction where $j = i$, we proceed as in the single variable case. This yields the forward Kolmogorov PDE

$$\sum_i (a_i - \sum_j \gamma_{ij})p + \sum_i \left((a_i - 2\sum_j \gamma_{ij})p_{x_i} \right) - \sum_{i,j} \left(\gamma_{ij} p_{x_i x_j} \right) + p_\tau = 0$$

and, at the boundary of the no trading region where $x_i = y_i$,

$$a_i p - 2\sum_j \Big(\gamma_{ij}(p_{x_j} + p)\Big) + \gamma_{ii}\left(-p - p_{y_i} + G_i(y_i,\tau)p_z\right) = 0.$$

To transform our PDE to a form without the p and p_{x_i} terms, we apply the transformation

$$u\left(\underline{x}, \underline{y}, z, \tau, \underline{n}\right) = e^{\frac{1}{4}\sum_{i,j}\gamma_{ij}^{-1}a_i a_j \tau + \sum_i \left(\left(1-\frac{1}{2}\sum_j \gamma_{ij}^{-1}a_j\right)x_i\right)} p\left(\underline{x}, \underline{y}, z, \tau, \underline{n}\right),$$

which yields our simplified final form of the forward Kolmogorov PDE in the no trading region

$$u_\tau = \sum_{i,j} \gamma_{ij}^{-1} a_j \gamma_{ij} u_{x_i x_j} \tag{20}$$

along with the boundary condition at $x_i = y_i$

$$\frac{2}{\gamma_{ii}}\sum_j \gamma_{ij}u_{x_j} + u_{y_i} - G_i(y_i,\tau)u_z = -u, \quad \text{where } i = 1, 2, ..., m \tag{21}$$

and the initial condition

$$u\left(\underline{x}, \underline{y}, z, 0, \underline{n}\right) = \delta(z)\prod_{i=1}^{m}\left[\frac{\delta\left(x_i - \ln\left(S_i(0)\right)\right)\delta\left(y_i - \ln\left(B_i(0)\right)\right)}{\left(S_i(0)\right)^{\frac{1}{2}\sum_j \gamma_{ij}^{-1}a_j} B_i(0)}\right], \tag{22}$$

where $\underline{x}$, $\underline{y}$, and $\underline{n}$ are m-vectors. Therefore, our goal is to find the passive strategy vector $\underline{n}$ that maximizes

$$\begin{aligned}
& V\left(\ln(S_1(0)), \cdots, \ln(S_m(0)), \ln(B_1(0)), \cdots \ln(B_m(0)), 0, 0, \underline{n}\right) \\
= \; & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\int_{y_m}^{\infty}\cdots\int_{y_1}^{\infty} \\
& U\left(\begin{array}{c} c^{old} + c^{init} + z \\ + \sum_i \int_{\eta_i = y_i}^{\eta_i=-\infty}\left[e^{x_i} - R^{liq}\left(e^{x_i} - e^{\eta_i}\right)\right] dN_i(e^{\eta_i}) \end{array}\right) \\
& \times \left(e^{-\frac{1}{4}\sum_{i,j}\gamma_{ij}^{-1}a_i a_j T + \frac{1}{2}\sum_j \gamma_{ij}^{-1}a_j x_i + y_i}\right) u\left(\underline{x}, \underline{y}, z, T, \underline{n}\right) \\
& dx_1 \cdots dx_m dy_1 \cdots dy_m dz,
\end{aligned}$$

where $u\left(\underline{x}, \underline{y}, z, T, \underline{n}\right)$ is determined by solving the PDE (20) subject to conditions (21) and (22).

Note that each component of $\underline{n}$ in the analysis above is assumed to be non-negative. Of course, unlike the single stock case, the optimal strategy for multiple stock portfolios may require some stock positions to be short instead of long. To accommodate short positions, which correspond to components of $\underline{n}$ that are negative, we must switch the model so that, for each of these short positions, $B_i(t)$ now equals the highest, instead of the lowest, price at which any stock i in the portfolio at time t was originally sold, instead of bought. The trading strategy now, of course, is to buy and resell stock i only when $S_i(t) > B_i(t)$ to collect capital losses. We can continue to alter the long position formulation in this symmetric fashion to obtain the formulation of the model with short positions, since we assume the tax rate, R, is the same for all capital gains and losses. However, if we distinguish between various capital gain and capital loss tax rates, as we do in Subsection 5.1.4, the symmetry may fail in small ways that are easily addressed. In the American tax system, for example, the short term, as opposed to long term, capital gains rate applies to gains derived from liquidating any short position. This difference in tax treatment, however, is easily accommodated by adjusting the values of R and R^{liq} at appropriate places in the model, using an approach similar to the one taken in Subsection 5.1.4.

3.2 Probability Distribution for the Liquidation Time of the Portfolio

Since our algorithm first computes the probability distribution, u, instead of directly computing the expected portfolio utility, V, we can easily accommodate a probability distribution for the liquidation time of the portfolio. If liquidation is to occur at the time of the investor's death when all capital gains are forgiven, incorporating the probability distribution into our model allows us to work with available life expectancy data in determining the optimal initial investment strategy.

Let $f(\tau)$ be the probability density function for the (scaled) time of liquidation, $\tau = \frac{\sigma^2}{2}t$. Previously, T represented the (unscaled) time of liquidation, and cash was expressed in T dollars. Since we want to keep the convention of expressing cash in T dollars where T is fixed, but the time of liquidation is now a random variable, we let T represent any specific fixed time. For example, we might choose T to be the expected liquidation time $\frac{2}{\sigma^2}\int_0^\infty \tau f(\tau)d\tau$ or we might choose T to be 0.

With our T fixed, we look to find the n that maximizes the following generalization of (18):

$$\begin{aligned} & V\left(\ln(S(0)), \ln(B(0)), 0, 0, n\right) \qquad (23) \\ = \; & \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_y^\infty U \left(\begin{array}{c} (c^{old} + c^{init} + z) e^{r\left(\frac{2}{\sigma^2}\tau - T\right)} \\ + \int_{\eta=y}^{\eta=-\infty} \left[e^x - R^{liq}\left(e^x - e^\eta\right)\right] dN(e^\eta) \end{array} \right) \\ & \times \left(e^{-\frac{a^2}{4}\tau + \frac{a}{2}x + y} \right) u\left(x, y, z, \tau, n\right) f(\tau) dx dy dz d\tau \end{aligned}$$

The computation of u proceeds exactly as before, but, of course, we now stop the computation at a value of τ where $f(\tau)$ becomes so small that further computation would have no observable effect on the computation of V in (23).

3.3 Buying Stock with Tax Money Culled from Losses

An investor may decide to use all the proceeds from tax losses to buy more stock instead of keeping the proceeds as cash. This makes sense if a passive investor wishes to keep the ratio of stock to cash closer to the initial ratio, since the presence of losses means the stock has lost value while the cash, as always, has gained value. The model detailed in Section 2 for keeping losses as cash can be converted to a model for keeping losses as stock using the following changes.

Since no money is generated by losses in this model, the variable $C(t)$ is no longer helpful since $C(t) = 0$. Instead, we require a new variable $N^a(t)$, which represents the *a*ccumulated number of shares of stock bought using tax losses during the time interval $(0, t]$. The purchasing history at time t now becomes $N(b) + N^a(t)$ if $b < B(t)$ and, as before, 0 if $b \geq B(t)$.

Since C is no longer useful and N^a is, we replace $\tilde{V}(s, b, c, t, n)$, the old function for the expected portfolio utility at the liquidation time T, with the new function $\tilde{V}^a(s, b, n^a, t, n)$.

As in the analysis where losses convert to cash, the expected utility of the portfolio before and after culling tax losses is equal, so the analog of (3) for converting losses to stock is

$$\tilde{V}^a\left(s, s + \Delta b, n^a, t, n\right) = \tilde{V}^a\left(s, s, n^a + \frac{R}{s}\int_s^{s+\Delta b} N(b) + n^a db, t, n\right). \quad (24)$$

Now we take the limit of (24) as $\Delta b \to 0$ to obtain the analog of equation (4), the first order differential condition on the $s = b$ boundary:

$$\tilde{V}^a_b = \tilde{V}^a_{n^a} \frac{R}{b}(N(b) + n^a). \tag{25}$$

As before we transform the independent variables. The transformations to (x, y, τ) from (s, b, t) are exactly as before, but now we substitute z^a for n^a just as we substituted z for c before. In the new variable coordinate system, the boundary condition (25) becomes

$$\tilde{V}^a_y = G^a(y, z^a)\tilde{V}^a_z, \tag{26}$$

where $G^a(y, z^a) = R\,(N(e^y) + z^a)$.

Now we can proceed exactly as before, and we get essentially the same equations for u^a, the probability density function for converting to stock, as we did in (16) and (17) for u, the probability density function for converting to cash. Specifically, we get the heat equation

$$u^a_\tau = u^a_{xx}$$

in the no-trading region subject to the boundary condition

$$2u^a_x + u^a_y - G^a(y, z^a)u^a_z = -u^a$$

at $x = y$. The initial condition for u^a is also identical to (19), the initial condition for u. We just replace z with z^a.

All that remains is to update (18), the form of V that must be maximized. There are two changes in the update. First, $z = 0$ since no cash has accumulated. Second, we must liquidate the accumulated stock, all of which was bought at the lowest stock price in the time interval $(0, T)$, which, of course, corresponds to the value of y. This means the form of V^a to be maximized is

$$\begin{aligned} & V^a\left(\ln(S(0)), \ln(B(0)), 0, 0, n\right) \\ = \; & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{y}^{\infty} U\left(\begin{array}{c} c^{old} + c^{init} \\ + \int_{\eta=y}^{\eta=-\infty}\left[e^x - R^{liq}\left(e^x - e^\eta\right)\right] dN(e^\eta) \\ + n^a\left(e^x - R^{liq}\left(e^x - e^y\right)\right) \end{array} \right) \\ & \times \left(e^{-\frac{a^2}{4}\frac{\sigma^2}{2}T + \frac{a}{2}x + y}\right) u\left(x, y, z^a, \frac{\sigma^2}{2}T, n\right) dx dy dz^a. \end{aligned} \tag{27}$$

By combining the method here and in Section 2, it is straightforward to develop models in which losses are partially converted to cash with the remainder converted to stock. However, since both c and n^a must be used in such a model, the PDE gains a dimension, making computation more difficult.

4 Computation of the Optimal Strategy for the Single Stock Model

4.1 Numerical Algorithm

In this subsection we detail a dynamic finite difference method that simulates the solution to the heat equation (16) given the initial condition (19) and the boundary condition (17). We define the numerical approximation to this PDE's solution on a four dimensional rectangular mesh by the standard notation

$$u^n_{i,j,k} \approx u(i\Delta x, j\Delta y, k\Delta z, n\Delta\tau),$$

where i, j, k and n are integers. (Note that in this section n is no longer used to denote the initial stock purchase and, further, we suppress showing the dependence of u on this purchase.) We leave Δx as a free parameter, and then choose $\Delta t < \frac{1}{2}(\Delta x)^2$, the classic stability condition to the simple forward difference scheme we use for the heat equation:

$$u^{n+1}_{i,j,k} = u^n_{i,j,k} + \frac{\Delta t}{\Delta x^2}\left(u^n_{i-1,j,k} - 2u^n_{i,j,k} + u^n_{i+1,j,k}\right).$$

The Δy and Δz are determined by examining the boundary condition (17). The general solution to the boundary condition equation is

$$u\left(x_0 + 2s, y_0 + s, z_0 - G(y_0 + s, \tau_0)s, \tau_0\right)e^s = u\left(x_0, y_0, z_0, \tau_0\right), \tag{28}$$

where s parameterizes the characteristic path of the solution. By setting $s = \Delta x$ and choosing $\Delta y = \Delta x$, we ensure that both the x and y values at which the left hand side is evaluated correspond to mesh points. With these choices, to keep the CFL stability condition for the algorithm satisfied, we must make sure that Δz is chosen so that

$$\Delta z \geq G(y_0 + \Delta x, \tau_0)\Delta x$$

always holds. This is accomplished by letting $\Delta z = \gamma \Delta x$ where

$$\begin{aligned}
\gamma &= \max_{y,\tau} G(y + \Delta x, \tau) \\
&= \max_{y,\tau} \left[N\left(e^{y+\Delta x}\right) Re^{\left[y+\Delta x + r\left(T - \frac{2\tau}{\sigma^2}\right)\right]} \right] \\
&= N(0)\, RB(0) e^{rT + \Delta x},
\end{aligned}$$

since $y = \ln(B(t)) \leq \ln(B(0))$ and $N(0)$ is the total amount of stock in the portfolio after the initial rebalance.

We use the boundary condition equation's solution (28) to determine $u^n_{j-1,j,k}$, the value of the approximated solution at the mesh points just behind the boundary $x = y$. Since the characteristic curve emanating from any one of these mesh points passes at $s = \Delta x$ directly between two mesh points with different z values, we compute $u^n_{j-1,j,k}$ using a weighted average of the computed solution between these two points, where the weights depend upon the proximity of the characteristic to each point. This yields our boundary condition

$$u^n_{j-1,j,k} = \left[\begin{array}{c} \left(1 - \frac{G((j+1)\Delta x, n\Delta t)}{\gamma}\right) u^n_{j+1,j+1,k} \\ + \frac{G((j+1)\Delta x, n\Delta t)}{\gamma} u^n_{j+1,j+1,k-1} \end{array} \right] e^{\Delta x}. \tag{29}$$

The initial condition is treated in a similar manner. Since the location of the Dirac delta's mass generally lies between three mesh points with differing x and y values, we distribute the mass among the three points again with weights that depend upon the location's proximity to the three points. Attempting to replace this three mesh point method by instead putting the full mass at the single closest mesh point can introduce significant error.

We note that our program's results cannot be better than first order accurate in Δx since the N function is discontinuous (in general), which induces an error of order Δx in any integration scheme for approximating the expected utility in (18). Since this is the case, we employ a simple Riemann sum to approximate this final expected utility integral.

Because the heat equation only involves diffusion in the x direction, we compute the evolution of the heat equation along a dynamically sized x array. This array doubles in size whenever the computed solution at either end of the array becomes larger than a chosen threshold for x growth. In the experiments that follow, we select this threshold for x growth to be 10^{-25}, which saves significant computational time and has no effect on the final computed utility.

The heat equation is run on separately sized x arrays for each "alive" yz node. We start with two alive yz nodes, which correspond to the two y values of the three mesh points used for the initial condition along with $z = 0$. At the end of each time step, the boundary condition (29) is checked and updated. If the left hand side of (29) is greater than a chosen threshold for yz growth and corresponds to a currently unused yz value, then we create a new alive yz node for it. This is added to a linked list of the other currently alive yz nodes.

Experimentally, the threshold for yz growth can be quite important. For the base case described in the next subsection, we use a threshold for yz growth of 10^{-12}. If we try to use a threshold of 10^{-11}, we introduce significant noise in the second significant digit of our results. If we use a threshold of 10^{-13}, the calculation takes 50% longer while only slightly affecting the third significant digit of our results. Since we use 10^{-12}, all results in the next section are expressed with two significant digits. Alterations of the base case, particularly alterations to the utility function, can require a decrease in the value of the threshold for yz growth, as we detail further in the next subsection.

Finally, we note that we use successive parabolic interpolation to locate the strategy that optimizes the final expected utility. Experimentally, the parabolic interpolation usually converges to the optimal strategy within a few iterations.

4.2 Computational Results

Although any utility function is easily accommodated by our algorithm, our examples in this section will consider power law utility functions:

$$U(W) = \frac{W^{1-\rho}}{1-\rho}.$$

Here W is the worth of the portfolio and ρ, which is a positive number not equal to 1, measures risk aversion. Power law utility functions correspond to constant relative risk aversion (CRRA). When there are no taxes, we know from Merton [15] that the optimal strategy for power law utilities is to continuously rebalance the portfolio so that the fraction of the portfolio in stock remains constant for all time. This optimal fraction, f^{opt}_{Merton}, is given by

$$f^{opt}_{Merton} = \frac{\mu - r}{\rho\sigma^2}. \tag{30}$$

Since we want to compare our optimal results in the taxable case to Merton's result, we will express our control n in terms of f, the fraction of the portfolio in stock just after the $t = 0$ rebalance. For a portfolio with no stock just prior to the $t = 0$ rebalance, f and n are related by

$$n = f \frac{c^{old} e^{-rT}}{S(0)}.$$

In this context c^{old} is the initial worth of the portfolio in $t = T$ dollars, so $c^{old} e^{-rT}$ is the initial worth in $t = 0$ dollars.

We next detail a base case. We will analyze this case and alterations of it. For this base case, we consider a portfolio that, just prior to $t = 0$, has no stock and \$50,000 in cash in $t = 0$ dollars. The fact that power law utilities correspond to constant relative risk aversion will mean that the results in this subsection will be the same if we replace \$50,000 with any other amount. The value, \$50,000, will matter, however, at times in Section 5. We start with an initial stock price $S(0) = 100$ and so, once stock is bought at $t = 0$, we also have that $B(0) = 100$. The specific value, 100, has no effect on our results, only the fact that $S(0) = B(0)$.

For our base case, we assume that the stock evolves with an expected return and volatility of $\mu = 0.07 \frac{1}{\text{yr}}$ and $\sigma = 0.20 \frac{1}{\sqrt{\text{yr}}}$. We assume the risk-free, tax-exempt interest rate $r = 0.04 \frac{1}{\text{yr}}$. We assume a 30% combined federal and state tax rate, so $R = R^{liq} = 0.3$. We consider a liquidation time of $T = 40$ years. Finally, we use a risk aversion corresponding to $\rho = 0.9$ in our utility function so the optimal Merton fraction is $f^{opt}_{Merton} = 0.83$.

For this base case, we apply our PDE based algorithm from the previous subsection to compute f^{opt}, the optimal f, for each of four decreasing values for Δx. In each case, converging to the optimal fraction required evaluating the expected utility at four values of f, so a single run takes one fourth the computational time presented in the following table:

Δx	f^{opt}	computation time (in seconds)
0.10	0.8247	7
0.08	0.8302	16
0.06	0.8322	58
0.04	0.8324	352

Since the computation time is of order $\frac{1}{(\Delta x)^5}$, the extreme sensitivity of the computational time to Δx is expected. For all subsequent calculations, we use $\Delta x = 0.08$, since our chosen threshold for yz growth, 10^{-12}, already restricts us to two significant digits for f^{opt}. The algorithm was programmed in C++ and run on an Apple iMac desktop computer with a 3.06 GHz Intel Core 2 Duo processor and 3 MB of RAM.

By removing the effect of the culled losses in our algorithm, we can also find f^{opt}_{wol}, the optimal fraction *w*ith*o*ut *l*osses; that is, assuming we never buy or sell stock at any $t \in (0, T)$ even when the stock is at a loss. We know that $f^{opt}_{wol} < f^{opt}$ since realized losses get converted to cash, but we can now quantify this difference. We compute that $f^{opt}_{wol} = 0.70$, so we have a significant shift: we should place 70% of our initial portfolio in stocks if we will not take advantage of losses, but if we do take advantage of losses that percentage jumps up to 83%. The fact that f^{opt} and f^{opt}_{Merton} are both 0.83 is, of course, coincidence, however, we will see that, despite the significant differences in both their meaning and their underlying models, f^{opt} and f^{opt}_{Merton} generally also stay reasonably close to each other when the base case is altered, as will be shown in the tables below.

In these tables, for a given initial stock fraction f, we express the expected utility of the portfolio worth after liquidation, that is, $E[U(W)] = \tilde{V} = V$, in terms of the following *normalized certainty equivalent*:

$$\mathcal{C}(f) = \frac{U^{-1}\left(E[U(W)]\right)}{c^{old}}. \tag{31}$$

Clearly, $\mathcal{C}(f)$ is maximized at $f = f^{opt}$. At $t = 0$ an investor with a given utility preference is indifferent to having the portfolio with the fraction f in stock versus being guaranteed at time T having the certainty equivalent, $U^{-1}\left(E[U(W)]\right)$. Therefore, $\mathcal{C}(f)$ measures the advantage, from the investor's viewpoint, of a portfolio with a fraction f chosen in stock initially versus having to keep the same portfolio strictly in cash. Since we are employing power law utilities, $\mathcal{C}(f)$ is independent of the initial portfolio worth. The tables below for alterations to the base case provide $\mathcal{C}(f^{opt})$ and $\mathcal{C}(0.83)$, since $f^{opt} = 0.83$ for the base case. By definition, $\mathcal{C}(f^{opt}) \geq \mathcal{C}(0.83)$. When $\mathcal{C}(0.83) < 1$, having the portfolio completely in cash, that is, $f = 0$, is preferable to having the primarily stock oriented portfolio where $f = 0.83$.

The tables provide two measures of the advantage, or premium, that reaping losses generates for the portfolio with an inital stock fraction, f.

The first measure is the *dollar premium*:

$$\mathcal{D}^{prem}(f) = \frac{E[W] - E^{wol}[W]}{c^{old}}. \tag{32}$$

Here $E[W]$ is the expected worth of the portfolio if we harvest losses and $E^{wol}[W]$ is the expected worth of the portfolio if we do not harvest losses. Therefore, the dollar premium is the (normalized) expected difference in cash at time T due to our investing harvested losses in cash. Once the dollar premium is determined for one value of f, we can easily determine the dollar premium any other value of f from the fact that $\frac{\mathcal{D}^{prem}(f_2)}{\mathcal{D}^{prem}(f_1)} = \frac{f_2}{f_1}$.

The second measure of the premium generated by losses takes into account the investor's risk preferences. Here we use the difference in the normalized certainty equivalents with and without harvesting losses to define the *certainty equivalent premium*:

$$\mathcal{C}^{prem}(f) = \frac{U^{-1}\left(E[U(W)]\right) - U^{-1}\left(E^{wol}[U(W)]\right)}{c^{old}}.$$

Because we are using power law utilities, both $\mathcal{D}^{prem}(f)$ and $\mathcal{C}^{prem}(f)$ are independent of the value of c^{old}.

In the calculations below, five premia are calculated. We first determine $\mathcal{D}^{prem}(0.83)$, the dollar premium at $f = 0.83$, to give a constant stock fraction comparison. We calculate $\mathcal{D}^{prem}(0.83)$ for both $R^{liq} = R$ and $R^{liq} = 0$ to quantify the benefit of having $R^{liq} = 0$, where there are no capital gains at liquidation to offset the earlier use of capital losses. We then calculate $\mathcal{D}^{prem}(f^{opt})$, the dollar premium at the recommended stock fraction f^{opt}, when $R^{liq} = R$. Of course, as f^{opt} decreases, the dollar premium also decreases, since the portfolio has less stock. Finally, at $R^{liq} = R$, we compute $\mathcal{C}^{prem}(f^{opt})$ and $\mathcal{C}^{prem}(0.83)$ for comparison with their dollar premium counterparts.

We now alter each parameter away from its base case value to understand its effect on the three optimal fractions (f^{opt}, f^{opt}_{wol}, and f^{opt}_{Merton}), on the normalized certainty equivalents ($\mathcal{C}(f^{opt})$ and $\mathcal{C}(0.83)$), and on the five premia just described.

We begin by altering the expected return, μ:

μ	f^{opt}	f^{opt}_{wol}	f^{opt}_{Merton}	$\mathcal{C}(f^{opt})$	$\mathcal{C}(0.83)$
0.055	0.24	0.17	0.42	1.04	0.92
0.060	0.43	0.33	0.55	1.10	1.04
0.065	0.63	0.52	0.69	1.21	1.19
0.070	0.83	0.70	0.83	1.37	1.37

μ	$\mathcal{D}^{prem}(0.83)$ $(R^{liq}=R)$	$\mathcal{D}^{prem}(0.83)$ $(R^{liq}=0)$	$\mathcal{D}^{prem}(f^{opt})$ $(R^{liq}=R)$	$\mathcal{C}^{prem}(f^{opt})$ $(R^{liq}=R)$	$\mathcal{C}^{prem}(0.83)$ $(R^{liq}=R)$
0.055	5.0%	6.6%	1.4%	1.6%	7.6%
0.060	4.8%	6.2%	2.5%	3.0%	7.5%
0.065	4.5%	5.9%	3.4%	4.8%	7.4%
0.070	4.3%	5.5%	4.3%	7.3%	7.3%

As μ increases to its base case value, the three optimal fractions and $\mathcal{C}(f^{opt})$ increase to reflect the increased power of the stock, while the three premia at $f = 0.83$ decrease since there are fewer losses. The two premia at f^{opt} increase with μ, since the increase in f^{opt} has a stronger effect than the diminished loss per stock.

Next we consider the effect of the volatility, σ:

σ	f^{opt}	f^{opt}_{wol}	f^{opt}_{Merton}	$\mathcal{C}(f^{opt})$	$\mathcal{C}(0.83)$
0.19	0.91	0.79	0.92	1.41	1.41
0.20	0.83	0.70	0.83	1.37	1.37
0.25	0.50	0.40	0.53	1.19	1.14
0.30	0.31	0.20	0.37	1.09	0.95

σ	$\mathcal{D}^{prem}(0.83)$ $(R^{liq}=R)$	$\mathcal{D}^{prem}(0.83)$ $(R^{liq}=0)$	$\mathcal{D}^{prem}(f^{opt})$ $(R^{liq}=R)$	$\mathcal{C}^{prem}(f^{opt})$ $(R^{liq}=R)$	$\mathcal{C}^{prem}(0.83)$ $(R^{liq}=R)$
0.19	3.9%	5.0%	4.3%	7.7%	6.5%
0.20	4.3%	5.5%	4.3%	7.3%	7.3%
0.25	6.3%	8.2%	3.8%	5.1%	10.8%
0.30	8.2%	10.8%	3.1%	3.6%	13.5%

As σ increases, the utility function penalizes the increased uncertainty, lowering the three optimal fractions and $\mathcal{C}(f^{opt})$, while the three premia at $f = 0.83$ increase since losses now become more likely. The two premia at f^{opt} decrease as σ increases, because, as before, the effect from f^{opt} is stronger than the effect from the losses per stock.

The tax rate R has no effect on the optimal Merton fraction formula (30) so, $f^{opt}_{Merton} = 0.83$ for all values of R.

R	f^{opt}	f^{opt}_{wol}	$\mathcal{C}(f^{opt})$	$\mathcal{C}(0.83)$
0.20	0.87	0.79	1.46	1.46
0.30	0.83	0.70	1.37	1.37
0.40	0.76	0.59	1.27	1.26
0.50	0.64	0.45	1.17	1.16

R	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = R)$	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = 0)$	$\mathcal{D}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(0.83)$ $(R^{liq} = R)$
0.20	2.9%	3.7%	3.0%	5.6%	5.1%
0.30	4.3%	5.5%	4.3%	7.3%	7.3%
0.40	5.7%	7.4%	5.2%	7.9%	9.1%
0.50	7.1%	9.2%	5.5%	7.4%	10.7%

The increased tax rate makes stocks less attractive, but has no effect on the tax-free cash growth. Therefore $\mathcal{C}(f^{opt})$ and the two optimal fractions decrease, but the return on losses increases, causing the three premia at $f = 0.83$ to increase. In general, the premia at f^{opt} increase with R, since, unlike before, the increased return on losses is now a stronger effect than the change in f^{opt}. An exception is between $R = 0.40$ and $R = 0.50$ where the large drop in f^{opt} becomes significant enough to cause $\mathcal{C}^{prem}(f^{opt})$ to decrease.

For the risk-free, tax-free interest rate, r, we have

r	f^{opt}	f^{opt}_{wol}	f^{opt}_{Merton}	$\mathcal{C}(f^{opt})$	$\mathcal{C}(0.83)$
0.038	0.91	0.78	0.88	1.46	1.46
0.040	0.83	0.70	0.83	1.37	1.37
0.045	0.61	0.50	0.69	1.19	1.17
0.050	0.39	0.31	0.55	1.09	1.01
0.055	0.20	0.14	0.42	1.03	0.88

r	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = R)$	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = 0)$	$\mathcal{D}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(0.83)$ $(R^{liq} = R)$
0.038	4.2%	5.6%	4.6%	8.6%	7.3%
0.040	4.3%	5.5%	4.3%	7.3%	7.3%
0.045	4.4%	5.5%	3.2%	4.5%	7.2%
0.050	4.5%	5.4%	2.1%	2.5%	7.1%
0.055	4.6%	5.3%	1.1%	1.2%	6.9%

As r increases, cash becomes more valuable, so the three optimal fractions decrease. $\mathcal{C}(f^{opt})$ also decreases because when r increases, the denominator, which corresponds to an all cash portfolio, increases faster than the numerator, which corresponds to a portfolio whose initial cash fraction is only $1 - f^{opt}$. Because both the numerator and denominator of the three premia at $f = 0.83$ correspond completely to cash differences, the general effect of r on these premia is small. However, the slight decrease in $\mathcal{D}^{prem}(0.83)$ when $R^{liq} = 0$ is expected due to the denominator benefiting from the interest over the entire 40 year investment horizon, while the numerator only benefits from the interest after losses are harvested. The smallness of this decrease suggests that most losses are harvested early in the 40 year investment period. We will see more about this phenomenon later. The two premia at f^{opt}, of course, decrease due to the decrease in f^{opt}.

As ρ in the utility function increases, the investor becomes more risk averse. We note that to calculate f^{opt} and f^{opt}_{wol} as ρ increased, we needed to decrease the threshold for yz growth. For example, at $\rho = 5.0$, the threshold was reduced from 10^{-12}, its value in the base case calculations, to 10^{-16}.

ρ	f^{opt}	f^{opt}_{wol}	f^{opt}_{Merton}	$\mathcal{C}(f^{opt})$	$\mathcal{C}(0.83)$
0.8	0.94	0.80	0.94	1.43	1.42
0.9	0.83	0.70	0.83	1.37	1.37
1.1	0.65	0.55	0.68	1.26	1.24
2.0	0.31	0.26	0.38	1.11	0.94
3.0	0.19	0.16	0.25	1.07	0.76
5.0	0.10	0.09	0.15	1.04	0.60

ρ	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = R)$	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = 0)$	$\mathcal{D}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(0.83)$ $(R^{liq} = R)$
0.8	4.3%	5.5%	4.9%	8.7%	6.9%
0.9	4.3%	5.5%	4.3%	7.3%	7.3%
1.1	4.3%	5.5%	3.4%	5.0%	7.5%
2.0	4.3%	5.5%	1.6%	2.1%	9.0%
3.0	4.3%	5.5%	1.0%	1.2%	9.5%
5.0	4.3%	5.5%	0.5%	0.6%	9.4%

As ρ increases and we become more risk averse, the three optimal fractions decrease and the values for $\mathcal{C}(f^{opt})$ converge to 1. Also as ρ increases, f^{opt} vanishes, which causes both $\mathcal{D}^{prem}(f^{opt})$ and $\mathcal{C}^{prem}(f^{opt})$ to converge to zero at about the same rate. On the other hand, as ρ decreases to 0, $\mathcal{C}^{prem}(0.83)$ must, by definition, converge to $\mathcal{D}^{prem}(0.83) = 4.3\%$.

Finally, we consider the effect of the time horizon, T. Since T has no effect on the optimal Merton fraction formula (30), $f^{opt}_{Merton} = 0.83$ for all values of T.

T	f^{opt}	f^{opt}_{wol}	$\mathcal{C}(f^{opt})$	$\mathcal{C}(0.83)$
5	0.70	0.57	1.03	1.03
10	0.72	0.58	1.06	1.06
20	0.77	0.63	1.14	1.14
30	0.80	0.67	1.24	1.24
40	0.83	0.70	1.37	1.37
50	0.86	0.73	1.51	1.51
75	0.90	0.79	1.97	1.96
100	0.93	0.83	2.60	2.58
125	0.96	0.87	3.44	3.38

T	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = R)$	$\mathcal{D}^{prem}(0.83)$ $(R^{liq} = 0)$	$\mathcal{D}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(f^{opt})$ $(R^{liq} = R)$	$\mathcal{C}^{prem}(0.83)$ $(R^{liq} = R)$
5	0.8%	4.3%	0.7%	0.8%	1.0%
10	1.6%	5.0%	1.4%	1.6%	2.0%
20	2.8%	5.4%	2.6%	3.5%	3.9%
30	3.7%	5.5%	3.6%	5.4%	5.7%
40	4.3%	5.5%	4.3%	7.3%	7.3%
50	4.7%	5.6%	4.9%	9.3%	8.7%
75	5.2%	5.6%	5.6%	14.2%	11.8%
100	5.4%	5.6%	6.1%	19.9%	14.6%
125	5.5%	5.6%	6.4%	27.1%	17.2%

Over time, the fact that μ, the expected return for stock, is greater than r, the interest rate for cash, becomes a more important factor than the stock's volatility, so the two optimal fractions in the chart increase with T. Of course $\mathcal{C}(f^{opt})$ increases with T as well. Note from the dollar premium for the $R^{liq} = 0$ case that even by $T = 5$, a significant fraction of the total eventual losses appears to have already been realized. This is the same phenomenon we saw in the r table. We will explore this further in the next section. For $R^{liq} = R$, we can see that the premia diminish quickly as T gets small. This is because the money generated by the losses has little time to grow, so the beneficial effects of taking capital losses are largely cancelled by the corresponding increase in capital gains this creates at the early liquidation time T. On the other hand, when T is large, the losses, as we have observed, primarily occur early on, so they have time to generate interest. For large T, this interest dwarfs the increase in capital gains owed so much later, so the $f = 0.83$ dollar premia for $R^{liq} = R$ and for $R^{liq} = 0$ converge.

5 Approximate Effect of Model Assumptions

In this chapter we use Monte Carlo simulation with a minimum of 100,000 runs to approximate the effect of many of our model's assumptions on capital loss harvesting. We will generally continue to work with our base case, where all stock is purchased at $t = 0$, the liquidation time is $T = 40$ years, and the stock's expected return and volatility are $\mu = 0.07\frac{1}{\text{yr}}$ and $\sigma = 0.20\frac{1}{\sqrt{\text{yr}}}$.

For this base case, the distribution of harvested losses at the end of each of the 40 years is contained in Figure 2. The quantity $1 - B(t)/S(0)$ on the

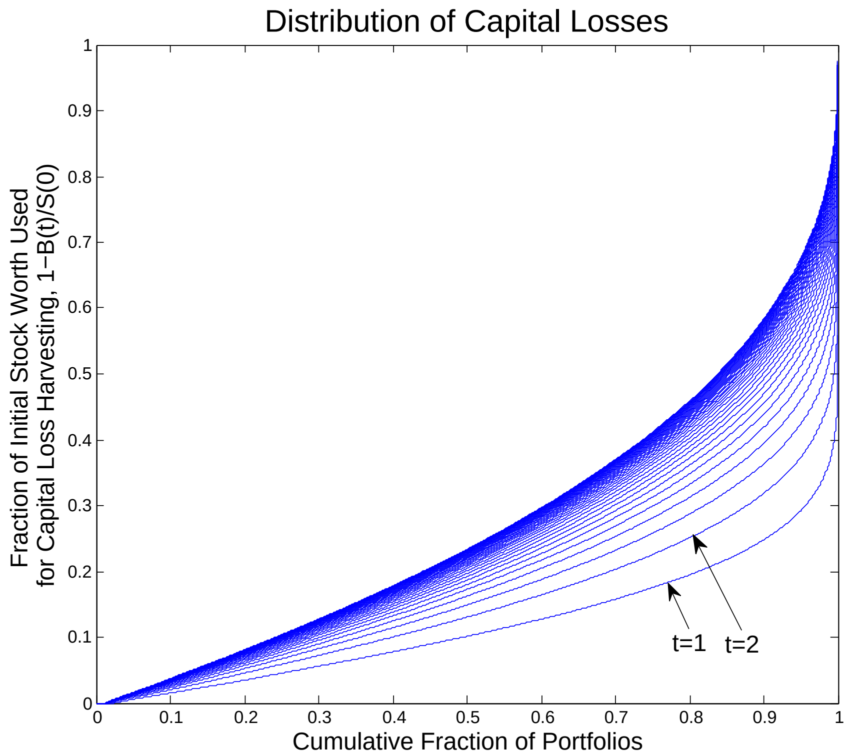

Figure 2: *Distribution of Capital Losses over each of 40 years. For example, a cumulative fraction of 0.8 at* $t = 1$ *corresponds to a ratio,* $1-B(1)/S(0)$*, of approximately 0.19, which means that within the first year, the stock position of 80% of portfolios will never dip below a 19% loss from its initial value. The area under the curve gives the average fraction of losses. Since the area under the top curve is equal to 0.276, we have that the average worst dip in stock worth throughout 40 years is 27.6% from the stock's initial worth.*

y-axis of the graph in Figure 2 is the fraction of the original stock's worth harvested as a loss by time t. The average of this fractional loss is equal to the area under the curve corresponding to time t. In particular, we define F_{avg} to be the value of this fractional loss quantity at the final time T, so, for our base case shown in Figure 2, F_{avg} is the area under the top curve in the graph, which equals 0.276.

From the areas in Figure 2, we see that in the first year, on average, 11.9% of the eventual total harvested losses have already been taken. By the end of the fifth year, this percentage increases to 73.1% of the eventual total harvested losses. This heavy preponderance towards expected losses occurring in the first five years was previously indicated by the $R^{liq} = 0$ column of the table where T varied, as discussed at the end of the previous section. This clustering of trades towards the initial years is due to the long term effects of the expected return, so, as μ increases, these percentages also increase.

The size of the initial stock worth, $nS(0)$, is a major factor in determining the effect of some of our assumptions. According to the Survey of Consumer Finances run by the Federal Reserve Board [16], the median amount of stock held by American households with taxable accounts is in the neighborhood of \$50,000.[3] We will continue to use our base case assumption that the risk-free, tax-exempt interest rate is $r = 0.04\frac{1}{\text{yr}}$, and, at first, we will also continue to assume the base case value $R = 0.3$, which corresponds to a 30% combined federal and state tax rate for both capital gains and losses.

Given these parameters, we will determine factors in Subsection 5.1 that approximate the ramifications of removing many of our model's assumptions. We use these factors to replace the R in (3), which is the beneficial tax rate that generates cash from losses, with an appropriately smaller effective rate, R_{Eff}. We then consider the effect of the fact that, in the current American tax system, the beneficial tax rate for losses, that is, the value of R in (3), is larger than the detrimental tax rate for gains used both at the initial rebalance, which is the value of R in (1), and at liquidation, which is the value of R^{liq}. In Subsection 5.2, we will approximate the effect of rebalancing the portfolio at the halfway time, $t = T/2$, to the same fraction of stock in

[3]Specifically, from the 2007 Survey of Consumer Finances: 17.9% of households have non-retirement stock accounts. These stock accounts have a median worth of \$17,000. Fewer households, 11.4%, have non-retirement pooled investment funds. The median worth of these funds per household is \$56,000. Of these pooled investment funds, approximately 70% of the funds are in stocks.

the portfolio that we selected at $t = 0$.

5.1 The Effect of Wash Sale Rules, Not Trading Continuously, Transaction Costs, the Annual $3000 Limit on Losses and The Difference between Capital Gain and Capital Loss Tax Rates

In 5.1.1 – 5.1.3 below we determine factors, $\mathcal{F}$, that approximate the ramifications of not trading continuously, $\mathcal{F}_{Trade}$, including transaction costs, $\mathcal{F}_{Trans}$, and enforcing the $3000 annual limit on undeferred losses, $\mathcal{F}_{3000}$. The *effective tax rate for capital losses*, R_{Eff}, is computed from these factors by the formula

$$R_{Eff} = \mathcal{F}_{Trade}\mathcal{F}_{Trans}\mathcal{F}_{3000}R, \qquad (33)$$

where R is the actual tax rate for capital losses. If we replace R in (3) with R_{Eff}, we can apply our PDE model as before to obtain an approximate solution.

Our base case assumption is that $R = 0.3$, which was determined by the approximate average of the tax rates for capital gains and capital losses. In 5.1.4, we take into account the fact that the rate for capital losses is higher than the rate for capital gains and examine the increase this creates to the value of R (and therefore R_{Eff}) in (33).

5.1.1 Wash Sale Rules and Not Trading Continuously

Wash sale rules prevent collecting capital losses if an investor sells and rebuys the same stock within 31 days. For index ETFs, the investor cannot sell and rebuy an identical index within 31 days, even if they are different ETFs like IVV and SPY, which both track the S&P 500. However, wash sale rules can be avoided by using similar index funds. Say, for example, an investor uses three similar iShares ETFs to track most of the American domestic stock market: IVV (S&P 500), IWB (Russell 1000) and IWV (Russell 3000). By having all stock holdings in only one of these three ETFs, the investor can completely sell their holdings in one of these ETFs and use all the proceeds to buy a different ETF twice per month without triggering wash sale rules.

By not trading continuously, the investor misses some of the stock dips. The effect of only trading at specific time intervals (yearly, monthly, etc.) on loss harvesting is given by $\mathcal{F}_{Trade}$ in the table below. $\mathcal{F}_{Trade}$ is computed from

100,000 Monte Carlo runs. We take the average over these runs of the money generated in the final portfolio if harvesting losses can only occur at the beginning of an interval and then divide by the average of the money generated with these same runs if we harvest losses continuously. Since continuous trading cannot be exactly simulated, we approximate it by 5000 trading intervals per year. Repeating this method with more runs confirms the three digits of accuracy given in the following table:

Transaction interval	Average number of transactions, A	Factor, $\mathcal{F}_{Trade}$, due to not trading continuously
yearly	2.0	0.673
quarterly	4.6	0.832
monthly	8.6	0.901
twice monthly	12.5	0.931
thrice monthly	15.4	0.947
10 times monthly	28.9	0.973
daily (252 trading days/year)	42.2	0.984

Of course, the shorter the transaction interval is, the less stock dips are missed, so $\mathcal{F}_{Trade}$ gets closer to 1, but also the higher the number of transactions and, therefore, transaction costs. The average number of transactions, A, for each transaction interval is given above. To be conservative in our approximation, we only address the detrimental effect of not trading continuously here, which is through the capital loss tax rate via (33), but not the smaller beneficial effect from the corresponding reduction in capital gains at liquidation. In the case where all stock is purchased at $t = 0$, this reduction at liquidation can be approximated by replacing R^{liq} with $\mathcal{F}_{Trade}R^{liq}$.

5.1.2 Transaction Costs

Transaction costs have two sources: the proportional cost per trade, generally due to the bid-ask spread, and the fixed cost per trade, generally due to commissions. From the analysis for (3), where there are no transaction costs, we know that ΔC, the money in T dollars added into the portfolio from a single harvesting of losses at time t, equals $-n\Delta BRe^{r(T-t)}$. With transaction costs this becomes

$$\Delta C = -n\Delta BRe^{r(T-t)} - (\text{bid-ask spread})\, ne^{r(T-t)} - (\text{fixed cost})e^{r(T-t)}.$$

To get $C(T)$, we sum this expression over all trades. To approximate this sum, we assume the average number of trades, A. We also approximate $e^{r(T-t)}$ by its average value, $\left(e^{r(T-t)}\right)_{avg}$, and we approximate the sum of $-\Delta B$ over all trades by its average value, $F_{avg}S(0)$. Given these approximations, the sum over all ΔC yields

$$\begin{aligned} C(T) &= nF_{avg}S(0)R\left(e^{r(T-t)}\right)_{avg} \\ &\quad -A\,(\text{bid-ask spread})\,n\left(e^{r(T-t)}\right)_{avg} \\ &\quad -A(\text{fixed cost})\left(e^{r(T-t)}\right)_{avg}. \end{aligned} \tag{34}$$

The first line of this equation, (34), corresponds to the case of no transactions, therefore, since we want the full equation to take the form

$$C(T) = \mathcal{F}_{Trans}nF_{avg}S(0)R\left(e^{r(T-t)}\right)_{avg},$$

we must define

$$\mathcal{F}_{Trans} = 1 - \frac{A}{F_{avg}S(0)R}(\text{bid-ask spread}) - \frac{A}{nF_{avg}S(0)R}(\text{fixed cost}).$$

Consider the case of trading twice monthly, so $A = 12.5$. In March 2010, there were no commission costs for trading the three iShares ETFs discussed above on the Fidelity platform, so the fixed cost for these ETFs in our $\mathcal{F}_{Trans}$ formula is zero. Also, these ETFs had an average bid-ask spread of approximately 1.5 cents and a stock price of (at least) \$60. We can therefore estimate that $\frac{(\text{bid-ask spread})}{S(0)} = \frac{(\text{bid-ask spread})}{S(t)} \cdot \frac{S(t)}{S(0)} = \frac{0.015}{60} \cdot (1 - F_{avg}/2)$. Using that $F_{avg} = 0.276$ and $R = .3$, we get $\mathcal{F}_{Trans} = 0.967$. If we are not on the Fidelity platform, we have more standard commission costs, like \$8 per trade, which yields a fixed cost of \$16 for both selling and buying during a transaction. For an initial stock worth of \$50,000 (near the median value), the value of $\mathcal{F}_{Trans}$ would be reduced to 0.919.

5.1.3 Annual \$3000 Limit on Losses

We now consider the effect of the requirement in American tax law that net losses over \$3000 must be deferred to future years. Given an initial stock worth, $nS(0)$, we can determine the average over 100,000 runs of the harvested losses at time T with the \$3000 limit and, over the same runs, the harvested losses at time T without the limit. We set $\mathcal{F}_{3000}$ equal to the ratio of these two losses. For our base case, we find that 100,000 runs gives 3 digits of accuracy for computing $\mathcal{F}_{3000}$, which are given in the following table:

Initial Stock Worth, $nS(0)$	$\mathcal{F}_{3000}$
\$10,000	0.999
\$25,000	0.986
\$50,000	0.946
\$75,000	0.901
\$100,000	0.855
\$250,000	0.619
\$500,000	0.391

It is unsurprising that the \$3000 annual limit has almost no effect on smaller stock portfolios. Even at an initial worth of \$50,000, near the median in America, there is only a 5.4% loss due to this factor.

As the worth of the stock in the portfolio increases past \$100,000, the \$3000 limit is overrun more often, which decreases $\mathcal{F}_{3000}$. As $\mathcal{F}_{3000}$ decreases significantly, inserting the R_{Eff} that the $\mathcal{F}_{3000}$ factor generates into our PDE model becomes more questionable. Therefore, for portfolios with stock positions larger than \$100,000, it is better to alter our PDE model via Marekwica's method [14], as described in Subsection 1.3. Since Marekwica's method requires the addition of a state variable for losses, the computation of the solution is slowed, but this is necessary to maintain accuracy with these large stock positions.

5.1.4 The Combined Effect, R_{Eff}, and Capital Losses Versus Capital Gains Tax Rates

For a taxable portfolio starting with \$50,000 in stock (the approximate median value) and a trader using the three iShares ETFs discussed above on a Fidelity platform, we have

$$R_{Eff} = \mathcal{F}_{Trade}\mathcal{F}_{Trans}\mathcal{F}_{3000}R = 0.931 \times 0.967 \times 0.946 \times 0.3 = 0.255,$$

which is a 15.0% combined loss from the three $\mathcal{F}$ factors.

We note, however, that the effect of the tax rate being different for gains and losses can easily be as big a factor as all of these $\mathcal{F}$ factors combined. In America, capital losses are treated from a tax viewpoint as income loss, which has a higher tax rate than the long term capital gains rate. As of 2010, an American in the 28% federal tax bracket living in the most populous state, California, had a total income tax rate of 28% federal + 9.55% state[4] = 37.55%, compared to a long term capital gains tax rate of 15% federal + 9.55% state = 24.55%. Our base case, $R = 0.3$, is the approximate average of these two values. Using the more accurate income tax rate, $R = 0.3755$, $\mathcal{F}_{Trans}$ changes slightly to 0.974, while $\mathcal{F}_{Trade}$ and $\mathcal{F}_{3000}$ are unchanged. This gives $R_{Eff} = 0.322$, which is higher than the base case $R = 0.3$. For the highest federal income tax bracket, 35%, this increases to $R_{Eff} = 0.383$. Assuming the investor is alive when the portfolio is liquidated, the smaller capital gains rate can be accommodated by replacing $R^{liq} = 0.3$ with $R^{liq} = 0.2455$, which further increases the value of reaping losses. Similarly, the capital gains rate for the initial rebalance, R in (1), also becomes 0.2455.

5.2 Effect of a Single Rebalance

For solving this paper's primary optimization problem, where there is no rebalancing, the PDE based method of Chapter 4 is far superior, both in speed and precision, to Monte Carlo methods. We demonstrate and explain this in 5.2.1.

However, neither the PDE based method nor Monte Carlo methods are tractable for this optimization problem if we extend our model to allow even a single optimal rebalance at $T/2$, the halfway-to-liquidation time (or any other time). Therefore, we consider instead the more common method of rebalancing where we readjust the stock fraction at $t = T/2$ to equal the stock fraction chosen at $t = 0$. Because our examples use power law utilities, this also reflects Merton's strategy of maintaining a constant stock fraction.

Even with this simplified rebalancing strategy, the PDE based method is essentially intractable. At rebalancing, this method requires adding one dimension to the PDE for the amount of stock bought or sold in the rebalance and another dimension for the lowest stock price attained after $T/2$. The resulting 4 space + 1 time dimensional PDE now becomes far less desirable than Monte Carlo methods, which remain almost as fast as they were without rebalancing. Therefore, in 5.2.2 we will apply Monte Carlo methods to our simplified rebalancing strategy. We will see that, when including taxes,

[4]This is the California tax rate for single filers with incomes between $47,055 and $1,000,000.

this rebalancing can have a beneficial or detrimental effect on the optimal expected utility.

5.2.1 Monte Carlo Optimization without Rebalancing

We return to the base case with no rebalancing from Section 4. As in Section 4, we are looking to find f^{opt}, the initial stock fraction that optimizes the normalized certainty equivalent, $\mathcal{C}(f)$, given in (31). Since $\mathcal{C}'(f^{opt}) = 0$, the values of $\mathcal{C}$ for f near f^{opt} are very close to each other. This means that attempting to determine f^{opt} requires determining many significant digits for the value of $\mathcal{C}$. Monte Carlo methods converge to the value of $\mathcal{C}$ at a rate of order $\frac{1}{\sqrt{m}}$, where m is the number of runs in the simulation. Therefore, computational time quickly increases as we try to obtain more significant digits for $\mathcal{C}$. Also, as we compare computed $\mathcal{C}$ outputs for progressively close f values, the noise in the Monte Carlo results becomes a progressively dominant effect, so it is impossible to use efficient optimization algorithms like the successive parabolic interpolation used in the PDE based algorithm of Section 4.

These effects take a computational toll. Each Monte Carlo evaluation of $\mathcal{C}$ at a given f was computed from the mean of 2500 simulations where each simulation used 20,000 runs with 100 trading intervals per year. This total of $2500 \times 20,000 = 50$ million runs required, on average, 7730 seconds using MATLAB R2009b. The standard deviation for $\mathcal{C}$ was computed by taking the standard deviation of these 2500 simulations and dividing by $\sqrt{2500}$. The following 7 evaluations were performed[5]:

f	mean for $\mathcal{C}$	std. dev. for $\mathcal{C}$
0.80	1.364139	0.000176
0.81	1.364740	0.000182
0.82	1.364918	0.000180
0.83	1.365107	0.000182
0.84	1.364913	0.000186
0.85	1.364966	0.000191
0.86	1.364269	0.000184

[5]Note that our results are sensitive to the total number of runs (50 million here), but not the choice for dividing the runs into the number of simulations (2500 here) versus the number of runs per simulation (20,000 here), as long as both of these last two numbers remain reasonably large.

Even at this level of computational time, the standard deviation is large compared to the difference in $\mathcal{C}$ outputs. Of course, reducing the standard deviation further would require a significant multiple of the more than 15 hours required to obtain the above data. From pairwise tests (at the 95% level of significance) for the hypothesis that the means of $\mathcal{C}$ are unequal, we are able to conclude from our data only that f^{opt} is somewhere between 0.80 and 0.86.

For comparison, recall that the PDE based algorithm in Section 4 converged quickly. Even for the most computationally expensive run, $\Delta x = 0.04$, each evaluation of $\mathcal{C}$ took only 88 seconds and, for the base case, we converged after only four evaluations to $f^{opt} = 0.8324$.

5.2.2 Monte Carlo Optimization with a Single Rebalance

We now consider the base case again, but allow, as stated before, for a single rebalancing at $t = T/2$ to the same fraction of stock chosen at $t = 0$. The Monte Carlo computational time slightly increases to, on average, 8090 seconds per evaluation of $\mathcal{C}$. Our results are similar to the case without rebalancing:

f	mean for $\mathcal{C}$	std. dev. for $\mathcal{C}$
0.79	1.365365	0.000170
0.80	1.365867	0.000176
0.81	1.366028	0.000175
0.82	1.366279	0.000176
0.83	1.366023	0.000177
0.84	1.365989	0.000187
0.85	1.365773	0.000184

We see that optimal rebalancing generates a very small increase in the expected utility for this case, and we can only conclude that f^{opt} is somewhere between 0.79 and 0.85. Therefore, the effect of the rebalancing is small, but slightly beneficial.

The fact that capital gains taxes must be paid when stock is sold during rebalancing has a detrimental effect on the expected utility. This detrimental effect can be stronger than the beneficial effect from recapturing the optimal stock fraction in rebalancing.

For example, we alter the base case so that $R = R^{liq} = 0.50$, which increases the capital gains paid during rebalancing. From the table in Subsection 4.2 where R varied, we know that if there is no rebalancing, then $f^{opt} = 0.64$, and we can compute by either the PDE or Monte Carlo method that $\mathcal{C}(0.64) = 1.173$. But if we allow for a rebalancing then, by Monte Carlo, we have

f	mean for $\mathcal{C}$	std. dev. for $\mathcal{C}$
0.64	1.160353	0.000107
0.67	1.161020	0.000111
0.68	1.161302	0.000113
0.69	1.161414	0.000115
0.70	1.161352	0.000115
0.71	1.161095	0.000119
0.72	1.160878	0.000119

and we conclude that f^{opt} is somewhere between 0.67 and 0.72. That is, when we include rebalancing in this case, we see a statistically significant increase in f^{opt} but a statistically significant *decrease* in the optimal value for $\mathcal{C}$ due to the early payment of capital gains. In contrast, because the rebalancing strategy outlined earlier for the PDE based algorithm defers rebalancing decisions until they must be made, it automatically avoids any rebalance that would diminish the expected utility at liquidation.

6 Concluding Remarks

In this paper, we showed how to quantify the effects of including the continuous realization of capital loss arbitrage into an otherwise passive investment strategy. This allowed us to determine the optimal initial rebalance with this strategy and also allowed us to determine optimal subsequent rebalances. We demonstrated that incorporating capital loss arbitrage leads to an oblique reflecting boundary condition for the PDE describing the evolution of the probability density function. Extensions of the model were presented for multiple stocks in the portfolio, probability distributions for the portfolio liquidation time, and money from capital loss arbitrage being put back into stock instead of cash.

For the model without extensions, we detailed a stable, first order accurate numerical algorithm that computes the solution to our derived PDE and boundary conditions. We established that this PDE based algorithm yields significantly more precise results than Monte Carlo methods in a fraction of the computational time. While our PDE based algorithm can be used with any utility function at the time of liquidation, our computed examples used power law utility functions. This allowed for comparisons with the optimal constant fraction to be kept in stock at all times, as determined by Merton when there are no taxes. In our examples the optimal initial fraction for our portfolio is not drastically different from the Merton fraction despite the fundamental differences between the two cases (e.g., taxes vs. no taxes, single initial rebalance vs. continuous rebalancing).

For a base case with a 40 year time horizon and a reasonable set of parameters, including a capital gains and losses tax rate of 30%, we showed that harvesting capital losses moves the optimal initial asset allocation from 70% in stock to 83% in stock. For the same time horizon and parameters, we showed that harvesting capital losses is equivalent to a portfolio starting with, on average, 4.3% more money, where this extra money is put in cash. If capital gains are forgiven when the portfolio is liquidated due to an investor's death, we showed this figure increases to 5.5%. We demonstrated how these numbers change as we alter each of the model parameters: the stock's expected return μ, the stock's volatility σ, the tax rate R, the risk-free tax-free interest rate r, the power law utility function exponent $1-\rho$, and the liquidation time T.

Finally, we approximated the effect of removing many of the major assumptions used in our PDE model. We used Monte Carlo methods to examine the effect of not trading continuously to avoid wash sales, transaction costs, and the \$3000 annual limit on claimed losses. We applied the same 40 year time horizon and base case parameter values previously used to a \$50,000 taxable stock account, which is near the median value of such accounts in America. For this account, the approximate combined effect of removing these assumptions reduced the average rate at which money is collected from losses by 15%. On the other hand, we saw that more realistic values for the capital loss tax rate, which in America is equal to the income tax rate, often more than compensate for this reduction. Finally, we considered the effect of rebalancing the portfolio to its initial stock fraction at the halfway time to liquidation, $T/2$. For our examples, we showed this rebalancing can have small beneficial or detrimental effects depending on which of two factors are stronger: the desirable restoration of the optimal stock fraction or the undesirable early payment of capital gains from the rebalance.

There are a number of future questions and directions this paper suggests. For example:

- Deciding how to optimally allow for continuous rebalancing in the presence of taxes while still using the full purchase history remains, largely, an open question.

- Geometric Brownian motion underestimates the minimum stock price over time, causing our model to underestimate the beneficial effects of harvesting losses. Our framework can be extended to using more complex, accurate models for stock dynamics. However, determining the adjoint problem for the correspondingly more complex PDE can pose a difficult challenge.

- Determining how to optimally realize both gains and losses when the tax rate for gains is smaller than the rate for losses is still, by and large, an open question. The optimal strategy is heavily dependent upon the fact that gains and losses cancel until the end of each calendar year, after which gains are fully paid but losses over the annual limit are deferred. This can cause important variation within the course of each year in the optimal trading strategy.

- Our estimate of the total transaction costs is too high because we chose a conservative model that trades whenever there is a loss of any size. In reality, small losses would be ignored, since transaction costs would outweigh the benefit of reaping these losses. It would be helpful to determine the optimal size that a loss must attain before it should be reaped.

7 Bibliography

References

[1] Ben Tahar, I., Soner, H. M., Touzi, N. Merton problem with taxes: characterization, computation, and approximation. *SIAM J. Financial Math.* **1** (2010), 366–395

[2] Ben Tahar, I., Soner, H. M., Touzi, N. The dynamic programming equation for the problem of optimal investment under capital gains taxes. *SIAM J. Control Optim.* **46** (2007), no. 5, 1779–1801

[3] Chang, M.-H. Hereditary portfolio optimization with taxes and fixed plus proportional transaction costs. I. *J. Appl. Math. Stoch. Anal.* (2007), Art. ID 82753, 33 pp.

[4] Chang, M.-H. Hereditary portfolio optimization with taxes and fixed plus proportional transaction costs. II. *J. Appl. Math. Stoch. Anal.* (2007), Art. ID 40149, 25 pp.

[5] Constantinides, G. Capital market equilibrium with personal taxes. *Econometrica* **51** (1983), 611–636.

[6] Dammon, R., Spatt, C., and Zhang, H. Optimal consumption and investment with capital gains taxes. *Review of Financial Studies* **14** (2001) 583–616.

[7] Dammon, R., Spatt, C., and Zhang, H. Optimal asset location and allocation with taxable and tax-deferred investing. *Journal of Finance* **59** (2004) 999–1037.

[8] DeMiguel, A. and Uppal, R. Portfolio investment with the exact tax basis via nonlinear programming. *Management Science* **51** (2005) 277–290.

[9] Dumas, B., Super Contact and Related Optimality Conditions, *J. of Economic Dynamics and Control*, **15**, (1991) 675–685.

[10] Dybvig, P. and Koo, H. Investment with taxes. Working paper (1996) Washington University.

[11] Gallmeyer, M., Kaniel, R., and Tompaidis, S. Tax management strategies with multiple risky assets. *Journal of Financial Economics* **80** (2006) 243–291.

[12] Krugman, P. Target Zones and Exchange Rate Dynamics. *Quarterly Journal of Economics*, **106**, no. 3 (1991) 669-682.

[13] Gill, P., Murray, W., and Saunders, M. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM J. Optim.* **12** (2002) 979–1006.

[14] Marekwica, M. Optimal Tax-Timing and Asset Allocation when Tax Rebates on Capital Losses are Limited. Working Paper (January 10, 2010). Available at SSRN: `http://ssrn.com/abstract=997813`

[15] Merton, R. Continuous-Time Finance. (1992) revised edition, Blackwell, Oxford.

[16] Survey of Consumer Finances, Federal Reserve Bulletin (2009), Federal Reserve Board, available at `www.federalreserve.gov/pubs/bulletin/2009/pdf/scf09.pdf`

[17] Trefethen, L. and Williams, R. Conformal mapping solution of Laplace's equation on a polygon with oblique derivative boundary conditions. Special issue on numerical conformal mapping. *J. Comput. Appl. Math.* **14** (1986) 227–249.