RFS Advance Access published May 3, 2015

Optimal Tax Timing with Asymmetric Long-Term/Short-Term Capital Gains Tax

Min Dai
Department of Mathematics and Risk Management Institute, NUS

Hong Liu
Olin Business School, Washington University in St. Louis and CAFR

Chen Yang
Department of Mathematics, NUS

Yifei Zhong
Mathematical Institute, Oxford University

We develop an optimal tax-timing model that takes into account asymmetric long-term and short-term tax rates for positive capital gains and limited tax deductibility of capital losses. In contrast to the existing literature, this model can help explain why many investors not only defer short-term capital losses to long term but also defer large long-term capital gains and losses. Because the benefit of tax deductibility of capital losses increases with the short-term tax rates, effective tax rates can decrease as short-term capital gains tax rates increase. (*JEL* G11, H24, K34, D91)

Investors in U.S. stock markets are subject to capital gains tax when gains or losses are realized. When gains are realized, a lower long-term tax rate applies if and only if the stock holding period is at least one year. In contrast, when losses are realized, a higher short-term rate applies regardless of the length of the holding period, and investors effectively get a tax rebate. The short-term rate is set to investors' marginal ordinary income tax rate, and the rebate is implemented through deducting the losses from their taxable ordinary income. Assuming that a long-term tax rate applies to long-term losses, the existing literature on optimal investment with capital gains tax argues that

We thank Pietro Veronesi (the Editor) and an anonymous referee for very helpful comments. We are grateful to Bob Dammon, Phil Dybvig, Marcel Fischer, Jennifer Huang (AFA Discussant), Pete Kyle, Mark Loewenstein, Chester Spatt, Raman Uppal, Harold Zhang, and seminar participants at 2011 CICF, 2013 AFA, University of Maryland, and Washington University in St. Louis for useful discussions. We thank Yingshan Chen and Yaoting Lei for excellent research assistance. Min Dai acknowledges support from Singapore AcRF grant R-146-000-188/201-112 and NUS Global Asia Institute -LCF Fund R-146-000-160-646. Chen Yang acknowledges support from Wee Cho Yaw Finance and Banking Scholarship Award. Send correspondence to Hong Liu, Olin Business School, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130; telephone: (314) 935-5883. E-mail: liuh@wustl.edu.

doi:10.1093/rfs/hhv024

investors should realize all losses before they turn long term and realize all gains right after they turn long term.[1] In contrast, empirical evidence shows that many investors defer not only short-term losses beyond one year but also large long-term gains and losses.[2]

In this paper, we propose an optimal tax-timing model that can help explain this puzzle. In contrast to the large amount of existing literature,[3] our model takes into account three important features of the current tax code: (i) the tax rates for long-term gains can be lower than the rates for short-term gains; (ii) capital losses allowed to offset taxable ordinary income are capped at $3,000 per year, with the rest carried forward indefinitely for offsetting future gains and/or income; and (iii) short-term tax rates apply to both long-term and short-term losses.

More specifically, we consider an optimal capital gains tax-timing problem of a small (i.e., no price impact), constant relative risk averse investor who can continuously trade a risk-free asset and a stock to maximize expected utility from intertemporal consumption and bequest. According to the current tax code, if an investor bought shares of a stock at different times, the capital gain for a particular share sold is equal to the difference between the sale price and the original purchase price of this share, and the applicable tax rate is determined by whether it is a loss and whether the holding period of this share is at least one year. Therefore, one needs to keep track of the exact original purchase price ("exact basis") and the exact holding period of each share, which causes path dependency of the optimal investment policy. Because this path dependency makes the optimal investment problem infinitely dimensional, we approximate the exact basis using the average basis of all the shares held, as in most of the existing literature, and we approximate the exact holding period using the basis-weighted average of the holding periods of these shares ("average holding period").[4]

---

[1] This way, they receive a tax rebate at the higher short-term rate for losses and enjoy the lower long-term rate for the realized gains and reestablish the short-term status for potential subsequent losses (see, e.g., Constantinides 1983, 1984; Dammon and Spatt 1996; and Dammon, Spatt, and Zhang 2001).

[2] For example, Wilson and Liddell (2010) report that among all 2007 U.S. tax returns, there were 53,403 long-term gain transactions, of which 63.8% (30.0% of all gain transactions) had a holding period of 18 months or longer with an average gain of $7,964.33 per transaction, and there were 19,186 long-term loss transactions, of which 62.9% (20.0% of all loss transactions) had a holding period of 18 months or longer with an average loss of $2,741.33 per transaction.

[3] For example, Cadenillas and Pliska (1999), Dammon, Spatt, and Zhang (2004), Gallmeyer, Kaniel, and Tompaidis (2006), Ben Tahar, Soner, and Touzi (2010), Ehling et al. (2010), Marekwica (2012), and Fischer and Gallmeyer (2012).

[4] As shown in DeMiguel and Uppal (2005), an investor rarely has more than one cost basis, and as shown in the Appendix, the certainty equivalent wealth loss from following a single-basis strategy (which is a feasible, but suboptimal strategy in our model) is almost negligible (to keep a single basis, one needs to liquidate the entire position before any additional purchases can be made). A previous version of this paper conducts the same analysis as this paper with restriction to the class of single-basis strategies and finds qualitatively the same results as this paper.

For positive capital gains, a long-term tax rate applies if and only if the average holding period is at least one year. Different from the existing literature but consistent with the tax code, a higher short-term rate applies to both long-term and short-term capital losses. We consider both the full rebate (FR) case where an investor can use all capital losses to offset taxable ordinary income and the full carry-forward (FC) case where the investor can only carry forward capital losses to offset future gains and/or income.[5] The FR case applies better to lower-income investors whose capital losses are likely less than $3,000 per year, while the FC case is more suitable for high-income investors for whom capital losses can be much more than $3,000 per year and a tax rebate (which is capped at $3,000 × 39.6% = $1,188 per year) is relatively unimportant. The optimal tax-timing and trading strategy is characterized by a no-trade region, a buy region, and a sell region that vary through time. In both the buy region and the sell region, the investor trades to the no-trade region to achieve optimal risk exposure and optimal tax timing. Outside the sell region, capital gains tax is deferred.

In contrast to the existing literature (e.g., Constantinides 1984; Dammon and Spatt 1996; Ehling et al. 2010; Marekwica 2012), we find that it may be optimal for investors to defer not only short-term losses but also large long-term gains and long-term losses. Intuitively, different from what is assumed in the existing literature, the higher short-term rates apply to both long-term and short-term losses under the current law. Thus, the long-term status strictly dominates the short-term status. Therefore, it may be optimal for investors to defer some possibly large gains and losses regardless of the length of the holding period. In addition, high-income investors may optimally defer larger long-term gains and losses than would lower-income investors.[6] The main intuition is that there is an additional benefit of deferring the realization of gains for high-income investors: it makes incremental losses effectively tax rebatable without limit.[7] When there is a large long-term loss, and the long-term rate is much lower than the short-term rate, keeping the long-term status by deferring realization can provide significant benefit from the much lower long-term rate when stock prices rise and current losses turn into gains. In addition, the benefit of realizing long-term gains or losses to reestablish the short-term status for future losses is small for high-income investors because only a small fraction of losses can be tax deductible at the higher short-term rate for these investors.

---

[5] Our main qualitative results such as an investor may defer long-term gains, and long-term losses are valid in both of these polar cases. Therefore, a model with positive but limited rebate for losses would unlikely produce qualitatively different results.

[6] The few existing studies that consider limited tax deductibility of losses (e.g., Ehling et al. 2010, Marekwica 2012) consider only symmetric tax rates and are thus silent on the optimal deferring strategy of long-term gains and losses when short-term rates are higher.

[7] To help understand this additional benefit, consider a simple example where a high-income investor realizes a gain of $1, pays the capital gain tax, reestablishes a stock position, and immediately loses $1. If the investor did not realize the gain, then the subsequent loss would offset the original gain, and the investor would not need to pay any tax if the stock were sold after the stock price decreased.

We also show that adopting the optimal trading strategy can be economically important. For example, consider the alternative strategy of immediately realizing all losses and long-term gains but deferring all short-term gains, as most of the existing literature recommends. We find that the certainty equivalent wealth loss (CEWL) from following this alternative strategy is about 0.84% of the initial wealth for lower-income investors and as much as 5.20% for high-income investors, given reasonable parameter values.

Because lower-income investors can effectively obtain a tax rebate at the short-term tax rate for a high percentage of their capital losses and can defer short-term capital gains to long term, a higher short-term tax rate would enable them to get a higher rebate in case of a loss without paying much more in case of a gain, and thus could make them better off. This implies that effective tax rates on equity securities for lower-income investors can actually decline in short-term capital gains tax rates. Even when other capital gains tax rates (e.g., long-term tax rates) are also increased, lower-income investors may still become better off because marginal utility of wealth is higher when there is a capital loss, and thus the tax rebate effect of a higher short-term rate can dominate. Indeed, we find that keeping everything else (including the ordinary income tax rate) constant, lower-income investors can be significantly better off with higher capital gains tax rates. In addition, with higher capital gains tax rates, lower-income investors generally invest more and consume more, because the after-tax stock return becomes less risky. As Wilson and Liddell (2010) reported, in 2007 tax returns with an adjusted gross income of $100,000 or less had short-term net losses on average. These tax returns, about six million in total, accounted for more than half of all the returns that had short-term gains or losses, which suggests that many lower-income investors could indeed benefit from higher short-term capital gains tax rates.[8]

Additionally, we analyze the sources of the value of tax deferral. The value of tax deferral comes from (i) saving the time value of capital gains tax; (ii) realizing gains at the lower long-term rate in the future; and (iii) in the FC case, making a capital loss effectively rebatable while deferring. We show that for lower-income investors this value mainly comes from realizing gains at a lower long-term rate because the interest rate is usually much lower than the difference between the short-term rate and the long-term rate. In contrast, for high-income investors the value comes mainly from making a capital loss effectively rebatable at the short-term rate because the short-term rate is much higher than both the long-term rate and the interest rate for these investors.

[8] In contrast, high-income investors for whom the FC case fits better are worse off with higher capital gains tax rates, and their stock investment and consumption are almost insensitive to changes in short-term tax rates because they defer most of the short-term gains.

1. The Model

There are two assets that an investor who is subject to capital gains tax can trade without any transaction costs. The first asset is a money market account growing at a continuously compounded, constant rate of r. The second asset ("the stock") is a risky investment.[9] The stock pays a constant, continuous dividend yield of δ. The ex-dividend stock price S_t follows the process

$$\frac{dS_t}{S_t} = \mu dt + \sigma dw_t, \tag{1}$$

where μ and σ are constants with $\mu + \delta > r$, and w_t is a one-dimensional Brownian motion.

The tax rates for long-term investment may be lower than those for short-term investment. According to the tax code, capital gains tax depends on the final sale price, the exact initial purchase price ("exact basis"), and the exact holding period of each share sold. Therefore, the optimal investment strategy becomes path dependent (e.g., Dybvig and Koo 1996), and the optimization problem is of infinite dimension.[10] To simplify analysis, we approximate the exact cost basis using the average cost basis of a position, as in most of the existing literature (e.g., Dammon, Spatt, and Zhang 2001, 2004; Gallmeyer, Kaniel, and Tompaidis 2006), and in addition, we approximate the exact holding period using the basis-weighted average of the holding periods of the shares in the position ("average holding period"). Using an exact-basis and exact-holding-period model similar to that of DeMiguel and Uppal (2005) but with asymmetric tax rates and multiple trading opportunities within a year, we find that the optimal strategy using the average-basis and average-holding-period rule yields almost the same utility as using the exact-basis and exact-holding-period rule (see Appendix A.5).

As the findings in DeMiguel and Uppal (2005) and our analysis in Appendix A.5 suggest, an investor rarely holds more than one cost basis. Accordingly, an alternative approximation approach is to restrict feasible trading strategies to those that result in a single cost basis—that is, an investor must liquidate the entire stock position before making any additional purchase so that at any point

[9] The risky asset can be interpreted as an exchange traded fund (ETF) that represents a diversified portfolio. Although ETFs are also pass-through entities like open-end mutual funds, ETFs pass through smaller amounts of capital gains because they are typically passively managed and because they are more likely to use in-kind redemptions that reduce the required distributions (e.g., Poterba and Shoven 2002). As a result, most of the capital gains tax for an ETF investor is realized at sale, like a stock. An extension to a multi-stock case might help understand cross-stock tax management strategy, but it would unlikely change our main qualitative results.

[10] As an example of the exact-basis and exact-holding-time system, suppose an investor bought 10 shares at \$50/share one and half years ago and purchased 20 more shares at \$60/share three months ago. The first 10 shares have a cost basis of \$50/share and a holding period of 1.5 years, and the remaining 20 shares have a cost basis of \$60/share and a holding period of 0.25 years. If the investor sells the entire position at \$65/share, the early purchased 10 shares have a capital gain of $65 \times 10 - 50 \times 10 = \$150$ and will be taxed at the long-term tax rate, and the remaining 20 shares have a capital gain of $65 \times 20 - 60 \times 20 = \$100$ and will be taxed at the short-term tax rate.

in time all shares in the entire stock position were purchased at the same time and at the same initial cost. Clearly, a single-basis strategy is a feasible strategy in our average-basis and average-holding-period model. One advantage of the single-basis model over the average-basis and average-holding-period model is that no approximation is needed for the holding period. However, restriction to a single-basis strategy biases against purchases because any purchases require that the investor first realize all capital gains or losses.[11]

We consider both the full rebate (FR) case where an investor can use all capital losses to offset taxable ordinary income and the full carry-forward (FC) case where the investor can only carry forward capital losses to offset future gains and/or income.[12]

Let L be the shortest holding period required to qualify for a long-term tax status and h_t be the basis-weighted average holding period at time t. To reduce the (unintended) incentive to hold some shares for a long time just to make the average holding period of a position greater than the long-term threshold L, we cap the average holding period h_t at $\bar{h} > L$.[13] Let $\tau(h)$ be the tax rate function defined as follows:

$$\tau(h) = \begin{cases} \tau_S & if \quad h < L \\ \tau_L & if \quad h \geq L, \end{cases} \tag{2}$$

where $\tau_S \geq \tau_L \geq 0$ are constants.

As in most models on optimal investment with capital gains tax, we further assume (i) the tax on dividend and interest is due when they are paid;[14] (ii) capital gains tax is realized immediately after sale; (iii) there is no wash sale restriction; and (iv) shorting against the box is prohibited.

[11] In addition, using a single-basis model does not change our qualitative results, as we have shown in a previous version of the paper.

[12] For a high-income investor with over \$1.2 million and an investment horizon of ten years, the certainty equivalent wealth gain from a tax rebate is less than 1% of the initial wealth because the maximum tax rebate the investor can get is capped at $\$3,000 \times 39.6\% = \$1,188$ a year. This suggests that for millionaire investors, the FC case likely applies well. From our simulation results using the optimal trading strategy under the FR model given the default parameter values in the numerical analysis section, for an investor with less than \$34,642 initial wealth and an investment horizon of 10 years, the probability that the investor has no loss above \$3,000 any time in the 10-year horizon is greater than 0.99. For these lower-wealth investors, the FR case likely applies well. We also solved a similar problem where an investor switches from the FR case in the first half of the investment horizon to the FC case in the second half and vice versa. These switches do not affect any of our main qualitative results, and the initial trading strategies stay virtually the same.

[13] Setting the upper bound $\bar{h} = L$ would imply if one buys one additional share, then the short-term rate applies to the entire position regardless of the size of the existing position. Choosing $\bar{h} > L$ implies that when there are both long-term and short-term positions, one can realize capital gains sometimes at the long-term rate and sometimes at the short-term rate, depending on whether $h \geq L$. This better approximates current tax code: when there are both long-term and short-term shares in a position, one can realize those long-term gains at the long-term rate and those short-term gains at the short-term rate. We find that our main results are robust to the choice of this upper bound that varies from 1 to 10, which suggests that the incentive to hold some shares for a long time just to make the average holding period $h \geq L$ seems small in our model.

[14] Interest paid on margin loan for stock purchasing is tax deductible.

The investor is endowed with x_0 dollars cash and y_0 dollars worth of stock at time 0.[15] Let x_t denote the dollar amount invested in the riskless asset, y_t denote the dollar value of the stock holding, B_t be the total cost basis for the stock holding, and $H_t \equiv B_t h_t$ be the basis-weighted total holding time, all at time t.[16] We first state the evolution equations for the state variables and then provide explanations below.

$$dx_t = (1-\tau_i) r x_t dt + (1-\tau_d)\delta y_t dt - c_t dt + f\left(0, y_t, B_t, \frac{H_t}{B_t}; 1\right) dM_t - dI_t, \quad (3)$$

$$dy_t = \mu y_t dt + \sigma y_t dw_t - y_t dM_t + dI_t, \quad (4)$$

$$dB_t = -B_t dM_t + \omega (B_t - y_t)^+ dM_t + dI_t, \quad (5)$$

$$dH_t = B_t \mathbf{1}_{\{H_t < \bar{h} B_t\}} dt - H_t dM_t, \quad (6)$$

where

$$f(x, y, B, h; \iota)$$
$$\equiv x + \left[y - \iota\left(\tau(h)(y-B) - (1-\omega)\kappa(\tau_S - \tau(h))(B-y)^+ + \omega\tau(h)(B-y)^+\right)\right] \quad (7)$$

is the after-tax wealth (zero tax if $\iota = 0$, as in the case of tax forgiving at death); dM_t represents the fraction of the current stock position that is sold; dI_t denotes the dollar amount purchased; $\mathbf{1}_{\{H_t < \bar{h} B_t\}}$ is an indicator function that is equal to one if the average holding period $h_t = H_t / B_t$ is below $\bar{h}$ and zero otherwise;[17] τ_i and τ_d are the tax rates for interest and dividend, respectively; $\omega = 0$ or 1 corresponds respectively to the FR case or the FC case, and $\kappa = 0$ or 1 corresponds respectively to applying the long-term tax rate to long-term losses and the short-term tax rate to short-term losses or applying the short-term tax rate to both long-term and short-term losses. Note that when a jump in a variable occurs at t (e.g., y_t, B_t), the variable value on the right-hand side of the equations represents the value just before the jump—that is, the time $t-$ value.

On the right-hand side of Equation (3), the first two terms are, respectively, the after-tax interest earned and dividend paid; the third term is the consumption flow; the fourth term denotes the after-tax dollar revenue from selling a fraction

15 This initial endowment includes the present value of all future after-tax ordinary income to which the investor can add the tax rebate from capital losses, if any.

16 Take the example in Footnote 10. In the average-basis and average-holding-period scheme, the total cost basis $B_t = \$50 \times 10 + \$60 \times 20 = \$1,700$, the average basis is $B_t/30 = \$56\frac{2}{3}$/share, the basis-weighted *total* holding time for the entire position is $H_t = 50 \times 10 \times 1.5 + 60 \times 20 \times 0.25 = 1,050$ (dollar year), the average holding period $h_t = H_t / B_t = 0.62$ years, and thus the total capital gain of $\$65 \times 30 - \$56\frac{2}{3} \times 30 = \$250$ is taxed at the short-term rate.

17 As a convention, if $H_t = 0$ and $B_t = 0$, then $h_t = H_t / B_t = 0$.

dM_t of the time t stock position, and the last term, dI_t, is the dollar cost of purchasing additional stock at time t. By Equation (7), $f(0, y, B, H/B; 1)$ in Equation (3) represents the after-tax dollar revenue from selling the entire stock position. To help understand this, we consider four cases. First, if there is a capital gain (i.e., $y > B$), then the bracketed term reduces to $y - \tau(h)(y - B)$, which is clearly the after-tax revenue from selling the entire position. Second, if there is a capital loss, and the investor can only carry forward the loss (i.e., $y < B$ and $\omega = 1$), then the revenue term becomes y, meaning the investor does not get any tax rebate for the loss. Third, if there is a capital loss, the investor gets a full rebate, and the short-term tax rate is used for both long-term and short-term losses (i.e., $y < B$, $\omega = 0$, and $\kappa = 1$), then the revenue term becomes $y - \tau_S(y - B)$, which implies that the investor gets a full tax rebate at the short-term rate. Fourth, if there is a capital loss, the investor gets a full rebate, and the short-term tax rate is used for short-term losses and the long-term rate is used for long-term losses (i.e., $y < B$, $\omega = 0$, and $\kappa = 0$), then the revenue term becomes $y - \tau(h)(y - B)$, which implies that the investor gets a full tax rebate at the long-term rate for long-term losses and at the short-term rate for short-term losses.[18]

Equation (4) states that the value of the stock position fluctuates with the stock price, decreases by the amount of sales $y_t dM_t$, and increases by the amount of purchases dI_t. Between trades, the dollar value of the stock holding follows a log normal distribution.

Equation (5) shows that the total cost basis B_t increases with purchases and decreases with sales. When a capital loss is fully rebatable ($\omega = 0$), the cost basis decreases proportionally with sales. For example, a sale of 50% of the current position (i.e., $dM_t = 0.5$) reduces the cost basis by 50%. When there is a loss (i.e., $B_t > y_t$) and a capital loss can only be carried forward ($\omega = 1$), the loss $(B_t - y_t)dM_t$ is added back to the remaining basis to be carried forward to offset future gains.[19]

Equation (6) implies that without a sale, the basis-weighted total holding time H_t is increased by the cost basis B_t multiplied by the time passed, up to a limit. If there is a sale, then the total holding time is reduced proportionally. On the other hand, H_t is not immediately affected by a purchase at time t (i.e., dI_t is absent in Equation (6)) because at the time of purchase, the holding

[18] To understand the average-basis approximation in Equation (3), let n be the number of shares sold at time t and N be the total number of shares the investor holds just before the sale. Then $dM_t = \frac{n}{N}$ and the realized capital gain is equal to $n \times (S_t - \bar{B}_t) = n \times (\frac{y_t}{N} - \frac{B_t}{N}) = (y_t - B_t)dM_t$, where $\bar{B}_t$ is the average basis.

[19] Note that the carried loss as modeled here does not differentiate long-term losses from short-term ones. According to the tax code, whether carried loss is long term or short term matters only if there is a future long-term net gain after offset by the carried loss and an investor can get a tax rebate for the carried loss at the time of the gain realization. Because any loss can only be carried forward in the FC model, whether carried loss is long term or short term does not matter for our model. In addition, our simulation results show that both the probability of a future long-term net gain after offset by the carried loss and the dollar value difference it makes when this occurs are small for reasonable parameter values. This suggests that even for a carry-forward model that allows a positive amount of rebate, the status of the carried loss is unlikely important.

period for the newly purchased shares is zero. The indicator function keeps the average holding period $h_t = H_t / B_t$ below $\bar{h}$. This is because if h_t reaches $\bar{h}$, then Equation (6) becomes $dH_t = -H_t dM_t$. Equations (5) and (6) imply that (i) without a purchase or a sale, both the total holding time H_t and the cost basis B_t stay the same, and so does the average holding period h_t; (ii) with a sale, if $\omega = 0$ or $y \geq B$, then both the total holding time and the cost basis go down by the same proportion, and so the average holding period stays at $\bar{h}$; if $\omega = 1$ and $y < B$, then the total holding time goes down by a greater proportion than the cost basis does, and so the average holding period goes down; (iii) with a purchase, the total holding time stays the same, but the cost basis increases, and so the average holding period goes down.

The investor maximizes expected utility from intertemporal consumption and the final after-tax wealth at the first jump time $\mathcal{T}$ of an independent Poisson process with intensity λ. If this jump time represents death time, the capital gains tax may be forgiven (e.g., in the United States) or may not be forgiven (e.g., in Canada) at the death time.[20] Let $V(x_0, y_0, B_0, H_0)$ be the time 0 value function, which is equal to

$$\sup_{\{c_t, M_t, I_t\}} E\left[\alpha \int_0^{\mathcal{T}} e^{-\beta t} u(c_t) dt + (1-\alpha) e^{-\beta \mathcal{T}} u\left(f\left(x_{\mathcal{T}}, y_{\mathcal{T}}, B_{\mathcal{T}}, \frac{H_{\mathcal{T}}}{B_{\mathcal{T}}}; \iota\right)\right)\right], \quad (8)$$

subject to (3)–(6) and the solvency constraint

$$f\left(x_t, y_t, B_t, \frac{H_t}{B_t}; 1\right) \geq 0, \ \forall t \geq 0, \quad (9)$$

where $\beta > 0$ is the subjective discount rate; $\alpha \in [0, 1]$ is the weight on intertemporal consumption; $\iota = 0$ or 1 indicates whether tax is forgiven or not at the jump time; and

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}$$

with the relative risk aversion coefficient γ being positive and not equal to 1.

1.1 Discussions on the assumptions of the model

Clearly, with us adopting several simplifying assumptions to make the analysis tractable, our model is only a broad approximation of reality. On the other hand, these assumptions most likely do not affect our main qualitative results. Take the main simplifying assumption of the average-basis and average-holding-time approximation as an example. Even with exact basis and exact holding period as stipulated by the tax code, an investor may still defer long-term gains

[20] This jump time can also represent the time of a liquidity shock upon which one must liquidate the entire stock position. As shown by Carr (1998) and Liu and Loewenstein (2002), one can use a series of random times to approximate a fixed horizon (e.g., of performance evaluation), and when the investment horizon is long, the approximation using one jump time is usually sufficient.

and losses because long-term status still strictly dominates short-term status. In addition, lower-income investors can get even greater benefit from higher short-term tax rates in the exact-basis and exact-holding-period model because they would be able to pick precisely the shares with the greatest losses to realize first, while the average-basis approximation essentially forces investors to sell the same proportion of the shares for each different cost basis whenever they sell stock.

Another simplifying assumption is that tax rates stay the same as an investor ages. Time-varying tax rates surely would quantitatively affect the optimal deferral strategy. For example, if an investor expects tax rates to decrease in the near future, then that investor has a stronger incentive to defer more capital gains and for a longer time. However, qualitative results in this paper still hold because the main driving forces of these results are still present. The impact of the assumption on the immediate realization of capital gains tax is probably small, especially when the interest rate is low. This is because if the interest rate is zero, then the investor can capitalize the tax rebate/payment to be received/paid later without interest cost. The assumption of no wash sale restriction is also unlikely to change our main results because, as Constantinides (1983) argued, an investor may purchase a different stock with similar risk and return characteristics to effectively bypass the wash sale restriction. Adding transaction cost to the model would likely widen the no-trade region, and thus make investors defer even larger capital gains and losses, thereby strengthening our qualitative results.

2. Theoretical Analysis

In this section we conduct some theoretical analysis that facilitates our subsequent analysis.

The associated Hamilton-Jacobi-Bellman (HJB) equation for the investor's optimization problem is

$$\max\left\{\mathbf{1}_{\{H<\bar{h}B\}}BV_H+\mathcal{L}_0V,\quad -V_x+V_y+V_B,\right.$$
$$\left. f\left(0,y,B,\frac{H}{B};1\right)V_x-yV_y-\left(B-\omega(B-y)^+\right)V_B-HV_H\right\}=0 \tag{10}$$

in the region where $H>0,\ B>0,\ y>0$, and $f(x,y,B,H/B;1)>0$, where

$$\mathcal{L}_0V=\frac{1}{2}\sigma^2y^2V_{yy}+\mu yV_y+((1-\tau_i)rx+(1-\tau_d)\delta y)V_x-(\beta+\lambda)V$$
$$+\alpha^{\frac{1}{\gamma}}\frac{\gamma}{1-\gamma}(V_x)^{-\frac{1-\gamma}{\gamma}}+\frac{(1-\alpha)\lambda}{1-\gamma}f(x,y,B,H/B;\iota)^{1-\gamma}.$$

Using the homogeneity property of the value function, we can reduce the dimensionality of the problem by the following transformation:

$$z=\frac{x}{y},\quad b=\frac{B}{y},\quad h=\frac{H}{B},\quad V(x,y,B,H)=y^{1-\gamma}\Phi(z,b,h),$$

for some function Φ, where b is equal to the average basis divided by the stock price and will be simply referred to as the basis-price ratio. Then Equation (10) can be reduced to

$$\max\left\{\mathbf{1}_{\{h<\bar{h}\}}\Phi_h+\mathcal{L}_1\Phi,(1-\gamma)\Phi-(z+1)\Phi_z+(1-b)\Phi_b-\frac{h}{b}\Phi_h,\right.$$
$$\left.-(1-\gamma)\Phi+f(z,1,b,h;1)\Phi_z+\omega(b-1)^+\left(\Phi_b-\frac{h}{b}\Phi_h\right)\right\}=0 \tag{11}$$

in the region where $h>0,\ b>0,\ f(z,1,b,h;1)>0,$ where

$$\begin{aligned}\mathcal{L}_1\Phi=&\frac{1}{2}\sigma^2z^2\Phi_{zz}+\frac{1}{2}\sigma^2b^2\Phi_{bb}+\sigma^2zb\Phi_{zb}-\left(\mu-\gamma\sigma^2\right)b\Phi_b\\&-\left[(\mu-(1-\tau_i)r-\gamma\sigma^2)z-(1-\tau_d)\delta\right]\Phi_z\\&+\left[(1-\gamma)(\mu-\frac{1}{2}\gamma\sigma^2)-\beta-\lambda\right]\Phi\\&+\frac{\gamma\alpha^{1/\gamma}}{1-\gamma}(\Phi_z)^{-\frac{1-\gamma}{\gamma}}+\frac{(1-\alpha)\lambda}{1-\gamma}f(z,1,b,h;\iota)^{1-\gamma}.\end{aligned}$$

The optimal trading strategy of the investor can be characterized by a no-trade region NT, a buy region BR, and a sell region SR, which are defined as follows:

$$NT=\left\{(z,b,h):\ \begin{array}{l}(1-\gamma)\Phi-(z+1)\Phi_z+(1-b)\Phi_b-\frac{h}{b}\Phi_h<0,\\-(1-\gamma)\Phi+f(z,1,b,h;1)\Phi_z+\omega(b-1)^+\left(\Phi_b-\frac{h}{b}\Phi_h\right)<0\end{array}\right\},$$

$$BR=\left\{(z,b,h):(1-\gamma)\Phi-(z+1)\Phi_z+(1-b)\Phi_b-\frac{h}{b}\Phi_h=0\right\},\ \text{and}$$

$$SR=\left\{(z,b,h):-(1-\gamma)\Phi+f(z,1,b,h;1)\Phi_z+\omega(b-1)^+\left(\Phi_b-\frac{h}{b}\Phi_h\right)=0\right\}.$$

Out of the no-trade region, buying to the buy boundary of NT, selling to the sell boundary of NT, or liquidating a fraction of the current position and then buying back some shares is optimal. We provide a verification theorem for the optimality of this trading strategy with its proof and a numerical algorithm for solving the investor's problem in the Appendix.

To provide a baseline result for computing the value of deferring tax realization, we next analyze the optimal strategy within the class of strategies that never defer any capital gains tax. Proposition 1 shows that within this class of trading strategies, keeping a constant fraction of after-tax wealth in stock is optimal in the FR case.

Proposition 1. Assume $\omega=0$, $\iota=1$ and

$$\rho\equiv\beta+\lambda-(1-\gamma)\left((1-\tau_i)r+\frac{((1-\tau_S)\mu+(1-\tau_d)\delta-(1-\tau_i)r)^2}{2\gamma(1-\tau_S)^2\sigma^2}\right)>0.$$

Within the class of strategies that never defer any capital gains tax, investing and consuming a constant fraction of after-tax wealth are optimal, where the optimal fractions are

$$\frac{y_t}{x_t+y_t-\tau_S(y_t-B_t)}=\frac{(1-\tau_S)\mu+(1-\tau_d)\delta-(1-\tau_i)r}{\gamma(1-\tau_S)^2\sigma^2},$$

$$\frac{c_t}{x_t+y_t-\tau_S(y_t-B_t)}=\alpha^{\frac{1}{\gamma}}\nu^{-\frac{1-\gamma}{\gamma}},$$

and the associated value function is

$$\frac{[\nu(x+y-\tau_S(y-B))]^{1-\gamma}}{1-\gamma},$$

where ν is the unique positive root of

$$-\rho\nu^{1-\gamma}+\gamma\alpha^{\frac{1}{\gamma}}\nu^{-\frac{(1-\gamma)^2}{\gamma}}+(1-\alpha)\lambda=0. \tag{12}$$

The following proposition indicates that in the FC case, if there is capital loss, then continuous trading is optimal when $h=0$ or tax rates are symmetric. This result is useful for finding numerical solutions for the FC case because it provides a way to compute the solution at the initial point $h=0$, which we can then use for solving the problem with a positive holding period.

Proposition 2. Suppose $\omega=1$ and, in addition, $h=0$ or $\tau_L=\tau_S$. Let $\Phi(z,b,h)$ be a solution to HJB Equation (11). Then,

1. at $h=0$,
$$\Phi(z,b,0)=(z+1)^{1-\gamma}\zeta(\theta),\text{for } b>1,$$
where $\theta=\frac{z+1}{z+b}\in[0,1]$ and $\zeta(\theta)$ satisfies Equations (A-2)–(A-3) in the Appendix.
2. the optimal trading strategy when there is a capital loss is to trade continuously to keep the fraction of wealth in stock $\frac{y}{x+y}$ equal to $\pi^{**}(b)$ as defined below Equation (A-4).

3. Numerical Analysis

In this section, we provide some numerical analysis on the solution of the investor's problem.

3.1 Optimal trading boundaries

In this subsection, we set the default parameter values as follows: relative risk aversion coefficient $\gamma = 3$, jump time intensity $\lambda = 0.04$ (i.e., an average investment horizon of 25 years), subjective discount rate $\beta = 0.01$, interest rate $r = 0.03$, expected stock return $\mu = 0.07$, dividend yield $\delta = 0.02$, stock return volatility $\sigma = 0.2$, intertemporal consumption utility weight $\alpha = 0.9$, short-term tax rate $\tau_S = 0.35$, long-term tax rate $\tau_L = 0.15$, interest and dividend tax rates $\tau_i = \tau_d = \tau_S$, threshold for long-term status $L = 1$, the upper bound for the average holding period $\bar{h} = 1.5$, and $\iota = 0$ (i.e., tax is forgiven at death). We also provide results with a different set of parameters to show comparative statics and robustness to the choice of parameter values.

Figure 1 plots the optimal trading boundaries against the basis-price ratio b for the FC case, with the round dots representing the optimal positions at $b = 1$ and $h = 0$. The vertical axis denotes the fraction of after-tax wealth invested in stock—that is, $\pi \equiv \frac{y}{f(x,y,b,h;1)}$. When tax rates are zero, we have the standard Merton solution where the investor invests a constant fraction 50% of wealth in the stock, as indicated by the thin Merton lines in Figures 1(a) and 1(b).

With positive tax rates, the investor may defer realizing capital gains as indicated by the no-trade regions in Figure 1. In addition, as tax rates increase, the buy boundaries go down and the sell boundaries go up due to the higher cost from realizing gains, as Figures 1(a) and 1(b) suggest. Deferring the realization of capital gains has three benefits. First, it defers the tax payment, and thus gains on the time value. Second, if the long-term rate is lower than the short-term rate, then deferring realization until it becomes long term enables the investor to realize gains at the lower long-term rate. Third, it can make some of the future losses rebatable.

To understand the third benefit, suppose the investor holds one share with capital gain. If the investor realizes the gain, pays the tax, and buys back some shares, but the stock price drops subsequently, then the investor can only carry forward the loss. If, instead, the investor did not realize the gain, then the subsequent loss from the drop in the stock price would offset some of the original gain, thus reducing the amount of remaining gain that is subject to tax and effectively making the subsequent loss rebatable. As we will show later, this third benefit can be the main source of the benefits from deferring tax in the FC case, even with asymmetric tax rates.

Even though deferring capital gains realization can have significant benefits, Figure 1 implies that realizing capital gains even when the time value is positive can still be optimal because the NT regions are bounded above. Intuitively, the no-trade region reflects the trade-off between the benefit of the deferral of capital gains and the cost of suboptimal risk exposure. When the fraction of wealth in stock is too high relative to the optimal fraction in the absence of tax, the cost of suboptimal risk exposure is greater than the benefit of the deferral. Therefore, the investor sells the stock to reduce the risk exposure. More specifically, if the fraction of wealth in stock is (vertically) above the sell boundary, then the

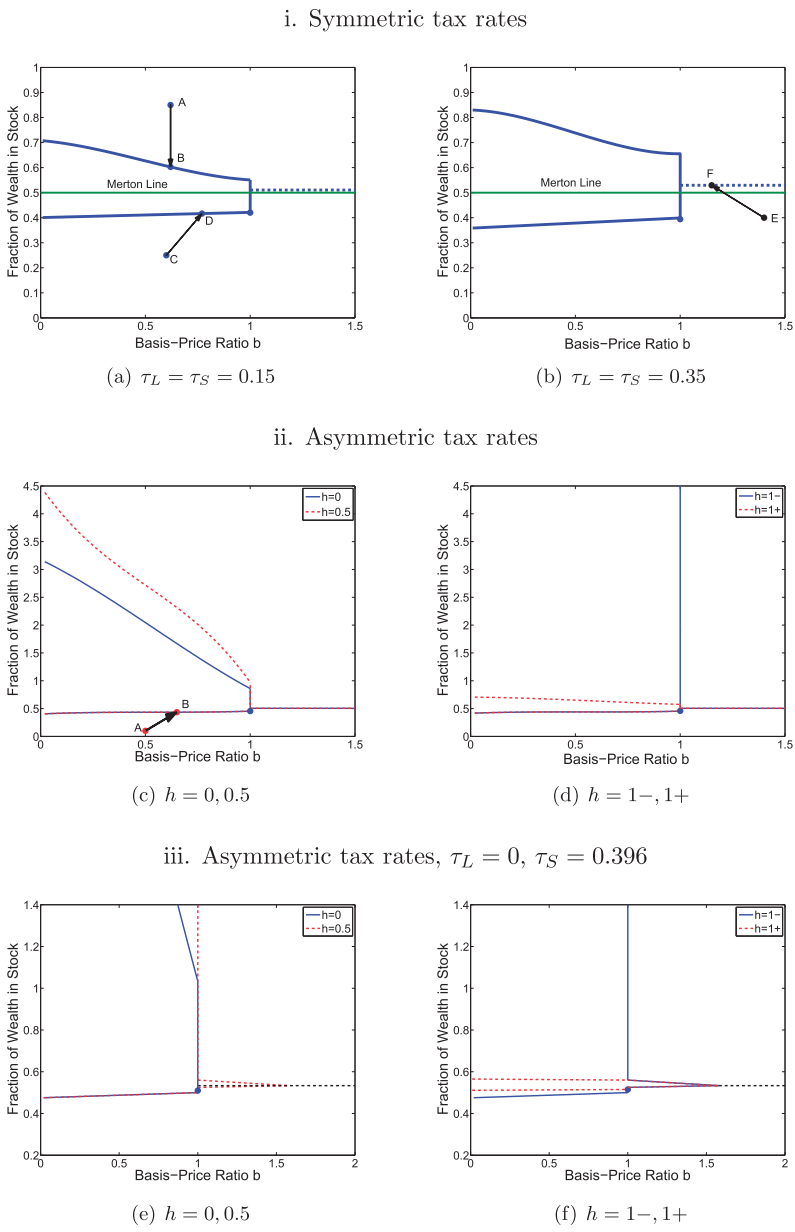


Figure 1
Optimal trading boundaries against basis-price ratio b, the FC case
Parameter default values: $\omega=1$, $\gamma=3$, $\lambda=0.04$, $\beta=0.01$, $\iota=0$, $L=1$, $\bar{h}=1.5$, $r=0.03$, $\mu=0.07$, $\sigma=0.2$, $\alpha=0.9$, $\delta=0.02$, $\tau_i=\tau_d=\tau_S=0.35$, $\tau_L=0.15$, and $\kappa=1$.

investor sells a minimum amount (and thus realizes some capital gains) to reach the sell boundary. The trading direction is vertically downward (e.g., A to B in Figure 1(a)) in the figures because as the investor sells, the total basis (B) and the dollar amount in the stock (y) decrease by the same proportion, and thus the basis-price ratio b does not change. However, if the fraction of wealth in the stock is (vertically) below the buy boundary, then the investor buys enough to reach the buy boundary (e.g., C to D in Figure 1(a) or E to F in Figure 1(b)). The direction of trade is no longer vertical because both the total basis B and the dollar amount y increase by the same dollar amount (instead of the same proportion), and thus the new basis-price ratio b gets closer to 1.

Consistent with the finding of Marekwica (2012), Figures 1(a) and 1(b) imply that investors should realize losses immediately. More specifically, if there is a net capital loss (i.e., $b > 1$), then as shown in Proposition 2, investors should continuously realize additional losses and gains to stay at the dotted lines. Even though capital losses are not eligible for a tax rebate, loss realization has a benefit of achieving a better risk exposure sooner. When the investor has subsequent positive gains, these gains offset some carried losses, and the investor's stock position moves left toward the $b = 1$ line. The distance between the optimal fraction at $b = 1$ (denoted by the round dot) and the dotted line for $b > 1$ suggests that the optimal fraction of wealth invested in the stock is discontinuous at $b = 1$. This is because the investor needs to pay tax for capital gains but can only carry forward capital losses. Because of this discrepancy, the investor tends to invest less with gains and more with losses, which can offset some subsequent gains. This asymmetric treatment of gains and losses also makes it optimal to defer even tiny capital gains (i.e., b is close to 1) as long as the fraction of wealth in stock is not too high.

Figures 1(c) through 1(f) plot the optimal trading boundaries for the asymmetric tax rates case. Compared with the symmetric tax case with $\tau_L = \tau_S = 0.35$, the no-trade region is much wider, which reflects the greater benefit of deferring capital gains tax in order to realize gains at the lower long-term tax rate. Still, in contrast to Constantinides (1984) and Dammon and Spatt (1996), Figure 1(c) implies that it can be optimal to realize short-term gains even when the long-term rate is much lower than the short-term rate, as long as the fraction of wealth in the stock becomes too high relative to the optimal risk exposure in the absence of tax and the holding period is not too close to one year. Different from the symmetric tax rate case, when the investor has held shares for some time ($h > 0$) and buys some additional shares, the purchase shortens the average holding period of the new position, and thus the end point (e.g., Point B in Figure 1(c)) lies on the buy boundary for some holding period $h < 0.5$, although this is not obvious in Figure 1(c) because the buy boundary at $h = 0$ is close to the boundary at $h = 0.5$.

When long-term rates are significantly lower than short-term rates and there is a capital gain ($b < 1$), the sell boundary goes up dramatically as the holding period increases because the benefit of deferring tax becomes greater. Figure

1(d) shows that just before the position turns long-term ($h = 1-$), the investor does not sell at all, even with a huge fraction of wealth invested in the stock, due to the imminent long-term status that entitles the investor to a lower tax rate. However, the buy boundary stays close to the Merton line and is almost insensitive to change in the holding period even when the position is about to become long term. This is because (i) buying stock gets the position closer to the Merton line and does not trigger capital gains tax, and (ii) as explained below, it is optimal to buy additional shares to stay close to the Merton line even when the gain becomes long term, as implied by the long-term buy boundary.

Although Figures 1(c) and 1(d) may suggest that realizing all losses immediately is optimal even with asymmetric tax rates, this is not true in general. Indeed, Figures 1(e) and 1(f) show that if the difference between the long-term rate and short-term rate is large enough, then deferring even large capital losses may be optimal. This is because deferring realization would entitle the investor to the much lower long-term rate sooner when the stock price rises and current losses turn into gains.

In contrast to the existing literature (e.g., Constantinides 1984), Figures 1(d) and 1(f) show that deferring even large long-term capital gains ($h > 1$) can also be optimal, although the no-trade region shrinks significantly for $h > 1$ because of the lower long-term capital gains tax rate. As discussed before, by deferring gains the investor can effectively make incremental losses rebatable. Therefore, a benefit of deferring the realization of any gains, long term or short term, always exists. To ensure that this benefit dominates the cost of having suboptimal risk exposure, the investor keeps the fraction of wealth in stock close to the Merton line by buying or selling whenever the fraction gets out of the narrow no-trade region. The standard argument for immediately realizing all long-term gains and losses is that one can reestablish short-term status so that the subsequent losses can be rebated at the higher short-term rate. But this argument no longer holds when the rebate is limited and relatively unimportant, as in the case for high-income investors. Given that high-income investors are still entitled to some tax rebate in practice, our findings suggest that it is optimal for these investors to realize a small fraction of long-term gains and losses to catch the limited rebate benefit but defer the rest.

Figure 2 plots the optimal trading boundaries against the basis-price ratio b for the FR case. Figures 2(a) and 2(b) show that with symmetric tax rates, the entire region with capital losses (i.e., $b > 1$) belongs to the sell region, which implies that immediately realizing any capital losses is always optimal, as predicted by the existing literature. In addition to the benefit of reducing the duration of a suboptimal position, immediately realizing losses can also earn interest on the tax rebate sooner. As the tax rate increases, the no-trade region widens due to the increased benefit of deferring.

As in the FC case, Figure 2(c) shows that if the long-term rate is lower than the short-term rate, then the no-trade region becomes much wider because the benefit of deferring short-term gains increases due to the lower long-term rate.

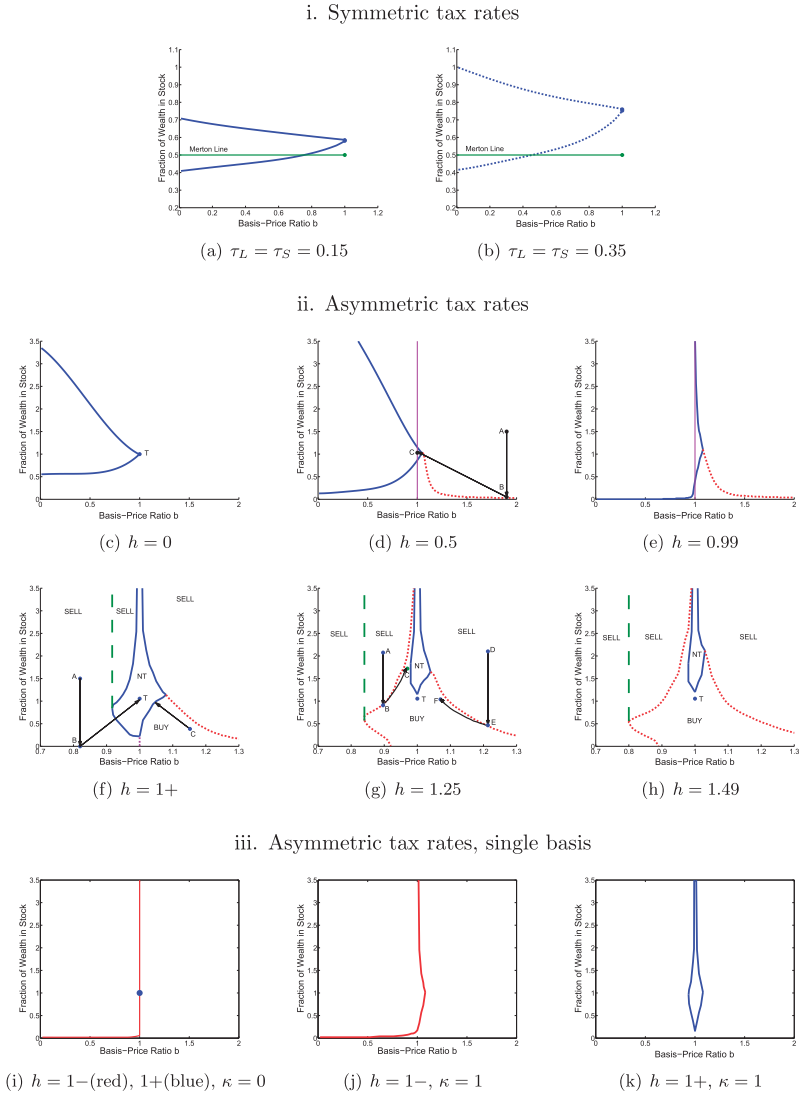


Figure 2
Optimal trading boundaries against basis-price ratio b, the FR case
Parameter default values: ω=0, γ=3, β=0.01, ι=0, λ=0.04, L=1, $\bar{h}$=1.5, r=0.03, μ=0.07, σ=0.2, α=0.9, δ=0.02, $\tau_i = \tau_d = \tau_S$=0.35, τ_L=0.15, and κ=1.

Figures 2(d) and 2(e) indicate that as in the FC case, the investor may find it optimal to defer loss realizations if he or she has already held the stock for some time (e.g., $h \geq 0.5$). In addition, as the holding period increases, the no-trade region widens, and when the holding period gets close to one year, it is rarely

optimal for the investor to realize short-term gains, as in the FC case. In contrast to the FC case, however, the buy boundary lowers significantly as the holding period increases. This is because with full rebate, realizing all large long-term gains to reestablish the short-term status is optimal, as explained below, and buying additional shares would shorten the average holding period and defer realization of gains at the lower long-term rate.

At $h=0$, the investor trades to Point T to realize all losses immediately, as in the symmetric rate case. When there is a large loss and $h>0$ (e.g., Point A in Figure 2(d)), an investor should realize some of the loss by selling (vertically) to the red curve and then buy back some shares. For example, at Point A in Figure 2(d), it is optimal to first sell to Point B and then buy to Point C, which is close to Point T in Figure 2(c). The reason that C is inside the no-trade region for $h=0.5$ is again that any purchase reduces the average holding period. Realizing all losses is not optimal when $h>0$ because the utility strictly increases in the average holding period, and by realizing only part of the losses, the average holding period for the new position after realization remains greater than zero.

Figures 2(c) through 2(h) imply that the optimal trading strategies for short-term status and long-term status are qualitatively different. An investor tends to defer realization of large short-term capital gains, as reflected by the wider no-trade region when b is small and $h<1$, but immediately realizes at least some large long-term capital gains. For example, Figure 2(f) shows that in the region to the left of the dashed line, selling the entire position before buying back some shares is optimal (e.g., A to B to T, where T is the same point as that in Figure 2(c)), while Figure 2(g) shows that in the sell region just to the right of the dashed line, it is optimal to sell a fraction of the current position before buying back a certain amount (e.g., A to B to C in Figure 2(g)). If an investor deferred a large long-term gain and experienced an incremental loss due to a subsequent price drop, but the investor still had a net gain because the original gain was large, then effectively the investor can use the entire incremental loss to offset the original gain, making it equivalent to being rebated at the lower long-term rate. In contrast, if it is realized, then the incremental loss can be rebated at the higher short-term rate.

Still, because the investor realizes only part of the long-term gains and losses (e.g., D to E to F in Figure 2(g)) except for huge long-term gains (to the left of the dashed lines), deferring some large long-term gains and losses is optimal, as in the FC case. The main intuition is as follows. First, long-term status strictly dominates short-term status because long-term gains can be realized at the lower long-term rate, and long-term losses can be rebated at the same short-term rate. Therefore, keeping long-term status by deferring long-term gains and losses always has a benefit. Second, the cost of deferring long-term gains explained in the previous paragraph is less for smaller gains. This is because an incremental loss can more likely turn a smaller gain into a net loss, and if the net loss is realized, then the part of the incremental loss that exceeds the

original gain is rebated at the short-term rate, while for a large gain, the entire incremental loss is effectively rebated at the lower long-term rate because a net gain still occurs after offset by the incremental loss.

As far as we know, all the existing literature on optimal consumption and investment with capital gains tax assumes long-term tax rates apply to long-term losses (i.e., $\kappa = 0$), while the tax code dictates that short-term rates apply (i.e., $\kappa = 1$). Does the assumption of $\kappa = 0$ significantly affect the optimal trading strategies? What is the value of being able to realize long-term losses at the short-term rate instead of at the long-term rate? We find that with the average-holding-period approximation, whether the short-term or long-term rate applies to long-term losses does not significantly change either the optimal trading strategy or the expected utility. This is because an investor with a long-term loss position can first change the position from long term to short term by buying enough additional shares and then realize the losses at the short-term rate. However, in practice, while buying shares changes the average holding period, it does not change the tax rate that applies to the shares with long-term losses. This suggests that for the purpose of examining the impact of different rates applicable to long-term losses, restricting feasible trading strategies to the class of single-basis strategies discussed before may yield a better assessment of this impact. This is because within the class of single-basis strategies, if the investor defers short-term losses to long term, then as in practice, these losses will stay long term until entirely realized. We find that with the restriction to single-basis strategies, when the holding period is short (e.g., $h = 0$, $h = 0.5$), the trading strategies for $\kappa = 1$ and $\kappa = 0$ are still virtually the same, as in the unrestricted case. However, Figures 2(i)–2(k) show that as the holding period approaches or exceeds the long-term threshold of one year, the optimal trading strategies for $\kappa = 1$ and $\kappa = 0$ become significantly different. Figure 2(i) suggests that when the holding period is just below 1 ($h = 1^-$), realizing all losses and deferring almost all gains is optimal if the long-term rate applies to long-term losses ($\kappa = 0$). In addition, right after the status becomes long term (i.e., $h = 1^+$), the investor should realize all long-term gains and losses immediately and rebalance to the dotted position, consistent with the recommendation of the existing literature. In contrast, if the short-term rate applies to long-term losses ($\kappa = 1$), the investor should defer some long-term gains and losses, as in the unrestricted case. However, the utility loss from having the long-term rate applying to long-term losses is still small. For example, for the default parameter values, it is about 0.5% of the initial wealth because an investor defers only small long-term losses in the single-basis case, as shown in Figure 2(k). Even though the utility loss from applying long-term rates to long-term losses is relatively small, it is still important that we model short-term rates as applicable to both short-term losses and long-term losses (as in the tax code). One reason is that doing so helps understand why it can be optimal for investors to defer long-term capital gains and losses and shows that our main results are not caused by assumptions that are inconsistent with the tax code.

Table 1
Three tax brackets and CEWLs

Ordinary income levels	τ_i	τ_d	τ_S	τ_L	CEWL
Lower income (low rates)	0.1	0	0.1	0	0.84% (FR)
Medium income (medium rates)	0.25	0.15	0.25	0.15	2.99% (FC)
High income (high rates)	0.434	0.238	0.434	0.238	5.20% (FC)

Parameter default values: $\gamma=3$, $\beta=0.01$, $L=1$, $\bar{h}=1.5$, $r=0.03$, $\mu=0.07$, $\sigma=0.2$, $\alpha=0.9$, $\delta=0.02$, $\iota=0$, $\kappa=1$, $\lambda=0.04$, $x_{0-}=1$, and $y_{0-}=0$.

So far we have shown that the optimal trading strategies are qualitatively different from those of the existing literature. We next examine whether it is also economically important to adopt the optimal strategies. For this purpose, we compute the certainty equivalent wealth loss (CEWL) as a fraction of the initial wealth from following the simple strategy most of the existing literature recommends: realize all losses and all long-term gains immediately, but defer all short-term gains. In Table 1, we report the tax rates and the CEWLs for investors at three ordinary income levels. Table 1 suggests that adopting the suboptimal strategy can be costly, especially for high-income investors for whom most losses are not tax deductible. For example, high-income investors are willing to pay as much as 5.20% of their initial wealth to adopt the optimal strategy. Even for lower-income investors, the value is still about 0.84% of their initial wealth. Further analysis shows that the main source of the cost comes from immediately realizing all long-term gains, whereas the optimal strategy is to defer some long-term gains, as shown in Figures 1 and 2.

3.2 Would higher capital gains tax rates make lower-income investors better off?

Because short-term capital gains and interest tax rates are set to investors' marginal ordinary income tax rates in the current tax code, these rates for lower-income investors are lower than those for higher-income investors (e.g., see Table 1). Because lower-income investors likely can have a high percentage of their capital losses rebated at the short-term tax rate and defer short-term capital gains to long term, higher short-term tax rates would enable them to get more rebate in case of losses, largely avoid paying tax at the short-term rates in case of positive gains, and thus could make them better off. Furthermore, even when other capital gains tax rates (e.g., long-term and dividend tax rates) are also increased, lower-income investors may still be better off because marginal utility of wealth is higher when there is a capital loss, and thus the tax rebate effect of a higher short-term rate can dominate.

To investigate this possibility, we compute the certainty equivalent wealth gain (CEWG, in terms of the fraction of the initial wealth) of a lower-income investor with the investor's short-term capital gains and interest tax rates increased to higher rates, keeping constant everything else (including ordinary income tax rate). More specifically, let $V_L(x, y, B, H)$ denote the value function of a lower-income investor with the tax rates as in the "Low rates" row of Table 1

Table 2
CEWGs and optimal initial policies of a lower-income investor with different short-term capital gains and interest tax rates

Cases	CEWG Δ		High rates		Medium rates		Low rates	
	L→M	L→H	$\frac{y_0^*}{x_0^*+y_0^*}$	$\frac{c_0^*}{x_0^*+y_0^*}$	$\frac{y_0^*}{x_0^*+y_0^*}$	$\frac{c_0^*}{x_0^*+y_0^*}$	$\frac{y_0^*}{x_0^*+y_0^*}$	$\frac{c_0^*}{x_0^*+y_0^*}$
Base case	0.315	0.914	1.390	0.078	0.930	0.061	0.610	0.051
$\delta=0$	0.223	0.688	1.070	0.061	0.690	0.049	0.430	0.043
$\delta=0.01$	0.272	0.808	1.230	0.069	0.810	0.055	0.530	0.046
$\mu=0.06$	0.291	0.873	1.270	0.071	0.830	0.055	0.530	0.047
$\sigma=0.3$	0.271	0.678	0.690	0.060	0.450	0.049	0.270	0.043
$\alpha=0.5$	0.315	0.914	1.390	0.078	0.930	0.061	0.610	0.051
$\lambda=0.1$	0.222	0.634	1.370	0.098	0.930	0.081	0.610	0.071
$\lambda=0.2$	0.149	0.417	1.350	0.131	0.910	0.114	0.610	0.104
$\gamma=5$	0.172	0.513	0.790	0.060	0.530	0.049	0.370	0.043
$\iota=1$	0.315	0.914	1.390	0.078	0.930	0.061	0.610	0.051
$\kappa=0$	0.315	0.913	1.390	0.078	0.930	0.061	0.600	0.051
$\delta=0$ $\mu=0.06$	0.185	0.611	0.960	0.055	0.590	0.045	0.350	0.040
$\iota=1$ $\lambda=0.5$	0.074	0.203	1.290	0.230	0.870	0.213	0.590	0.204
$\iota=1$ $\lambda=1$	0.039	0.105	1.210	0.394	0.830	0.378	0.570	0.368

This table presents CEWGs of a lower-income investor with long-term capital gains and dividend tax rates as in the "Low rates" row of Table 1 but with short-term capital gains and interest tax rates increased to those in the "Medium rates" row or in the "High rates" row of Table 1, in addition to optimal time 0 fractions of wealth invested in stock $\left(\frac{y_0^*}{x_0^*+y_0^*}\right)$ and time 0 consumption to wealth ratios $\left(\frac{c_0^*}{x_0^*+y_0^*}\right)$ of the investor given these different rates. Base case parameter values: $\omega=0$, $\gamma=3$, $\beta=0.01$, $L=1$, $\bar{h}=1.5$, $r=0.03$, $\mu=0.07$, $\sigma=0.2$, $\alpha=0.9$, $\delta=0.02$, $\iota=0$, $\kappa=1$, $\lambda=0.04$, $x_{0-}=1$, and $y_{0-}=0$.

and $V_H(x,y,B,H)$ denote the value function of the same lower-income investor but with the short-term capital gains and interest tax rates increased to those in the "High rates" row or the "Medium rates" row in Table 1. Let Δ denote the time 0 CEWG of the lower-income investor with all the initial wealth in the risk-free asset—that is, Δ solves[21]

$$V_L(1+\Delta,0,0,0)=V_H(1,0,0,0).$$

For a set of parameter values, we report in Table 2 the CEWGs of a lower-income investor when the short-term capital gains and interest tax rates increase from those in the "Low rates" row of Table 1 to those in the "Medium rates" or "High rates" rows of Table 1.[22] In Table 2, we also report the optimal time 0 fractions of wealth invested in stock $\left(\frac{y_0^*}{x_0^*+y_0^*}\right)$ and time 0 consumption to wealth

[21] Because of the homogeneity of the value functions, Δ is independent of the initial wealth, which can therefore be set to 1 without loss of generality.

[22] For the case where the lower-income investor's rates are changed to the "High rates," because the tax code stipulates that the Medicare surtax of 3.8% does not apply to losses (i.e., an investor can only get a tax rebate for losses at 39.6% instead of 43.4%), we use 39.6% for capital losses and 43.4% for short-term positive capital gains in the calculations in this case in Tables 2 and 3. This requires a slight extension of our main model because the short-term rate that applies to losses differs from the short-term rate that applies to short-term positive gains. Because the extension is minor and we do not want to further complicate the notations in the main model, we did not change the main model specification.

ratios ($\frac{c_0^*}{x_0^*+y_0^*}$) of the lower-income investor facing these different tax rates.[23] Table 2 suggests that a lower-income investor can be significantly better off with higher short-term capital gains and interest tax rates. For example, in the base case, increasing the short-term and interest tax rate from 10% to the medium rate of 25% (respectively the high rate of 39.6%) is equivalent to increasing the investor's initial wealth by 31.5% (respectively 91.4%). Even when an older lower-income investor has only an expected five-year remaining lifetime ($\lambda = 0.2$), having the medium rate of 25% (respectively the high rate of 43.4%) is equivalent to an increase of 14.9% (respectively 41.7%) of the initial wealth. An increase in stock volatility or risk aversion, or a decrease in expected return or dividend yield, decreases the CEWGs because the investor invests less in stock. However, the gains are still significant. For example, when the volatility increases to 30%, the CEWG is still 27.1% for an increase to the medium rate of 25% and 67.8% for an increase to the high rate of 43.4%. With an expected time to death of 25 years, the tax forgiveness at death has almost no impact on the CEWGs or optimal consumption or optimal investment, as shown by the row with $\iota = 1$.

An investor with a higher tax rate pays more tax when he or she realizes a positive gain, and thus if liquidity reasons force investors to liquidate, the value of higher short-term rates should decrease. To examine the impact of this forced liquidation, we report the results when an investor must liquidate the entire stock position due to a large liquidity shock that occurs once or twice a year on average ($\lambda = 0.5, 1$) and needs to pay tax at liquidation (i.e., $\iota = 1$). Indeed the CEWGs are much smaller because the investor has to realize short-term gains more often, and the expected investment horizon is much shorter. For example, when the large liquidity shock occurs once a year on average ($\lambda = 1$), the CEWG of the investor from having the medium rate of 25% (resp. high rate of 43.4%) decreases to 3.9% (resp. 10.5%). On the other hand, since the expected investment horizon is now only one year, the 3.9% gain remains economically significant. These findings indicate that liquidity-induced capital gain realization is unlikely to eliminate or reverse the result that lower-income investors can be better off with higher short-term and interest tax rates. These results show that effective tax rates on equity securities for lower-income investors can actually decline as short-term capital gains tax rates increase because of the option of realizing losses short-term to get greater tax rebates and realizing positive gains long-term to avoid paying tax at the higher short-term rates.

In the above analysis, we increase only the short-term capital gains and interest tax rates to those of higher-income investors. Table 3 reports the corresponding results when we also increase long-term capital gains and dividend tax rates to those of higher-income investors as specified in Table 2.

[23] At time 0, the after-tax wealth is the same as the before-tax wealth because the initial stock holding is 0.

Table 3
CEWGs and optimal initial policies of a lower-income investor with high, medium, or low rates as in Table 1

Cases	CEWG Δ		High rates		Medium rates		Low rates	
	L→M	L→H	$\frac{y_0^*}{x_0^*+y_0^*}$	$\frac{c_0^*}{x_0^*+y_0^*}$	$\frac{y_0^*}{x_0^*+y_0^*}$	$\frac{c_0^*}{x_0^*+y_0^*}$	$\frac{y_0^*}{x_0^*+y_0^*}$	$\frac{c_0^*}{x_0^*+y_0^*}$
Base case	−0.049	0.094	1.090	0.054	0.770	0.049	0.610	0.051
$\delta=0$	−0.070	0.014	0.810	0.043	0.550	0.041	0.430	0.043
$\delta=0.01$	−0.058	0.056	0.950	0.048	0.650	0.045	0.530	0.046
$\mu=0.06$	−0.056	0.067	0.970	0.049	0.670	0.045	0.530	0.047
$\sigma=0.3$	−0.071	0.007	0.510	0.043	0.350	0.041	0.270	0.043
$\alpha=0.5$	−0.047	0.094	1.090	0.054	0.770	0.049	0.610	0.051
$\lambda=0.1$	−0.033	0.067	1.090	0.074	0.770	0.069	0.610	0.071
$\lambda=0.2$	−0.023	0.045	1.070	0.107	0.740	0.102	0.610	0.104
$\gamma=5$	−0.074	−0.017	0.630	0.042	0.450	0.040	0.370	0.043
$\iota=1$	−0.047	0.094	1.090	0.054	0.770	0.049	0.610	0.051
$\kappa=0$	−0.047	0.094	1.090	0.054	0.770	0.049	0.600	0.051
$\delta=0$ $\mu=0.06$	−0.081	−0.022	0.690	0.040	0.450	0.038	0.350	0.040
$\iota=1$ $\lambda=0.5$	−0.012	0.022	1.030	0.207	0.730	0.202	0.590	0.204
$\iota=1$ $\lambda=1$	−0.007	0.011	0.970	0.371	0.690	0.367	0.570	0.368

This table presents CEWGs of a lower-income investor with the low rates as in Table 1 changed to the medium rates or high rates as in Table 1, in addition to optimal time 0 fractions of wealth invested in stock $\left(\frac{y_0^*}{x_0^*+y_0^*}\right)$ and time 0 consumption to wealth ratios $\left(\frac{c_0^*}{x_0^*+y_0^*}\right)$ of the investor given these different rates. Base case parameter values: $\omega=0$, $\gamma=3$, $\beta=0.01$, $L=1$, $\bar{h}=1.5$, $r=0.03$, $\mu=0.07$, $\sigma=0.2$, $\alpha=0.9$, $\delta=0.02$, $\iota=0$, $\kappa=1$, $\lambda=0.04$, $x_{0-}=1$, and $y_{0-}=0$.

Table 3 shows that while the lower-income investor is worse off with the medium rates because the short-term tax rate for medium-income investors is not sufficiently high, the lower-income investor is still better off with the high rates even though the investor must pay long-term gains and dividends at significantly higher tax rates. For example, in the base case, having the high rates is equivalent to increasing the lower-income investor's initial wealth by 9.4%, even though the investor must pay 23.8% instead of 0% for long-term gains and dividends.

Tables 2 and 3 also show that with higher tax rates, a lower-income investor generally invests more and consumes more at time 0 because the after-tax stock return is less risky with higher short-term tax rates.[24] In addition, investment and consumption decrease with volatility, increase with expected return, and are almost insensitive to whether tax is forgiven at death or not.

To summarize, we show that keeping everything else constant, raising capital gains tax rates for lower-income investors can make them consume more, invest more, and become significantly better off. However, according to the current law, the tax rebate for losses is through deduction from taxable ordinary income.

[24] In contrast, we show that for a wealthy investor, both initial investment and initial consumption are almost insensitive to an increase in the short-term capital gains tax rate, mainly because realized losses are mostly not tax rebatable, and thus the after-tax riskiness of the stock is largely unchanged, as first pointed out by Domar and Musgrave (1944).

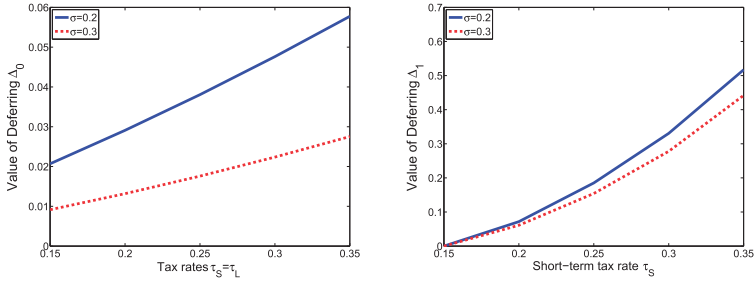
Thus, higher capital gains tax rates can benefit lower-income investors only through delinking short-term capital gains tax rates from ordinary income tax rates, such as changing the tax code from a taxable income deduction to a tax credit for capital losses.

In our model, we assume that high-income investors cannot get any tax rebate for capital losses. In practice, high-income investors can also get a tax rebate for up to $3,000 of capital losses per year. Because high-income investors invest more and are more likely to get the full benefit of the $3,000 capital loss deduction and at a higher ordinary income rate, they have a greater dollar amount benefit from the tax rebate for capital losses. Therefore, without delinking short-term capital gains tax rates from ordinary income tax rates, lower-income investors get less dollar amount benefit of a tax rebate than high-income investors get, even though the marginal utility of the benefit per dollar is higher for lower-income investors.

On the other hand, our model clearly does not consider all the factors that may decrease the benefit of higher short-term capital gains tax rates for lower-income investors. For example, some lower-income investors do not invest in stocks and their only investment income is from interest earned from savings in banks, in which case the interest would be taxed at the higher rates. Therefore, raising short-term capital gains tax rates for lower-income investors could indeed hurt some lower-income investors. Whether it is socially beneficial to raise short-term capital gains tax rates for lower-income investors is an important empirical question that is beyond the scope of this paper. On the other hand, our analysis suggests that any potential costs of raising short-term capital gains tax rates for lower-income investors should be compared with the above illustrated potential benefit.

3.3 The value of deferring capital gains realization

Because investors pay capital gains tax only when they realize capital gains, an investor has the option to defer capital gains tax. When long-term and short-term tax rates are the same, the value of this deferral comes from the time value of the capital gains tax (for both the FR and the FC cases) and from making losses effectively rebatable (for the FC case). When long-term rates are lower than short-term rates, the value of this deferral also comes from the benefit of realizing gains at the lower long-term rates. We next decompose the value of deferral into these sources to compare their relative magnitudes. More specifically, let $\underline{V}(x, y, B, H; \tau_S, \tau_S)$ be the value function when the investor cannot defer realizing any gains or losses—that is, when the investor must realize both gains and losses continuously (and thus always short term). We compute this value function using Proposition 1 for the FR case and using simulation for the FC case. Let $V(x, y, B, H; \tau_S, \tau_S)$ (respectively $V(x, y, B, H; \tau_S, \tau_L)$) be the value function when the investor can defer capital gains realization and the long-term rate is equal to the short-term rate (respectively the long-term rate is lower than the short-term rate). We use



(a) From saving the time value (b) From realizing gains at the lower long-term rate

Figure 3
The value of deferral: The FR case
Parameter default values: $\omega=1$, $\gamma=3$, $\beta=0.01$, $\iota=0$, $\lambda=0.04$, $L=1$, $\bar{h}=1.5$, $r=0.03$, $\mu=0.07$, $\alpha=0.9$, $\delta=0.02$, $\tau_i=\tau_d=\tau_S$, $\tau_L=0.15$, $\kappa=1$.

the time 0 CEWGs Δ_0 and Δ_1 to measure the values of deferral from these sources respectively, assuming all the initial wealth is in the risk-free asset—that is,

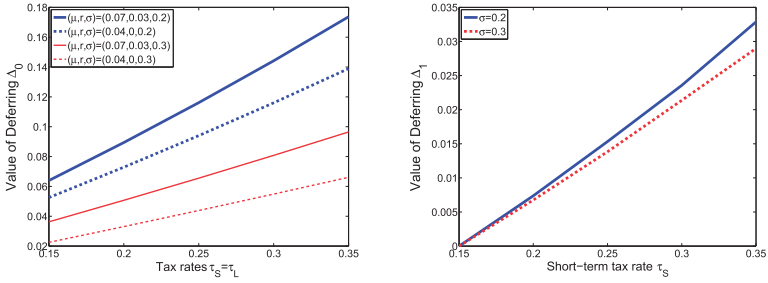
$$\underline{V}(1+\Delta_0,0,0,0;\tau_S,\tau_S)=V(1,0,0,0;\tau_S,\tau_S)$$

and

$$V(1+\Delta_1,0,0,0;\tau_S,\tau_S)=V(1,0,0,0;\tau_S,\tau_L).$$

For the FR case, Figure 3 plots the CEWG Δ_0 from saving the time value of tax (Figure 3(a)) and from realizing gains at a lower rate (Figure 3(b)), for two volatility levels $\sigma=0.2$ and $\sigma=0.3$. Figure 4 plots the corresponding results for the FC case. Figures 3(a) and 3(b) show that as the short-term tax rate increases, the values of deferral from saving the time value of tax and from realizing gains at a lower rate increase significantly, due to a higher tax cost of realizing gains. The CEWG from saving the time value varies from 2.1% to 5.8% of the initial wealth. Comparison between Figures 3(a) and 3(b) suggests that for lower-income investors, the value of deferral from realizing gains at a lower rate is much greater than that from saving the time value. For example, the CEWG from the former source can be as high as 51.7% of the initial wealth, compared with 5.8% from the latter source. Therefore, by ignoring the difference between the long-term and the short-term rates, most of the existing literature significantly underestimates the value of deferring capital gains tax and largely overestimates the effective tax rates for lower-income investors. Figure 3 also shows that as the stock volatility increases, the value of deferral decreases because the investor invests less in the stock, and thus the dollar amount of capital gains tax deferrable decreases on average.

Figure 4(a) shows that for the FC case, the CEWG Δ_0, which includes the part from saving the time value and the part from making losses effectively



(a) From saving the time value (b) From realizing gains at the lower long-term rate

Figure 4
The value of deferral: The FC case
Parameter default values: $\omega=1$, $\gamma=3$, $\beta=0.01$, $\iota=0$, $\lambda=0.04$, $L=1$, $\bar{h}=1.5$, $r=0.03$, $\mu=0.07$, $\alpha=0.9$, $\delta=0.02$, $\tau_i=\tau_d=\tau_S$, $\tau_L=0.15$, $\kappa=1$.

rebatable, ranges from 6.4% to 17.4% of the initial wealth. To help separate the time value source, we compute the values of deferral with the interest rate (and thus the time value) set to zero and plot the results using the dotted lines. Figure 4(a) indicates that the significantly greater value of deferral for the FC case compared with the FR case mainly comes from the additional benefit of making losses effectively rebatable. Compared with lower-income investors, for high-income investors, the CEWG Δ_1 from having the lower long-term rate is significantly smaller. This is because as shown in Figure 1, high-income investors defer even large long-term gains to make incremental losses effectively rebatable. Thus, these investors realize long-term gains less often and accordingly get smaller benefits from the lower long-term tax rate. Comparing Figures 4(a) and 4(b) suggests that the main source of the value of deferring for high-income investors is from making losses effectively rebatable, which differs from that for lower-income investors.

Similar to Figure 3, the value of deferral from realizing gains at a lower rate also decreases with volatility. This is because investment in the stock decreases as the volatility increases.

4. Concluding Summary

In this paper, we propose an optimal tax-timing model that takes into account three important features of the current tax law: (i) tax rates for long-term positive gains can be significantly lower than those for short-term positive gains, (ii) capital losses allowed to offset taxable ordinary income are capped at $3,000 per year, with the rest carried forward indefinitely to offset future gains and/or income, and (iii) short-term capital gains tax rates apply to both short-term and long-term losses. In contrast to the existing literature, this model can help explain the puzzle that many investors not only defer short-term losses beyond one year but also defer even large long-term losses and long-term gains. We find

that the impact of capital gains tax on lower-income investors for whom most losses are tax deductible can be qualitatively different from that on high-income investors for whom a majority of capital losses can only be carried forward. In addition, for lower-income investors, effective tax rates can decrease as short-term capital gains tax rates increase.

Appendix

A.1 Proof of Proposition 1

Let

$$W_t = x_t + y_t - \tau_S(y_t - B_t).$$

It is easy to verify that

$$dW_t = [(1-\tau_i)rx_t - c_t + ((1-\tau_S)\mu + (1-\tau_d)\delta)y_t]dt + (1-\tau_S)\sigma y_t dw_t \tag{A-1}$$

because $\tau\left(\frac{H}{B}\right) = \tau_S$. In order to never defer any capital gains tax, the investor needs to liquidate the entire stock holdings before making any purchase or sale; it follows that $y_t = B_t$, $x_t = W_t - y_t$. Problem (8) then reduces to a classic Merton's consumption-investment problem with interest rate $(1-\tau_i)r$, wealth process following Equation (A-1), and stock prices following

$$\frac{dS_t}{S_t} = [(1-\tau_S)\mu + (1-\tau_d)\delta]dt + (1-\tau_S)\sigma dw_t.$$

It is well known that Merton's problem's optimal consumption and investment strategy is as follows:

$$\begin{cases} \frac{c_t^*}{W_t} = \alpha^{\frac{1}{\gamma}} \nu^{-\frac{1-\gamma}{\gamma}}, \\ \frac{y_t^*}{W_t} = \frac{(1-\tau_S)\mu + (1-\tau_d)\delta - (1-\tau_i)r}{\gamma(1-\tau_S)^2\sigma^2}, \end{cases}$$

with ν solving Equation (12), which implies the value function $\frac{(\nu W)^{1-\gamma}}{1-\gamma}$. It is straightforward to show that Equation (12) has a unique positive root under the condition that $\rho > 0$. ■

A.2 Proof of Proposition 2

Let us start from HJB Equation (10) for $V(x, y, B, H)$. Note that by Equation (10), we have

$$-V_x + V_y + V_B \leq 0.$$

On the other hand, by Equation (10), because $\omega = 1$, we always have

$$yV_x - yV_y - yV_B - HV_H \leq 0, \text{ for } B > y.$$

Note that if $\tau_S = \tau_L$, the original problem degenerates to a single tax-rate problem; therefore, the value function V is independent of the holding time H—that is, $V_H = 0$. So,

$$V_x - V_y - V_B = 0 \text{ for } B > y,$$

when $H = 0$ or $\tau_S = \tau_L$, which implies continuous trading if $B > y$.

We will restrict attention to $B > y$. In this region, we have

$$V(x,y,B,0)=(x+y)^{1-\gamma}\zeta(\theta),\ \theta=\frac{x+y}{x+B},$$

for some function $\zeta(\cdot)$. Plugging into HJB Equation (10), we have

$$L\zeta=0 \quad \text{in} \quad 0<\theta<1, \text{with} \tag{A-2}$$

$$\zeta(0)=\frac{\nu^{1-\gamma}}{1-\gamma},\ \ \zeta(1)=\frac{M^{1-\gamma}}{1-\gamma}, \tag{A-3}$$

where ν solves Equation (12), M is such that $\Phi(z,1,0)=\frac{(M(z+1))^{1-\gamma}}{1-\gamma}$,

$$L\zeta=\frac{\alpha(c^*)^{1-\gamma}}{1-\gamma}+\frac{(1-\alpha)\lambda}{1-\gamma}+\frac{1}{2}\sigma^2(\pi^*(\theta))^2\theta^2\zeta_{\theta\theta}+\Big\{\sigma^2(\pi^*(\theta))^2(1-\gamma)$$
$$+\big[(1-\tau_i)r\big(1-\pi^*(\theta)\big)+(1-\tau_d)\delta\pi^*(\theta)-c^*\big](1-\theta)+\mu\pi^*(\theta)\Big\}\theta\zeta_\theta$$

$$+\Big\{\big[(1-\tau_i)r(1-\pi^*(\theta))+(\mu+(1-\tau_d)\delta)\pi^*(\theta)-\tfrac{1}{2}\gamma\sigma^2(\pi^*(\theta))^2-c^*\big](1-\gamma)$$
$$-\beta-\lambda\Big\}\zeta,\ \text{and}\quad c^*=[(1-\theta)\theta\zeta_\theta+(1-\gamma)\zeta]^{-1/\gamma}\alpha^{1/\gamma},$$

where

$$\pi^*(\theta)=\frac{[(1-\tau_d)\delta-(1-\tau_i)r]\theta^2\zeta_\theta-[\mu+(1-\tau_d)\delta-(1-\tau_i)r][\theta\zeta_\theta+(1-\gamma)\zeta]}{\sigma^2(\theta^2\zeta_{\theta\theta}+2(1-\gamma)\theta\zeta_\theta-\gamma(1-\gamma)\zeta)}. \tag{A-4}$$

Let $\pi\equiv\frac{y}{x+y}$. Note that because $\theta=\frac{1}{1-\pi+b\pi}$, after solving for the function ζ, we can solve for the optimal boundary $\pi^{**}(b)$ using the equality $\pi^*(\frac{1}{1-\pi^{**}+b\pi^{**}})=\pi^{**}$. ■

A.3 Verification Theorem

We now present the verification theorem.

Proposition 3. (Verification Theorem). Let $\Phi(z,b,h)$ be a solution to HJB Equation (11) satisfying certain regularity conditions. Denote $\partial B=\overline{NT}\cap BR$ and $\partial S=\overline{NT}\cap SR$. Define

$$V(x,y,B,H)=y^{1-\gamma}\Phi\left(\frac{x}{y},\frac{B}{y},\frac{H}{B}\right).$$

Assume that for any admissible controls, $(V(x_t,y_t,B_t,H_t))_{t\geq 0}$ is uniformly integrable. Then $V(x_0,y_0,B_0,H_0)$ equals the value function as defined in Equation (8), and the optimal policy is given by:

1. optimal consumption: $c_t^*=y_t^*\left(\frac{\Phi_z\left(\frac{x_t^*}{y_t^*},\frac{B_t^*}{y_t^*},\frac{H_t^*}{B_t^*}\right)}{\alpha}\right)^{-\frac{1}{\gamma}}$ with $(x_t^*,y_t^*,B_t^*,H_t^*)$ being the solution of Equations (3)–(6) given the optimal trading and consumption strategy (M_t^*,I_t^*,c_t^*).

2. optimal trading: if $\omega=1$, $b>1$, and $h=0$, then continuous trading to keep the fraction $\pi^{**}(b)$ (as defined below (A-4)) of wealth in stock; otherwise,

   (a) sell strategy M_t^*:

$$M_t^* = \int_0^t 1_{\left\{\left(\frac{x_\xi^*}{y_\xi^*}, \frac{B_\xi^*}{y_\xi^*}, \frac{H_\xi^*}{B_\xi^*}\right)\in\partial S\right\}} dM_\xi^*;$$

   (b) buy strategy I_t^*:

$$I_t^* = \int_0^t 1_{\left\{\left(\frac{x_\xi^*}{y_\xi^*}, \frac{B_\xi^*}{y_\xi^*}, \frac{H_\xi^*}{B_\xi^*}\right)\in\partial B\right\}} dI_\xi^*.$$

Proof of Proposition 3. The proof is similar to that in Davis and Norman (1990) and Cuoco and Liu (2000). Here we provide only the main steps for the proof.

Define

$$N_t = \int_0^t e^{-(\beta+\lambda)s}\left(\alpha u(c_s)+(1-\alpha)\lambda u\left(f\left(x_s, y_s, B_s, \frac{H_s}{B_s};\iota\right)\right)\right)ds$$
$$+e^{-(\beta+\lambda)t}V(x_t, y_t, B_t, H_t). \tag{A-5}$$

Let M_t^c and I_t^c be the continuous parts of M_t and I_t, respectively, and $\Delta M_s = M_s - M_{s-} \in (0,1]$ and $\Delta I_s = I_s - I_{s-}$ be the discrete jump at time s. An application of the generalized version of Ito's lemma implies that

$$\begin{aligned}
N_t = N_0 &+ \int_0^t e^{-(\beta+\lambda)s}(\mathbf{1}_{\{H_s<\bar{h}B_s\}}B_s V_H + \mathcal{L}_2 V)ds \\
&+ \int_0^t e^{-(\beta+\lambda)s}\left(f\left(0, y_s, B_s, \frac{H_s}{B_s};1\right)V_x - y_s V_y - (B_s-\omega(B_s-y_s)^+)V_B - H_s V_H\right)dM_s^c \\
&+ \sum_{0\le s\le t} e^{-(\beta+\lambda)s}\Big(V(x_{s-}+f\left(0, y_{s-}, B_{s-}, \frac{H_{s-}}{B_{s-}};1\right)\Delta M_s, y_{s-}-y_{s-}\Delta M_s, \\
&\quad B_{s-}-(B_{s-}-\omega(B_{s-}-y_{s-})^+)\Delta M_s, H_{s-}-H_{s-}\Delta M_s) - V(x_{s-}, y_{s-}, B_{s-}, H_{s-})\Big) \\
&+ \int_0^t e^{-(\beta+\lambda)s}(-V_x+V_y+V_B)dI_s^c \\
&+ \sum_{0\le s\le t} e^{-(\beta+\lambda)s}\Big(V(x_{s-}-\Delta I_s, y_{s-}+\Delta I_s, B_{s-}+\Delta I_s, H_{s-}) - V(x_{s-}, y_{s-}, B_{s-}, H_{s-})\Big) \\
&+ \int_0^t e^{-(\beta+\lambda)s}\sigma y_s V_y dw_s,
\end{aligned} \tag{A-6}$$

where

$$\mathcal{L}_2 V = \frac{1}{2}\sigma^2 y^2 V_{yy} + \mu y V_y + ((1-\tau_i)rx + (1-\tau_d)\delta y - c)V_x - (\beta+\lambda)V$$
$$+\alpha u(c) + (1-\alpha)\lambda u\left(f\left(x, y, B, \frac{H}{B};\iota\right)\right).$$

First, we show that the fourth term is nonpositive for any feasible trading strategy. By the mean-value theorem, there exists $0 \le m \le \Delta M_s \le 1$ such that the fourth term is equal to:[25]

$$\sum_{0\le s\le t} e^{-(\beta+\lambda)s}\left(f\left(0, y_{s-}, B_{s-}, \frac{H_{s-}}{B_{s-}}; 1\right)V_x - y_{s-}V_y - (B_{s-} - \omega(B_{s-} - y_{s-})^+)V_B - H_{s-}V_H\right)\Delta M_s$$

$$= \sum_{0\le s\le t} e^{-(\beta+\lambda)s}\frac{f\left(0, y_m, B_m, \frac{H_m}{B_m}; 1\right)V_x - y_m V_y - (B_m - \omega(B_m - y_m)^+)V_B - H_m V_H}{1-m}\Delta M_s$$

$$\le 0, \tag{A-7}$$

where V_x, V_y, and V_B are evaluated at (x_m, y_m, B_m, H_m) with

$$x_m = x_{s-} + f(0, y_{s-}, B_{s-}, h_{s-}; 1)m, \quad y_m = (1-m)y_{s-},$$

$$B_m = B_{s-} - (B_{s-} - \omega(B_{s-} - y_{s-})^+)m,$$

$$H_m = (1-m)H_{s-},$$

the second equality can be easily verified for (i) $\omega = 0$ or $B_{s-} \le y_{s-}$ and (ii) $\omega = 1$ and $B_{s-} > y_{s-}$, and the inequality follows from HJB Equation (10). By similar argument, we have

$$\sum_{0\le s\le t} e^{-(\beta+\lambda)s}\Big(V(x_{s-} - \Delta I_s, y_{s-} + \Delta I_s, B_{s-} + \Delta I_s, H_{s-}) - V(x_{s-}, y_{s-}, B_{s-}, H_{s-})\Big) \le 0.$$

Then N_t is a martingale under the proposed strategy and a supermartingale for any feasible strategy because V satisfies HJB Equation (10), $dM_t^c \ge 0$ and $dI_t^c \ge 0$ for any feasible M_t and I_t, and thus the second, the third, and the fifth terms are nonpositive for any feasible trading strategy and equal to zero for the proposed one; the fourth and the sixth terms are equal to zero for the proposed strategy and nonpositive for any feasible trading strategy by the above argument, and the last term is a martingale for any feasible strategy due to the lognormal distributed stock price and the boundedness of V_y.

We then have

$$\begin{aligned} &V(x_0, y_0, B_0, H_0) \\ &\quad = N_0 \ge E[N_T] \\ &\quad = E\Big[\int_0^T e^{-(\beta+\lambda)s}\big(\alpha u(c_s) + (1-\alpha)\lambda u(f(x_s, y_s, B_s, h_s; \iota))\big)ds \\ &\quad + e^{-(\beta+\lambda)T}V(x_T, y_T, B_T, H_T)\Big] \end{aligned} \tag{A-8}$$

for any feasible strategy (c_t, M_t, I_t), with equality for (c_t^*, M_t^*, I_t^*). Taking the limit as $T \to \infty$ and using the transversality condition

$$\lim_{T\to\infty} E\Big[e^{-(\beta+\lambda)T}V(x_T, y_T, B_T, H_T)\Big] = 0,$$

which follows from applying Ito's lemma to the discounted value function $e^{-(\beta+\lambda)t}V(x_t, y_t, B_t, H_t)$, the HJB equation, and the boundedness of Φ inside the no-transaction region, we get the desired result. ■

[25] If $m = 1$, the fourth term is also nonpositive because selling the entire position is a feasible strategy.

A.4 The algorithm of finding the solution numerically and its convergence

We use the following iterative algorithm to solve for Φ:

1. Set
$$M_0 = \text{initial guess}, \quad i = 0.$$

2. Given M_i, for $\omega = 0$ we denote
$$G_i(z,b,h) = \frac{(M_i f(z,1,b,h;1))^{1-\gamma}}{1-\gamma}, \quad b > 1$$
and for $\omega = 1$,
$$G_i(z,b,h) = (z+1)^{1-\gamma}\zeta\left(\frac{z+1}{z+b}\right), \quad b > 1$$
where ζ is obtained by solving Equation (A-2) with (A-3) and $\zeta(1) = \frac{M_i^{1-\gamma}}{1-\gamma}$.[26]
For b large enough, set boundary condition $\Phi(z,b,h) = G_i(z,b,h)$ and for $h \le \bar{h}$ solve
$$\max\Bigg\{\mathbf{1}_{\{h<\bar{h}\}}\Phi_h + \mathcal{L}_1\Phi,$$
$$(1-\gamma)\Phi - (z+1)\Phi_z + (1-b)\Phi_b - \frac{h}{b}\Phi_h,$$
$$-(1-\gamma)\Phi + f(z,1,b,h;1)\Phi_z + \omega(b-1)^+\left(\Phi_b - \frac{h}{b}\Phi_h\right)\Bigg\} = 0. \tag{A-9}$$
Denote the solution as Φ_i.

3. Set
$$M_{i+1} = \left((1-\gamma)\sup_{k\in(-1,+\infty)}(k+1)^{\gamma-1}\Phi_i(k,1,0)\right)^{1/(1-\gamma)}.$$

4. If $|M_{i+1} - M_i| <$ tolerance, then stop and set $\Phi = \Phi_i$; otherwise, set $M_i = M_{i+1}$, $i = i+1$, and go to Step 2.

We show next as long as the initial guess M_0 is small enough,[27] the above iterative procedure yields a monotonically increasing sequence $\{M_i\}_{i=1,2,\ldots}$.

For $\omega = 0$ denote
$$\bar{G}_i(z,b,h) = \frac{(M_i f(z,1,b,h;1))^{1-\gamma}}{1-\gamma},$$
and for $\omega = 1$
$$\bar{G}_i(z,b,h) = \begin{cases} \dfrac{(M_i f(z,1,b,h;1))^{1-\gamma}}{1-\gamma}, & b \le 1 \\ (z+1)^{1-\gamma}\zeta\left(\dfrac{z+1}{z+b}\right), & b > 1, \end{cases}$$
where $i \ge 0$.

[26] See the proof of Proposition 2 for justification.

[27] For example, one can set $M_0 = \nu$, where ν solves Equation (12) with $\rho = \beta + \lambda - (1-\gamma)(1-\tau_i)r$, corresponding to the feasible strategy of investing only in the risk-free asset.

We start from M_0. Because $\bar{G}_0(z,b,h)$ satisfies

$$(1-\gamma)\Phi-(z+1)\Phi_z+(1-b)\Phi_b-\frac{h}{b}\Phi_h=0, \tag{A-10}$$

it is a subsolution to Equation (A-9) with boundary condition

$$\Phi(z,b,h)=G_0(z,b,h)$$

for b large enough. Because Φ_0 is its solution, by the comparison principle we have

$$\Phi_0(z,b,h)\geq\bar{G}_0(z,b,h).$$

As a result, from Step 3,

$$\begin{aligned}M_1&=\left((1-\gamma)\sup_{k\in(-1,+\infty)}(k+1)^{\gamma-1}\Phi_0(k,1,0)\right)^{1/(1-\gamma)}\\&\geq\left((1-\gamma)\sup_{k\in(-1,+\infty)}(k+1)^{\gamma-1}\bar{G}_0(k,1,0)\right)^{1/(1-\gamma)}\\&=M_0.\end{aligned}$$

By similar arguments on $\bar{G}_i$ and Φ_i, we can also prove that

$$M_{i+1}\geq M_i.$$

Hence, $\{M_i\}_{i=1,2,\ldots}$ is a monotonically increasing sequence. For convergence, it remains to find an upper bound of the sequence.

Let $\tilde{\Phi}(z,b,h)$ be the solution to HJB Equation (11), as given in the verification theorem. For any $i\geq 0$, we see from the first part of the proof that $\bar{G}_i(z,b,h)$ satisfies Equation (A-10). Therefore, it is a subsolution to the transformed Equation (11). By applying the maximum principle (e.g., Friedman 1982), we see that

$$\bar{G}_i(z,b,h)\leq\tilde{\Phi}(z,b,h)\text{ for all }z,b,h,i.$$

Therefore,

$$\frac{M_i^{1-\gamma}}{1-\gamma}\leq\sup_{k\in(-1,+\infty)}(k+1)^{\gamma-1}\tilde{\Phi}(k,1,0),$$

or equivalently,

$$M_i\leq\left((1-\gamma)\sup_{k\in(-1,+\infty)}(k+1)^{\gamma-1}\tilde{\Phi}(k,1,0)\right)^{1/(1-\gamma)}$$

for all i, which is as desired. ■

A.5 Is the assumption of average basis and average holding period a good approximation? To keep tractability, we approximate the exact basis and exact holding period with the average basis and average holding period of the current position, respectively. If an investor always holds a single cost basis, then our approximation would be without any error. Thus, the finding of DeMiguel and Uppal (2005) that investors rarely hold more than one cost basis suggests that our assumption might be reasonable. However, DeMiguel and Uppal (2005) restrict their analysis to an annual trading frequency. With a higher trading frequency, as we allow, our approximation might not work as well. To address this concern, we extend DeMiguel and Uppal to examine whether our approximation is reasonably accurate in terms of an investor's utility loss from the approximation.

Because always keeping a single cost basis (by liquidating the entire position before buying any additional shares) is a feasible strategy in both the exact-basis and exact-holding-time model and the average-basis and average-holding-time model, an investor's utility in the single-basis model is bounded by the utilities in the two former models. In addition, when the upper bound $\bar{h}$ of the average holding period is equal to the long-term threshold L, the utility in the average-basis and average-holding-time model is smaller than that in the exact-basis and exact-holding-time model.[28] Because computationally it is much more tractable to compare the single-basis model with the exact-basis and exact-holding-time model, in the following we show that the utility loss from the single-basis model compared with the exact-basis and exact-holding-time model is small, which implies that the utility loss from the average-basis and average-holding-time model compared with the exact-basis and exact-holding-time model is even smaller.

A.5.1. The exact-basis model versus the single-basis model. First, let us give a brief introduction to the discrete-time exact-basis model DeMiguel and Uppal (2005) proposed.

If m is the number of discrete time points, we have time points $t_i = i\Delta t,\ i = 0, 1, \ldots, m$ with $\Delta t = T/m$. Let C_i, c_i, S_i denote cash in bank, consumption, and stock price, respectively, at time t_i. At each time point, the investor needs to determine stock holdings in addition to consumption. To keep track of the exact basis, we introduce the variable $N_{j,i}$ to represent the number of shares bought at time t_j and held at time t_i, where $j = 0, 1, \ldots, m$ and $i = j, \ldots, m$. When short sales are not optimal, such as when the risk premium is positive, we have

$$N_{j,i} \geq N_{j,i+1} \geq \ldots \geq N_{j,m} \geq 0 \tag{A-11}$$

for any $j \leq i$ and $i \in \{0, 1, \ldots, m\}$.

Given the trading and consumption strategy $\{N_{j,i}, c_i\}_{i=0,\ldots,m;\, j=0,\ldots,i}$, the cash amount at time t_i becomes

$$C_i = C_{i-1}e^{r\Delta t} - c_i - N_{i,i}S_i + \sum_{j=0}^{i-1}\left(N_{j,i-1} - N_{j,i}\right)\left[S_i - \left(S_i - S_j\right)\tau\left(t_i - t_j\right)\right] \tag{A-12}$$

for any $i \geq 1$, where $\tau(\cdot)$ is the tax rate as given in Section 2. We aim to choose $\{N_{j,i}, c_i\}$ to maximize

$$E\left[\sum_{i=1}^{m} e^{-\beta t_i}\frac{c_i^{1-\gamma}}{1-\gamma}\right] \tag{A-13}$$

subject to (A-11), (A-12), and the solvency constraint

$$C_i + \sum_{j=0}^{i} N_{j,i}\left[S_i - \left(S_i - S_j\right)\tau\left(t_i - t_j\right)\right] \geq 0 \ \text{ for any } i.$$

For further discretization, as in DeMiguel and Uppal (2005), we assume that the stock price S follows a binomial tree process. Then, the problem can be formulated as a constrained optimization

28 Intuitively, in an average-basis model, an investor is effectively forced to liquidate shares proportionally across all tax bases in the current position while it is better to first liquidate the shares with the greatest basis. Therefore, adopting an average-basis rule makes an investor worse off. In contrast, an average-holding-time rule may make an investor better off because some short-term gains can potentially be realized at the long-term rate. However, as the upper bound $\bar{h}$ of the average holding period decreases, this advantage of the average-holding-time rule decreases and can become a disadvantage. For example, if $\bar{h} = 1$, then unless all shares have long-term status, any gain will be taxed at the short-term rate, and thus the investor is worse off with the average-holding-time rule. Numerical results reveal that the utility differences among the cases with $\bar{h} = 1, 1.2, 1.5, 2, 10$ are negligible. Therefore, with the parameter values we used in our analysis, the utilities in our model are also smaller than those in the exact-basis and exact-holding-time model.

Table A.1
CEWL of the suboptimal single-basis strategy relative to the optimal exact-basis strategy

Parameter values	$\Delta t = 1/2, T = 4$ Single-basis CEWL	$\Delta t = 1/3, T = 3$ Single-basis CEWL
Default	0.00017%	0.03300%
$\mu = 0.04$	0.00000%	0.00960%
$\mu = 0.06$	0.00079%	0.39000%
$\sigma = 0.15$	0.00130%	0.55000%
$\sigma = 0.25$	0.00000%	0.00920%
$\tau_S = 0.30$	0.00059%	0.02600%
$\tau_S = 0.25$	0.00100%	0.02000%
$\tau_L = 0.20$	0.00060%	0.02700%
$\tau_L = 0.25$	0.00110%	0.02100%

Default parameters: $\mu = 0.05$, $\sigma = 0.20$, $r = 0.01$, $\tau_S = 0.35$, $\tau_L = 0.15$, $\gamma = 3$, and $\beta = 0.01$.

problem that can be solved by the Sparse Nonlinear Optimizer algorithm (SNOPT; see DeMiguel and Uppal 2005). It should be emphasized that the problem is strongly path-dependent and leads to 2^{i+1} states at time t_i. As a consequence, it is time consuming and can work only for a small number of time steps.

The single-basis model assumes that all stock holdings should be sold before purchase. Hence we need to add the following constraints

$$N_{i,i} \sum_{j=0}^{i-1} N_{j,i} = 0, \quad \forall i.$$

However, the solution to the optimization problem with the additional nonlinear constraints is not accurate enough to examine the wealth loss of the single-basis model from the exact-basis model. Hence, we use the following suboptimal single-basis strategy that clearly cannot be better than the optimal single-basis strategy: (i) if the second cost basis occurs in the optimal strategy with exact basis, we first sell all stock holdings to realize capital gains or losses and then buy back some stock to reach the same stock holding as in the strategy with exact basis; and (ii) the consumption remains unchanged. Using a suboptimal single-basis strategy biases against us in finding the small difference between the two models.

Table A.1 reports the CEWL from following the suboptimal single-basis strategy in terms of the percentage of the initial wealth. It shows that the loss relative to the optimal exact-basis strategy is small and seems almost negligible, even for the suboptimal single-basis strategy. Because a single-basis strategy is a feasible but suboptimal strategy in the average-basis and average-holding-time model, this seems to suggest that the average-basis and average-holding-time model is a reasonably good approximation.

References

Ben Tahar, I., H. M. Soner, and N. Touzi. 2010. Merton problem with taxes: Characterization, computation, and approximation. *SIAM Journal of Financial Mathematics* 1:366–95.

Cadenillas, A., and S. R. Pliska. 1999. Optimal trading of a security when there are taxes and transaction costs. *Finance and Stochastics* 3:137–65.

Carr, P. 1998, Randomization and the American put, *Review of Financial Studies* 11:597–626.

Constantinides, G. M. 1983. Capital market equilibrium with personal taxes. *Econometrica* 51:611–36.

———. 1984. Optimal stock trading with personal taxes: Implication for prices and the abnormal January returns. *Journal of Financial Economics* 13:65–89.

Cuoco, D., and H. Liu. 2000. Optimal consumption of a divisible durable good. *Journal of Economic Dynamics and Control* 24:561–613.

Dammon, R. M., and C. S. Spatt. 1996. The optimal trading and pricing of securities with asymmetric capital gains taxes and transaction costs. *Review of Financial Studies* 9:921–52.

Dammon, R. M., C. S. Spatt, and H. H. Zhang. 2001. Optimal consumption and investment with capital gains taxes. *Review of Financial Studies* 14:583–616.

———. 2004. Optimal asset location and allocation with taxable and tax-deferred investing. *Journal of Finance* 59:999–1037.

Davis, M. H. A., and A. R. Norman. 1990. Portfolio selection with transaction costs. *Mathematics of Operations Research* 15:676–713.

DeMiguel, V., and R. Uppal. 2005. Portfolio investment with the exact tax basis via nonlinear programming. *Management Science* 51:277–90.

Domar, E. D., and R. A. Musgrave. 1944. Proportional income taxation and risk-taking. *Quarterly Journal of Economics* 58:388–422.

Dybvig, P. H., and H. Koo. 1996. Investment with taxes. Working Paper, Washington University in St. Louis.

Ehling, P., M. Gallmeyer, S. Srivastava, S. Tompaidis, and C. Yang. 2010. Portfolio choice with capital gain taxation and the limited use of losses. Working Paper, University of Texas at Austin.

Fischer, M., and M. Gallmeyer. 2012. Heuristic portfolio trading rules with capital gain taxes. Working Paper, University of Virginia.

Friedman, A. 1982. *Variational principles and free-boundary problems*. New York: Wiley.

Gallmeyer, M., R. Kaniel, and S. Tompaidis. 2006. Tax management strategies with multiple risky assets. *Journal of Financial Economics* 80:243–91.

Liu, H., and M. Loewenstein. 2002. Optimal portfolio selection with transaction costs and finite horizons. *Review of Financial Studies* 15:805–35.

Marekwica, M. 2012. Optimal tax-timing and asset allocation when tax rebates on capital losses are limited. *Journal of Banking & Finance* 36:2048–63.

Merton, R. C. 1969. Lifetime portfolio selection under uncertainty: The continuous-time model. *Review of Economic Statistics* 51:247–57.

Poterba, J. M., and J. B. Shoven. 2002. Exchange-traded funds: A new investment option for taxable investors. *American Economic Review* 92:422–27.

Wilson, J., and P. Liddell. 2010. Sales of capital assets reported on individual tax returns, 2007. *Statistics of Income Bulletin*, Winter, 77–104.