ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Financial Economics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

JOURNAL OF Financial ECONOMICS

www.elsevier.com/locate/jfec

Tax management strategies with multiple risky assets ☆

Michael F. Gallmeyer[a,*], Ron Kaniel[b], Stathis Tompaidis[c]

[a]Mays Business School, Texas A&M University, College Station, TX 77843, USA
[b]Fuqua School of Business, Duke University, Durham, NC 27708, USA
[c]McCombs School of Business, University of Texas at Austin, Austin, TX 78712, USA

Received 13 March 2002; received in revised form 27 July 2004; accepted 2 August 2004

Abstract

We study the consumption-portfolio problem in a setting with capital gain taxes and multiple risky stocks to understand how short selling influences portfolio choice with a shorting-the-box restriction. Our analysis uncovers a novel *trading flexibility strategy* whereby, to minimize future tax-induced trading costs, the investor optimally shorts one of the stocks (or equivalently, buys put options) even when no stock has an embedded gain. Alternatively, an imperfect form of *shorting the box* can reduce aggregate equity exposure ex post. Given these two short selling strategies, it is common for an unconstrained investor to short some

☆We would like to thank Lorenzo Garlappi, Bruce Grundy, Campbell Harvey, Burton Hollifield, J. Spencer Martin, an anonymous referee, and seminar participants at the University of Florida, the University of South Carolina, the University of Texas at Austin, the Instituto Tecnológico Autónomo de México, the 2002 Bachelier World Congress meeting, the 2002 SIRIF Strategic Asset Allocation meeting, and the 2002 European Finance Association meeting for comments; Eric Bradley for conversations concerning the U.S. tax code; Patrick Jaillet for assisting us with obtaining computing resources; and Sergey Kolos for programming assistance. We would also would like to thank the Texas Advanced Computing Center for providing computing resources. The usual disclaimer applies.

*Corresponding author. Fax: +1 412 268 6837.
E-mail address: mikegallmeyer@gmail.com (M.F. Gallmeyer).

doi:10.1016/j.jfineco.2004.08.010

equity while a constrained investor holds a positive investment in all stocks. With no shorting, the benefit of trading separately in multiple stocks is not economically significant.

JEL classification: G11; H20

Keywords: Portfolio choice; Capital gain taxation; Short selling

1. Introduction

When investors are faced with asset allocation and consumption decisions, capital gain taxation plays an important role in the investor's optimal strategy. In his seminal work valuing the tax loss selling option from capital gain taxation, Constantinides (1983) shows that an investor's portfolio choice problem is integrally linked to realized capital gain taxation. With the only friction being taxation, the investor optimally defers all gains and immediately realizes all losses without influencing his optimal consumption strategy. This separation result is achieved by the investor rebalancing his portfolio without triggering a tax liability by engaging in a *shorting-the-box strategy*: if an investor is overexposed to a stock with a large embedded capital gain, he shorts that security instead of selling it so that his net position in the stock is optimal. By shorting, the investor has rebalanced and deferred realizing any taxable capital gains since none of the original position was sold. Before the 1997 Tax Reform Act, shorting the box was not viewed as a tax triggering transaction. Besides the collateral costs of shorting, the investor could effectively shield all gains from capital gain taxation over his lifetime given the U.S. tax code provision of resetting the tax basis of all securities to market prices at the time of death.

However, given that the shorting-the-box strategy for identical securities is no longer permitted under U.S. tax laws and that short selling is costly, investors do realize gains. For recent empirical evidence, see the references in Poterba (2001) as well as Auerbach and Siegel (2000). The work of Dammon et al. (2001b) uses this evidence as one motivation for studying capital gain tax portfolio-consumption problems where the separation result fails. They study a short-sale-constrained investor's consumption-portfolio problem with a single stock. Since the investor cannot trade without tax liabilities, the optimal policy is influenced by the current portfolio composition. They find results similar to portfolio problems with transaction costs—the optimal mix between the stock and a riskless money market account can deviate from the optimal policy with no capital gain tax due to the tax-induced costs of trading. A limitation of their work is their assumption of one risky stock. As a result, they are unable to analyze how the composition of a portfolio with multiple risky stocks is affected by realized capital gain taxation. With the introduction of taxes, risk-based motives for portfolio rebalancing now interact with motives for reducing realized capital gains.

In this paper, we study the role of realized capital gain taxation on an investor's consumption-portfolio problem with two risky stocks and a riskless money market account where costly shorting is allowed under a no shorting-the-box constraint. Our two-stock setting allows us to qualitatively analyze the tradeoff between diversification and minimizing tax liabilities, where we consider both diversification between the riskless money market account and stocks and diversification within the equity portion of the portfolio. The choice of two stocks is for tractability. However, the main features of our results should extend to portfolio choice with additional stocks. The setting we have in mind is one where an investor considers moving from investing in a single index fund and a money market account to a portfolio of two index funds and a money market account. Current investment vehicles make this transition particularly easy with the introduction of several exchange-traded funds (ETFs) in recent years like the SPDR and the DIAMONDS contract.[1]

As a baseline in our analysis, we start by studying optimal portfolio choice with a short sale constraint. Our results show that for stocks that are not highly correlated ($\rho = 0.4$), the asset allocation in one stock is largely unaffected by the embedded capital gain in the other stock. As in the single-stock case, the basis reset provision at the investor's death leads to holding more equity as the investor ages. However, if embedded gains are large enough, the investor holds an undiversified equity portfolio. When the stock return correlation rises to a level commonly observed between U.S. large-capitalization ETFs ($\rho = 0.8$ and above), the optimal portfolio policy is different since tax considerations now outweigh diversification costs. The portfolio allocation for one stock is not just driven by its own basis and position, but by the basis and position of the other stock. If initially overinvested in equity, the investor sells the stock with the lowest tax cost. If short sale constrained, the investor might entirely liquidate his position in one stock. This behavior leads to the investor holding a less diversified equity position than before.

Allowing the investor to short sell while still imposing a shorting-the-box constraint dramatically changes behavior. When the cost of shorting is not too large and the return correlation between the two stocks is high enough, the investor employs two tax management trading strategies that utilize short selling. The first strategy, new in our analysis, is an ex ante way of minimizing future tax-induced trading costs by shorting one of the stocks even when the stock portfolio contains no embedded capital gains. This trade, termed the *trading flexibility strategy*, is used when the benefit of holding a well-diversified stock portfolio is outweighed by the expected future rebalancing costs of such a position. From our parameterizations, the trading flexibility strategy is employed when the return correlation between the two stocks is greater than or equal to $\rho = 0.65$. The second strategy, present when the correlation between the two stocks is as low as $\rho = 0.4$, is an *imperfect form of shorting the box* used to reduce ex post the investor's total equity exposure by shorting the stock with the largest tax basis. Given that the two stocks are not

[1]Interestingly, many ETFs pass on lower taxable unrealized capital gains to investors than mutual funds. This is due to active creations and destructions of ETFs (Poterba and Shoven, 2002). Also, ETFs are marginable and most can be shorted without being subject to the uptick rule.

perfectly correlated, such a trade entails fundamental risk and is permitted under current U.S. tax laws. From these two incentives to short, the optimal equity portfolio is significantly different from both the no-tax well-diversified allocation and the allocation when short selling is disallowed. In particular, a short-selling investor is better able to manage his total equity exposure over his lifetime as compared to a short-sale-constrained investor.

The ability to short sell introduces another interesting feature to the trading strategy relative to the case of no short sales. Given that the net tax benefit of selling shares is not monotonic in the trading strategy when shorting is allowed, we show that it is common for an unconstrained investor to short equity while an otherwise identical constrained investor holds strictly positive positions in all stocks. This feature is especially common when the portfolio contains no embedded capital gains, but occurs even when the portfolio's stock positions contain capital gains.

Whether the portfolio strategies identified in our analysis should be used as normative advice to investors depends on two important factors. The first factor is the magnitude of the welfare improvement that can be obtained by following such strategies. Somewhat surprisingly, we find that when short sales are prohibited, the welfare benefit of using the optimal strategy is negligible relative to the case in which the investor invests in a single index fund and a money market account. On the other hand, with short selling the benefit can be significant. The second factor is the investor's type. We show that use of the trading flexibility and the imperfect shorting-the-box strategies can be beneficial to wealthy investors, but not to the same degree for small investors who pay higher shorting costs.

As an alternative to shorting, we also consider how derivative securities can be used by an investor to manage tax trading costs when rebalancing. Constantinides and Scholes (1980) discuss a similar trading strategy without exploring its feasibility. The introduction of derivatives in the opportunity set restores a small investor's flexibility to defer capital gains while keeping the equity exposure of the optimal portfolio closer to the no-tax benchmark. In particular, we consider a strategy in which the investor, in addition to trading a single risky stock, is able to trade a put option written on a highly correlated stock. This strategy is available to both small and large investors and is an implicit use of the trading flexibility strategy. We show that the welfare benefit of using puts is similar in magnitude to that of using low-cost short selling.

Two recent papers developed at the same time as this work, Dammon et al. (2001a) and Garlappi et al. (2001), have also numerically analyzed some aspects of the capital gain tax investment problem with multiple stocks. The focus of each of these papers is quite different in that neither studies the role of shorting in portfolio choice or the welfare benefits of investing in two stocks relative to one. The numerical analysis in Dammon et al. (2001a) focuses on demonstrating that the diversification benefit of reducing the exposure to a highly volatile concentrated position can significantly outweigh the tax cost of selling. The paper of Garlappi et al. (2001) mostly analyzes features of the ''no trade region'' in the presence of capital gain taxes when an investor maximizes the expected utility of terminal wealth over ten periods with tax forgiveness at the terminal date.

Our work is also related to several earlier papers characterizing portfolio choice with capital gain taxes or transaction costs. Building from Constantinides (1983), the pricing implications of optimal after-tax portfolios with shorting-the-box trades is studied in Constantinides (1984). Our analysis considers the case in which shorting-the-box trades are prohibited. Dybvig and Koo (1996) is one of the earliest numerical studies of after-tax portfolio choice in a single stock and bond setting with no short sales. Due to computational difficulties, they only study the portfolio problem for a limited number of time periods, in contrast to the lifetime portfolio problem considered here. Using the single stock and bond framework of Dammon et al. (2001b), Huang (2001) and Dammon et al. (2004) study the asset location decision when taxable and tax-deferred accounts are available to investors. By retaining the single-stock assumption, these studies face a less complex numerical characterization than the multiple-stock case. Other early works on after-tax portfolio choice but in a single-period setting are Elton and Gruber (1978) and Balcer and Judd (1987). For exact solutions to capital gain tax portfolio problems under restrictive conditions, see Cadenillas and Pliska (1999) and Jouini et al. (2000). Using results from the literature on portfolio problems with transaction costs,[2] Leland (2001) numerically characterizes a portfolio allocation problem for a stock and a bond with capital gain taxes when the objective is to minimize the deviation from exogenous portfolio weights subject to capital gain taxes and transaction costs. Finally, the numerical study of our capital gain tax problem is related to numerical characterizations of portfolio problems with transaction costs (Balduzzi and Lynch, 1999, 2000).

The remainder of the paper is organized as follows. Section 2 formulates the consumption-portfolio problem. Sections 3 and 4 present our numerical analysis where we characterize the trading strategies and provide comparative statics as well as a welfare analysis. An alternative trading strategy using derivative securities is discussed in Section 5. Section 6 concludes. Appendix A gives a formal mathematical definition of the problem studied in Sections 3 and 4. Appendix B modifies the portfolio problem to incorporate investment in a put option as discussed in Section 5. Appendix C discusses the numerical procedure.

2. The consumption-portfolio problem with taxes

We consider a discrete-time economy with trading dates $t = 0, \ldots, T$ in which an investor chooses an optimal consumption and investment policy in the presence of realized capital gain taxation. Our framework is an extension of the single risky asset model of Dammon et al. (2001b) modified to incorporate multiple risky assets and short sales with margin requirements and shorting collateral costs. These modifications greatly expand the opportunity set of the investor as compared to a setting with no short sales as will be described in Sections 3 and 4, where we provide

[2] See, for example, Constantinides (1986), Davis and Norman (1990), Davis et al. (1993), Shreve and Soner (1994), and Akian et al. (1996).

a numerical characterization of the investor's consumption-portfolio problem both with and without a short-sale constraint. Our assumptions concerning security prices, taxation, and the investor's portfolio problem are presented below. A mathematical description of our model is provided in Appendix A.

2.1. Security market

An investor derives utility from consuming a perishable good financed through yearly trade in the security market. This market consists of three assets: a riskless money market account and two risky dividend-paying stocks. The riskless money market pays a continuously compounded pre-tax interest rate r. The two risky stocks pay pre-tax dividends with constant dividend yields. The evolutions of the ex-dividend stock prices are described by binomial Markov chains, where the correlation between the two stocks is equal to a constant, ρ. When presenting our results, we will often refer to a benchmark economy with a single risky stock. This is an economy where the two stocks can only be traded as an index constructed from an equally weighted portfolio.

2.2. Taxation

Dividend and interest income are taxed as ordinary income on the date that they are paid at the constant rate τ_D. Realized capital gains and losses are subject to a constant capital gain tax rate of τ_C where we assume that the full proceeds of capital losses can be used. When an investor reduces his outstanding stock position by either selling his long position or buying back his short position, he incurs realized capital gains or losses subject to taxation. The tax basis used for computing these realized capital gains or losses is calculated as a weighted-average purchase price.[3] At the time of an investor's death, capital gain taxes are forgiven and the tax bases of the two stocks are reset to their current market prices. This is consistent with the reset provision in the U.S. tax code.[4] Dividend and interest taxes are still paid at the time of death.

While we allow an investor to wash sell to immediately realize capital losses, we do not allow him to "short the box" by taking an offsetting position in the same security to circumvent paying capital gain taxes on realized gains. Shorting the box involves

[3]The U.S. tax code allows investors a choice between the weighted-average price rule and the exact identification of the shares to be sold. While choosing to sell the shares with the smallest embedded gains using the exact identification rule is beneficial to the investor, solving for the optimal investment and liquidation strategy becomes numerically intractable for a large number of trading periods (Dybvig and Koo, 1996; Hur, 2001; DeMiguel and Uppal, 2003). Furthermore, for parameterizations similar to those in this paper, DeMiguel and Uppal numerically show that the certainty-equivalent wealth loss using the weighted-average price basis rule as compared to the exact identification rule is small. For a portfolio horizon of ten years, they find the certainty-equivalent wealth loss to be less than 0.5%.

[4]For long positions, the U.S. tax code is explicit that the basis is reset at death. For short positions, the basis is also reset to the stock price at the time of death (see the IRS Revenue Ruling 73-524 and the IRS Private Letter Rulings 9436017 and 9319005).

realizing a gain without tax consequences. Suppose an investor is currently long equity with a large embedded gain. Instead of selling this position, the investor could take an offsetting short position in the same security. He has now effectively sold his long position with no tax consequences. The 1997 Taxpayer Relief Act reclassified such a trade as a sale of the original position and thus subject to capital gain treatment.[5]

To accommodate short sales with a shorting-the-box restriction, the evolution of the tax basis for each stock includes a variety of cases. At each trading date, the tax basis for each stock position either evolves as a share-weighted average of the current stock price and the previous basis when increasing the absolute size of a stock's position, resets to the current stock price, or remains unchanged. The basis resets to the current stock price under two different scenarios: when an investor incurs a capital loss on his position, or when the investor's position changes sign from time $t-1$ to time t. If a position in the investor's portfolio incurs a capital loss, it is optimally liquidated to realize the loss given that wash sales are allowed. We assume that the full amount of this loss can be used immediately.[6] A transaction where the investor's position changes sign from time $t-1$ to t is treated as a closing of the $t-1$ position, since shorting the box is prohibited. Any gains or losses on this position are taxed at the capital gain rate. A stock's tax basis remains unchanged either when the investor does not trade in the stock or when the investor reduces but does not fully liquidate the absolute size of his stock holdings.

2.3. Investor problem

In order to finance consumption, an investor dynamically trades in the two risky stocks and a riskless money market account. Short sales of equity are allowed subject to collateral and margin requirements. The collateral and margin constraints lead to a constraint on the minimum amount invested in the money market account. Investors must also pay lending fees when shorting stocks. These fees are incorporated by reducing the rate of return received on the short-sale collateral as compared to money invested in the money market account. While small investors typically receive no interest on short-sale proceeds, large investors face much smaller fees. The size of these fees is discussed later when specific parameter values are presented.

Given an initial equity endowment, a consumption and security trading policy is an *admissible trading strategy* if it satisfies the collateral and margin requirements, is subject to lending fees, is self-financing, and leads to nonnegative wealth over the

[5]Strictly speaking, the 1997 Taxpayer Relief Act did not completely rule out shorting the box for deferring gains, but it seriously limited its effectiveness. Under the Act, shorting the box is still allowed to defer gains for one year but you must close your short position within 30 days after the end of the year, and then you must stay long in the stock unhedged for 60 days before closing your long position. To simplify our analysis, we assume that shorting the box is prohibited.

[6]Under the current U.S. tax code, realized losses can only offset up to $\$3,000$ of ordinary income, but can be carried forward indefinitely. Relaxing our full loss usage assumption would add one state variable to the formulation, significantly increasing the complexity of the problem. For an analysis of the role of capital losses with no after-tax arbitrage, see Gallmeyer and Srivastava (2003).

lifetime of the investor. The investor is assumed to live at most T periods and faces a positive probability of death each period. The probability that an investor lives up to period $t < T$ is given by a survival function, calibrated to the 1990 U.S. Life Table, compiled by the National Center for Health Statistics where we assume period $t = 0$ corresponds to age 20 and period $T = 80$ corresponds to age 100. At period $T = 80$, the investor exits the economy with certainty.

The investor's objective is to maximize his discounted expected utility of real lifetime consumption and a time of death bequest motive by choosing an admissible consumption-trading strategy given an initial endowment. For tractability and ease of comparison with no tax portfolio problems, the utility function for consumption and wealth is of the constant relative risk aversion form with a coefficient of relative risk aversion of γ. Using the principle of dynamic programming, the Bellman equation for the investor's optimization problem, derived in Appendix A, can be solved numerically by backward induction starting at time T. Details of the computational complexity of this problem are outlined in Appendix C.

2.4. Scenarios considered with parameter values

To understand how an expanded opportunity set due to shorting can influence the allocation decision and welfare of an investor, we focus on several cases where an investor has different investment opportunities and faces different tax trading costs. While a variety of different investor scenarios could be studied in the context of portfolio choice with multiple risky assets, we focus on an index investor who considers moving from investing in a single index fund and a money market account to a portfolio of two funds that compose the index and the money market account. Our benchmark is the case when the investor trades a money market and a single risky index fund with realized capital gain taxation. (When the investor is not subject to capital gain taxation, a mutual fund theorem results irrespective of the number of risky assets; he only trades in an appropriately weighted index fund of the two stocks and the money market.) We compare this benchmark to an investor who has access to two identical risky stocks that are subject to capital gain taxes. We consider investors who are restricted from shorting equity, as well as investors who can short subject to margin constraints and collateral costs.

Given that the main emphasis of our work is to understand the quantitative features of portfolio choice with taxes and short sales, especially the use of the flexibility and the imperfect shorting-the-box strategies, our index fund setting is chosen given the large number of exchange-traded funds (ETFs) that are now available for investing in broad-based market indices. Currently, roughly 40 different ETFs trade on the American Stock Exchange that are pegged to marketwide indices. All of these ETFs are marginable and can be shorted, while only a handful of them are subject to the "uptick rule." (Rule 10a-1 of the Securities Exchange Act of 1934, more commonly known as the "uptick" rule, precludes short selling when security prices are falling.) Additionally, the market for shorting ETFs is very active. For example, the NASDAQ 100 tracking stock, QQQ, had an average open interest of 27% of shares outstanding and an average days to cover of 2.60 over the year 2002.

Our base case choice of parameters, roughly comparable to the one used in Dammon et al. (2001b), considers an index fund with price dynamics consistent with large-capitalization U.S. stock indices given by an expected return due to capital gains of $\mu = 7\%$, a dividend yield $\delta = 2\%$, and a volatility $\sigma = 20\%$. To facilitate easy interpretation of the optimal portfolio choice, this index fund is composed of two ETFs in equal proportions with identical expected returns, dividend growths, and volatilities. For convenience throughout the rest of the paper, we will always refer to the ETF investments as stock investments. We allow the return correlation ρ between the two stocks to vary and report results for correlations $\rho = 0.4$, 0.8, and 0.9. In order to keep the pre-tax Sharpe ratio of the equally weighted risky stock portfolio fixed as the correlation varies, we set the volatility of each individual stock to $\sigma_i = \sigma/\sqrt{0.5(1+\rho)}$. For all parameterizations, the money market's return r_f equals 6%. The investor rebalances his portfolio once a year. The investor enters the economy at age 20 and exits no later than age 100.

When studying only the role of short selling on portfolio choice, we mainly focus on a setting where the correlation between the two risky stocks is at least $\rho = 80\%$. From our investment setting of a portfolio of ETFs written against broad market indices, such correlations are consistent with those seen among large market indices. Over the period 1962–2001, the correlation of returns between the S&P 500 and the value-weighted CRSP index, the Dow Jones Industrial Average, and the equal-weighted CRSP index was 99%, 95%, and 87% respectively. Over the period 1973–2001, the correlation of returns between the NASDAQ 100 index and the S&P 500 index, the value-weighted CRSP index, and the equal-weighted CRSP index was 80%, 73%, and 74%, respectively. Currently, several ETFs exist that are either directly pegged to these indices or are pegged to other large-capitalization indices.[7] Additionally, some ETFs exist that track a component of an index. For example, the iShares Russell 1000 and iShares Russell 2000 track portions of the iShares Russell 3000, while the iShares Russell 2000 Growth and iShares Russell 2000 Value track portions of the iShares Russell 2000. The correlations between these components of the two Russell ETFs are also high. Over the period 1995–2001, the return correlation between the Russell 1000 and Russell 2000 was 81% and between the Russell 2000 Growth and Russell 2000 Value was 87%.

The tax rates in the numerical examples are set to roughly match those faced by a wealthy investor. We assume that dividends and interest are taxed at the investor's marginal income rate $\tau_D = 36\%$. The capital gain tax rate is set to the long-term rate $\tau_C = 20\%$. The investor begins investing at age 20 and can live to a maximum of 100 years where the single-period hazard rates λ_n are calibrated to the 1990 U.S. Life Table compiled by the National Center for Health Statistics. Hence, the maximum horizon for an investor is $T = 80$. The inflation rate is assumed to be $i = 3.5\%$. The investor's constant relative risk averse preferences are calibrated with a time discount parameter $\beta = 0.96$ and relative risk aversion $\gamma = 3$. The bequest motive is calibrated

[7]Example ETFs include the SPDR, the DIAMOND, the Fortune 500 Index Tracking Stock, the Rydex S&P Equal Weight ETF, the Vanguard Total Stock Market VIPER, the Vanguard Extended Stock Market VIPER, the iShares S&P 500 Fund, and the streetTRACKS Dow Jones Global Titans 50 Index.

such that the investor plans to provide a perpetual real income stream to his heirs. This parameterization is consistent with the one used in Dammon et al. (2001b).

To short stock, U.S. investors must trade in a margin account and are required to deposit and maintain a minimum amount of cash or securities with their broker. The Federal Reserve Board's Regulation T sets the initial margin requirement for stock positions undertaken through brokers. The initial margin requirement is currently 50% for a long equity position and 150% for a short equity position. For a long position, the investor cannot borrow more than 50% of the market value of the stock. For a short position, 102% of the short sale proceeds must typically be held in cash as noted by Geczy et al. (2002) and Duffie et al. (2002). The remaining 48% needed to cover the margin requirement can be held in other securities such as U.S. Treasury Bills.[8] Small retail investors do not typically receive any interest on the cash collateral although large investors do. From data in Geczy et al., the rate of interest received on collateral, or the *general collateral rate*, is on average eight basis points below the federal funds effective rate, while for medium-size loans the general collateral rate is on average 15 basis points below the federal funds rate.

For our analysis, we consider conservative estimates for these lending rates where we assume that a large investor receives interest on his collateral at a rate of 30 basis points below the riskless money market rate. We frequently refer to the lending rate in terms of a *shorting cost* defined as the difference between the riskless money market rate and the general collateral rate. For tractability, we make no distinction between initial and maintenance margin collateral and assume that when rebalancing, the investor's portfolio must conform with Regulation T initial margin requirements.

3. Structure of optimal portfolios

We begin our numerical analysis by studying the structure of optimal portfolios. Specifically, our goal is to numerically characterize how and when investor behavior changes by expanding the investment opportunity set to include costly short selling where the flexibility and imperfect shorting-the-box strategies can be employed.

3.1. Single stock benchmark

To facilitate comparison with the two-stock setting, we first briefly analyze portfolio choice when the investor's only risky asset is a single index fund. Fig. 1 outlines the characteristics of this case both with and without realized capital gain taxation. Using the parameterization outlined above when the index volatility is $\sigma = 20\%$, the optimal equity allocation for an investor who faces no capital gain taxation but pays interest and dividend taxes is summarized by the solid line with

[8]For margin requirement institutional details, see Fortune, 2000 as well as the Federal Reserve Board's Regulation T available at http://www.gpoaccess.gov/cfr/index.html by searching the Code of Federal Regulations for 12CFR220.1.

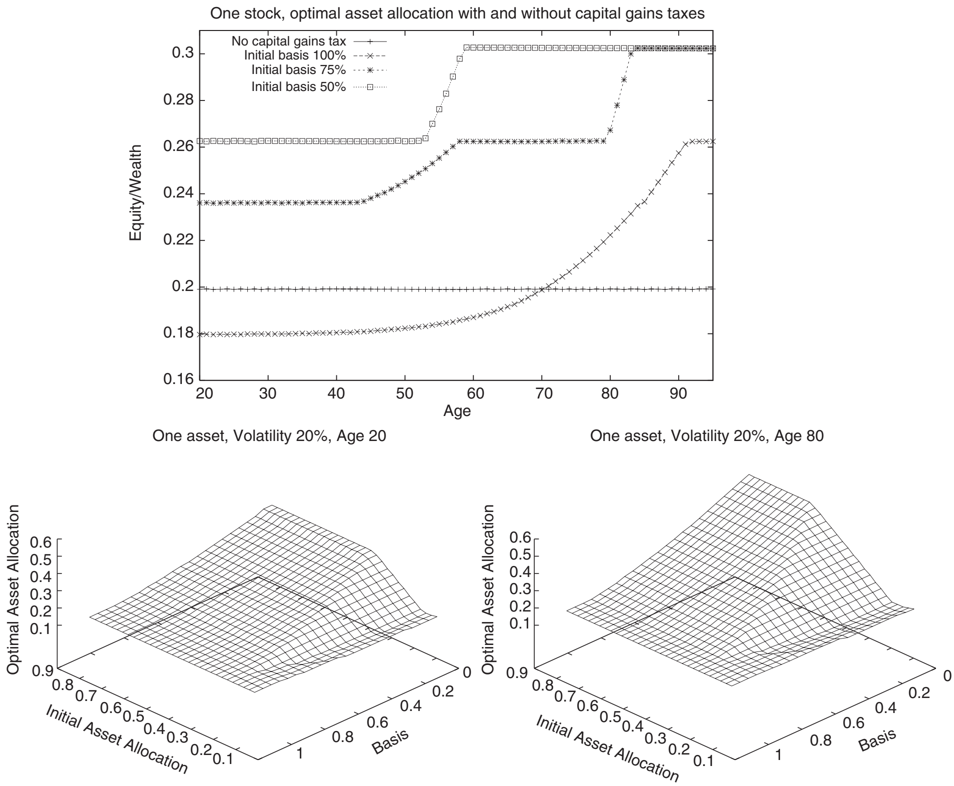


Fig. 1. *Benchmark portfolio choice with a single index fund.* The top panel of the figure summarizes the equity to wealth ratio as a function of age. The line labeled ''No capital gain tax'' presents the optimal equity allocation when the investor faces no capital gain tax ($\tau_D = 36\%$, $\tau_C = 0\%$, $\sigma = 20\%$). The other three lines present the optimal equity allocation when the investor enters age t with 30% of his wealth invested in equity and faces realized capital gain taxation ($\tau_C = 20\%$). The three lines plot different basis-price ratios (50%, 75%, and 100%) entering the trading period. The bottom panels of the figure summarize portfolio choice when the investor faces capital gain taxation on one risky stock. The left (right) panel presents the equity-to-wealth ratio as a function of the stock allocation and the basis-price ratio entering the trading period at age 20 (80). Parameters used for the bottom panels: $\tau_D = 36\%$, $\tau_C = 20\%$, $\sigma = 20\%$.

cross marks in the top panel of Fig. 1. Under this benchmark, the equity-to-wealth ratio is constant at 20% since the opportunity set with no capital gain tax is constant through time.

The bottom panels of Fig. 1 document optimal portfolio choice with realized capital gain taxation at ages 20 and 80. This is the setting studied by Dammon et al. (2001b). In this case, the investor's optimal equity exposure is a function of the beginning-period allocation and the basis-price ratio. When the marginal tax costs of trading are high due to a large embedded capital gain (a low basis-price ratio), the investor optimally holds more equity. This behavior occurs at a smaller embedded gain as the investor ages, since it is driven by the basis reset provision at death. The top panel of Fig. 1 demonstrates this age effect of embedded capital gains on equity choice conditional on three different basis-price ratios and a stock allocation entering age t of 30%. When the embedded gain in the stock portfolio is high (a 50% basis-price ratio), an age 20 investor reduces his equity exposure to 26.5% of wealth while an age 80 investor fully retains his 30% equity position. As the embedded gain in the stock position falls, the investor optimally liquidates more stock but less when older. This is captured in the basis-price ratio cases of 75% and 100% plotted in Fig. 1. Relative to the setting with no capital gain tax, the investor can be significantly overexposed to equity when older with a large embedded gain.

While examining optimal portfolio choice at a particular time and state is useful in understanding the conditional asset trading behavior of an investor, it provides only limited information about portfolio composition over the investor's lifetime. To gain insights about the unconditional optimal portfolio choice, we perform Monte Carlo simulations starting with no embedded stock gains at age 20 to track the evolution of the investor's optimal portfolio at ages 40, 60, and 80 conditional on the investor's survival. These results are reported in the lines labeled "One Stock Benchmark" in Panels A through C of Table 1. The columns labeled "Max Equity Allocation" and "Max Equity Basis" present the mean and standard deviation of equity exposure and the basis-price ratio, respectively. The equity exposure is expressed as a fraction of total financial wealth. The column "Embedded Gains" measures the fraction of financial wealth that is an unrealized capital gain. All simulations are over 50,000 paths. The standard error for each mean estimate can be computed by dividing the Monte Carlo standard deviation by $\sqrt{50,000} = 223.6$. Given the largest standard deviation in the table is 0.28, the largest standard error for the mean estimate of any quantity in the table is 0.00125.

The simulation analysis provides insights into the magnitude of the investor's equity position as he ages relative to the no-tax benchmark. At age 40 in the "One Stock Benchmark" (Panel A), the allocation in equity increases on average to 24% from 18% at age 20, while the average basis-price ratio drops to 0.48 from 1.0 at age 20. The evolution of the optimal portfolio leads to an average embedded gain in the risky stock of 13% of the investor's wealth, indicating that the investor's portfolio has substantial embedded capital gains. As the investor grows older, his fraction of wealth invested in the stock and embedded gain continues to grow as can be seen in the age 60 and 80 simulations. By age 80, the investor holds on average 29% of his wealth in equity with an average embedded gain of 19% of his wealth due to his bequest motive and capital gain tax forgiveness at death.

3.2. Optimal portfolio composition with two stocks and no short sales

To facilitate disentangling the role of short selling from the role of additional stocks in optimal portfolio choice, we study the effect of introducing a second stock with no short sales. Intuitively, by being able to trade the components of the stock index individually, the investor should be able to rebalance his portfolio in a more tax-efficient manner as compared to only trading the entire index. However, such rebalancing is costly given that the investor still has an incentive to maintain a well-diversified portfolio for risk exposure purposes. To understand how these two incentives quantitatively determine portfolio composition, we consider two different scenarios. Under the first scenario, the investor is grossly overinvested in equity compared to the single-stock benchmark. Specifically, we consider an investor at age t who holds 70% of his wealth in equity with 40% in stock 1 and 30% in stock 2. This allows us to study the tradeoff between holding the optimal mix of equity and money market holdings and minimizing tax-induced trading costs. To capture the costs of holding an undiversified equity position, our second scenario assumes that the investor's investment in equity at age t is 20% of wealth, which is roughly equal to the optimal total equity exposure with no capital gain tax. However, the investor holds only one of the stocks at the start of age t, which makes him grossly undiversified but not overexposed to equity versus the money market account.

Starting when the investor is overexposed to equity relative to the money market account, the optimal strategies for a stock return correlation of 80% are presented in Fig. 2. The left (right) panel describes the optimal equity allocation at age 20 (80) as a function of the equity basis-price ratios. From the figure, the optimal trading strategy is sensitive to tax trading costs. The optimal portfolio choice in one stock is not independent of the investor's position in the other. For example, the optimal allocation for stock 2 is weakly increasing in the basis-price ratio of stock 1 for both young and old investors. Given that the two stocks are highly correlated, the investor sells the stock with the smallest embedded gain to reduce the total equity exposure. The smallest optimal position in stock 2 occurs when its basis is in the tax-loss selling region and the embedded gain of stock 1 is high. Here, the investor completely liquidates his position in the stock in order to reduce his total equity exposure as cheaply as possible. When the basis-price ratios are close to each other, the optimal allocation can change dramatically for small perturbations in the initial bases. For example, at age 20, stock 2's optimal allocation is 8.1% of wealth for a basis-price distribution of $b_1 = 0.6$ and $b_2 = 0.8$, while it changes to 17.3% of wealth when the initial basis-price ratios are reversed to $b_1 = 0.8$ and $b_2 = 0.6$. In unreported results, when the correlation between the two stocks is reduced to 40%, the optimal strategy mirrors the strategy of the investor who only has access to a single stock. The investor optimally sells more of the stock when its embedded gains are smaller. In this case, the existence of a second stock does not appear to significantly influence the investor's action in the other stock. Summarizing these results, as the correlation between the two stocks increases, the investor sells the stock with the smallest cost to trade to reduce his total equity exposure.

We now examine optimal portfolio choice under our second scenario when the investor enters age t with an equity position that is of the appropriate magnitude, but is grossly undiversified. Fig. 3 presents the optimal equity position in each stock as well as the aggregate equity position as a function of stock 2's basis-price ratio. The investor enters the period holding only an equity position of 20% of his wealth in stock 2. He holds no position in stock 1 when he enters the period. The top panel gives the optimal allocations for a return correlation of 40% while the bottom panel gives the optimal allocation for a return correlation of 80% for an age 20 investor. At a correlation of 40% and a low stock 2 basis-price ratio, the tax trading costs

Table 1
Optimal portfolio choice characteristics

This table presents simulation results for the mean and standard deviation of portfolio characteristics at ages 40, 60, and 80 over 50,000 paths. The simulations consider several scenarios: an investor who trades one risky stock (“One Stock Benchmark”), an investor who trades two risky stocks and is short sale constrained (“Two Stock No Short Sales”) at correlations of 0.4 and 0.8, and an investor who trades two risky stocks at a correlation of 0.8 and can sell short at a cost (0, 30, 50, 100, 600 b.p.). In all cases, interest and dividend income is taxed at $\tau_D = 0.36$ and realized capital gains are taxed at $\tau_C = 0.2$. The column labeled “Max Equity Allocation” (“Min Equity Allocation”) records the simulation characteristics of the largest (smallest) stock position. The stock positions and total equity (labeled “Total Equity”) are computed as fraction of wealth. The bases (“Max Equity Basis” and “Min Equity Basis”) are basis-price ratios when the stock position is positive. When a stock position is negative, the basis is a price-basis ratio. Embedded gains (labeled “Embedded Gains”) are calculated as the fraction of wealth that is a capital gain. Time shorting (labeled “Time Shorting”) is the fraction of simulation paths in which the investor is shorting at each age.

Simulation results	Max equity allocation		Min equity allocation		Max equity basis		Min equity basis		Total equity mean	Embedded gains		Time shorting mean
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		Mean	Std. Dev.	
Panel A: Age 40												
One stock benchmark	0.24	0.05			0.48	0.26			0.24	0.13	0.08	
Two stock no short sales – 40% correlation	0.14	0.04	0.09	0.02	0.41	0.23	0.63	0.26	0.23	0.12	0.06	
Two stock no short sales – 80% correlation	0.13	0.03	0.09	0.02	0.48	0.22	0.64	0.24	0.22	0.10	0.06	
Two stock short sales with shorting costs												
0 b.p. – 80% correlation	0.31	0.06	−0.15	0.03	0.45	0.25	0.83	0.14	0.17	0.22	0.10	100.0%
30 b.p. – 80% correlation	0.21	0.10	−0.03	0.10	0.45	0.23	0.76	0.22	0.18	0.15	0.09	55.3%
50 b.p. – 80% correlation	0.16	0.07	0.05	0.06	0.46	0.23	0.74	0.22	0.21	0.12	0.07	20.0%
100 b.p. – 80% correlation	0.14	0.05	0.07	0.04	0.48	0.22	0.71	0.23	0.22	0.11	0.06	7.7%
600 b.p. – 80% correlation	0.13	0.03	0.09	0.02	0.48	0.22	0.64	0.24	0.22	0.11	0.06	0.0%

Panel B: *Age* 60												
One stock benchmark	0.25	0.05			0.35	0.26			0.25	0.17	0.09	
Two stock no short sales – 40% correlation	0.15	0.04	0.10	0.03	0.34	0.23	0.51	0.26	0.25	0.16	0.07	
Two stock no short sales – 80% correlation	0.15	0.05	0.08	0.03	0.42	0.21	0.62	0.24	0.24	0.13	0.07	
Two stock short sales with shorting costs												
0 b.p. – 80% correlation	0.37	0.11	−0.20	0.06	0.43	0.23	0.77	0.15	0.17	0.29	0.11	100.0%
30 b.p. – 80% correlation	0.30	0.09	−0.13	0.07	0.42	0.24	0.79	0.18	0.18	0.22	0.11	95.0%
50 b.p. – 80% correlation	0.20	0.09	0.01	0.10	0.37	0.22	0.75	0.21	0.21	0.16	0.09	36.5%
100 b.p. – 80% correlation	0.18	0.08	0.04	0.07	0.40	0.22	0.73	0.21	0.22	0.14	0.08	29.3%
600 b.p. – 80% correlation	0.15	0.05	0.08	0.03	0.41	0.21	0.61	0.24	0.24	0.13	0.07	0.0%
Panel C: *Age* 80												
One stock benchmark	0.29	0.08			0.36	0.23			0.29	0.19	0.10	
Two stock no short sales – 40% correlation	0.19	0.06	0.11	0.03	0.34	0.21	0.49	0.25	0.30	0.19	0.08	
Two stock no short sales – 80% correlation	0.18	0.07	0.10	0.04	0.37	0.20	0.57	0.24	0.28	0.17	0.09	
Two stock short sales with shorting costs												
0 b.p. – 80% correlation	0.47	0.25	−0.28	0.16	0.43	0.26	0.72	0.16	0.19	0.40	0.26	100.0%
30 b.p. – 80% correlation	0.41	0.23	−0.20	0.12	0.43	0.25	0.72	0.16	0.21	0.34	0.17	99.8%
50 b.p. – 80% correlation	0.36	0.12	−0.13	0.08	0.46	0.28	0.76	0.14	0.23	0.27	0.16	99.2%
100 b.p. – 80% correlation	0.23	0.11	0.00	0.07	0.37	0.23	0.72	0.21	0.22	0.19	0.11	46.5%
600 b.p. – 80% correlation	0.19	0.07	0.10	0.04	0.37	0.20	0.56	0.24	0.29	0.17	0.09	0.0%

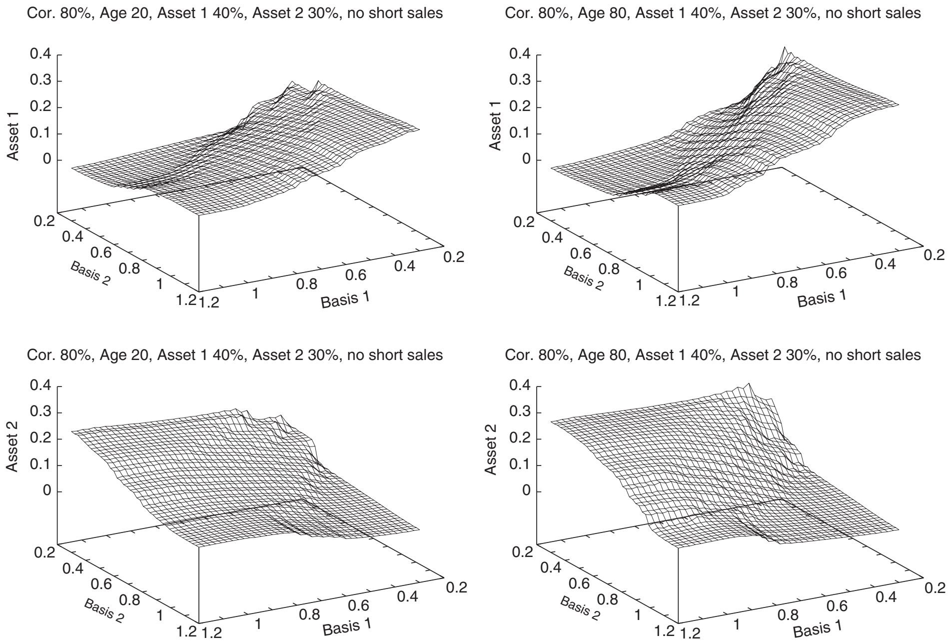


Fig. 2. *Optimal portfolio choice with short sales prohibited and* $\underline{\pi}_1 = 0.4$, $\underline{\pi}_2 = 0.3$, $\rho = 0.8$. The investor enters the age t trading period with 40% (30%) of his wealth invested in stock 1 (2). In the left panels the investor is age 20, while in the right panels the investor is age 80. The top (bottom) panels plot the optimal allocation of stock 1 (2) as a function of the basis-price ratios of the two stocks.

keep the investor from rebalancing back to a well-diversified equity portfolio. To diversify this risky stock investment, the investor purchases some stock 1 and only slightly liquidates stock 2. Since it is too costly to liquidate stock 2, this leads to being overexposed to equity with a total equity exposure of 25.5% of wealth at a basis-price ratio of 0.3. When stock 2 no longer has an embedded capital gain, the investor rebalances to a well-diversified portfolio that has an overall equity exposure of 18.0% with equal investments in each stock. When the return correlation increases to 0.8 (bottom panel), tax trading costs become more important since the diversification benefits have fallen. The investor remains undiversified and does not trade until the basis-price ratio is greater than 0.6. When the basis-price ratio reaches 1.0, the investor rebalances back to an equally weighted portfolio of the two stocks without paying capital gain taxes.

From these conditional snapshots of optimal portfolio choice, the 40% correlation case seems to exhibit few cross-equity effects, while the 80% correlation case exhibits cross-equity effects when the basis-price ratios are sufficiently different across the two stocks. Returning to the simulation analysis presented in Table 1, the effect of these changes in optimal portfolio choice relative to the one-stock case can be studied across the investor's lifetime. As in the one stock benchmark, we perform simulations starting at age 20 with no embedded gains. Results are presented for ages

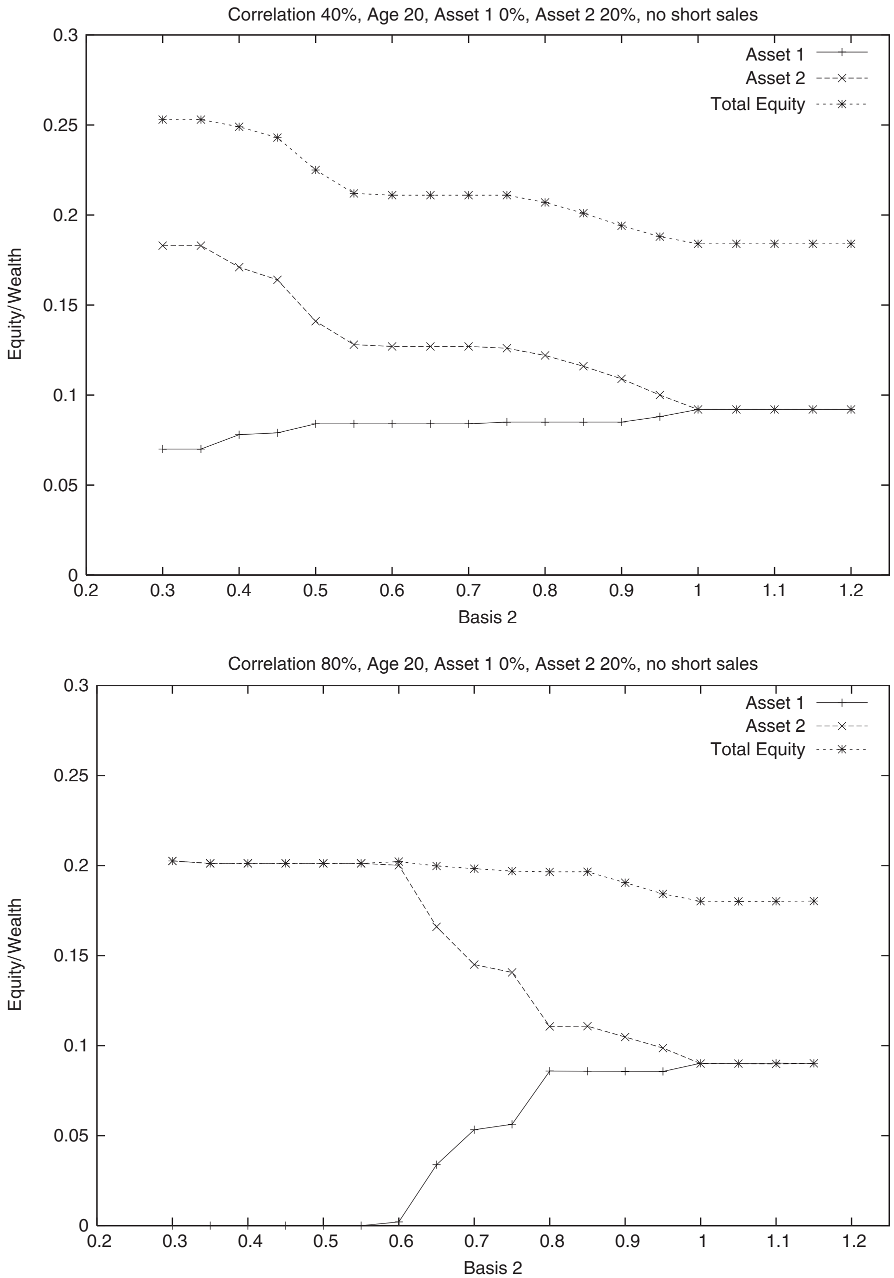


Fig. 3. *Optimal portfolio choice with short sales prohibited and* $\underline{\pi}_1 = 0.0$, $\underline{\pi}_2 = 0.2$ *at age 20*. The investor enters the trading period with 0% (20%) of his wealth invested in stock 1 (2). The top (bottom) panel is for the case when the correlation between the two stocks is $\rho = 0.4$ ($\rho = 0.8$).

40, 60, and 80 for both the 40% and 80% correlation cases in the lines labeled "Two Stock No Short Sales." The column labeled "Max Equity Allocation" records the simulation characteristics of the largest stock position, while "Min Equity Allocation" records the smallest stock position's characteristics. Given that the two stocks are ex ante identical, we arrive at the same statistics for each stock if the allocation characteristics are recorded on a stock-by-stock basis.

From the simulations, trading in two stocks is quite similar to trading in the index, as the overall mean equity allocations for both correlations are only slightly lower than the index case. However, the equity portfolio can deviate from the no-tax benchmark of equal investments in each stock. For example, at age 40, an investor who trades two stocks with an 80% correlation on average holds 22% of his wealth in equity as compared to holding 24% of his wealth in equity if he just invests in the index. He does, however, hold unequal positions in the two stocks on average. His average maximum equity allocation in one of the stocks is 13% of wealth, while his average minimum equity allocation in one of the stocks is 9% of wealth. His embedded gains in the portfolio are slightly lower than the index case: 10% of wealth as compared to 13% of wealth for the index investor. As in the single-stock case, the investor tends to hold more equity as he ages. For example, from the age 80 simulations, the investor holds 40% more equity on average than his untaxed counterpart when the correlation between the two stocks is 80%. This overexposure to equity is only slightly lower than when trading in the index and taxed on realized capital gains.

3.3. *Optimal portfolio composition with two stocks and short sales*

By allowing short selling, the investor's after-tax opportunity set is expanded. Short selling allows two additional trading strategies: a *trading flexibility strategy* in which an investor currently not overexposed to total equity ex ante shorts one stock to optimally manage realized capital gains when portfolio rebalancing in the future; and an *imperfect shorting-the-box strategy* in which an investor overexposed to equity with embedded gains ex post trades to reduce the exposure by shorting the cheaper-to-trade stock. These strategies are more effective for stocks that are highly correlated where the costs of not being well diversified are low. At a correlation between the two stocks of 40% as studied in the case with no short sales, our numerical analysis verifies that it is rarely optimal to short except at later ages when overexposed to one stock with a large embedded gain. Most of the time, an unconstrained investor acts like his constrained counterpart. Given that our setting is one where the investor's portfolio holdings are in exchange traded funds where highly correlated substitutes for particular securities are common, our discussion is focused on a setting where the correlation between the two stocks is $\rho = 80\%$ and higher.

3.3.1. *Optimal strategies*

Fig. 4 presents the optimal portfolio choice with shorting for an investor that is overinvested in equity entering age 20 or 80 with 40% of his wealth in stock 1 and

30% of his wealth in stock 2. The investor faces a shorting cost where the general collateral rate is 30 basis points below the riskless interest rate. Compared to the case with no short sales documented in Fig. 2, the optimal trading strategies are strikingly different. With shorting available, the investor no longer holds positive positions in both stocks.

As compared to the no-tax benchmark, surprisingly, the investor may choose to optimally short stock even when he has no embedded gains in either stock. For example, at age 80 when the basis-price ratio is one for both stocks, the investor invests 32% of his wealth in stock 1 and shorts 14% of his wealth in stock 2 for a net equity exposure of 18%. This *trading flexibility strategy* preserves the investor's flexibility for future asset reallocation by minimizing realized capital gains when rebalancing. This strategy leads to additional trading flexibility by providing capital losses in the portfolio when they are most needed. For example, consider two different ways of holding a net equity position of 18% of wealth as the investor does at age 80 in Fig. 4 with no embedded gains. To simplify the discussion, assume that the two stocks are perfectly correlated. In the first strategy, the investor holds 9% of his wealth in each stock. In the second strategy, the investor holds 32% of his wealth

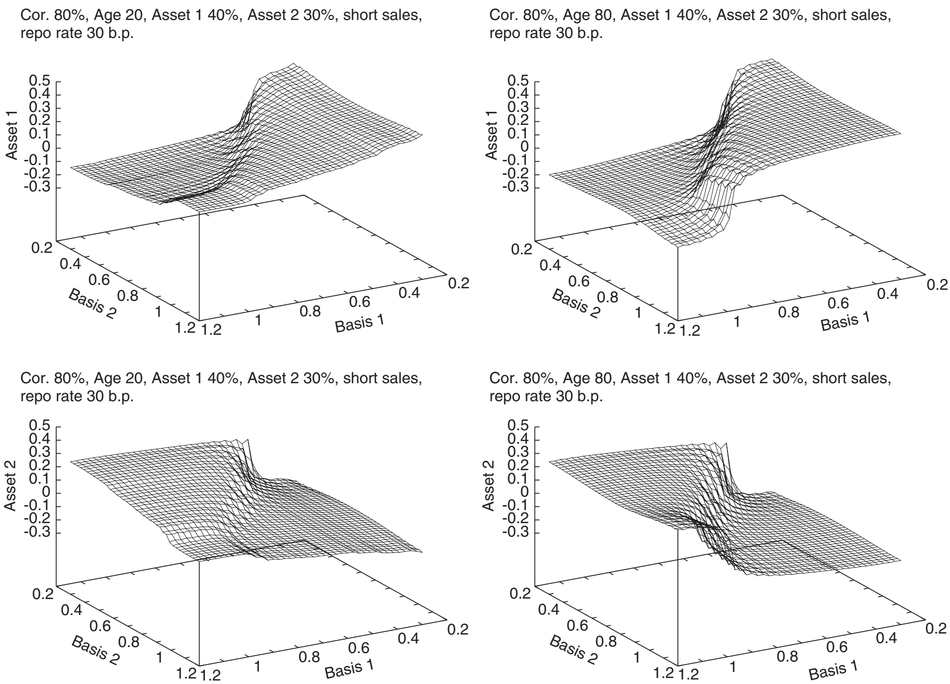


Fig. 4. *Optimal portfolio choice with short sales allowed*, $\underline{\pi}_1 = 0.4$, $\underline{\pi}_2 = 0.3$, $\rho = 0.8$, *and shorting costs of 30 basis points*. The investor enters the age t trading period with 40% (30%) of his wealth invested in stock 1 (2). In the left panels the investor is age 20, while in the right panels the investor is age 80. The top (bottom) panels plot the optimal allocation of stock 1 (2) as a function of the basis-price ratios of the two stocks. The investor can sell short subject to a shorting cost of 30 basis points.

in stock 1 and −14% of his wealth in stock 2 as he does at age 80 in the no-embedded-gains region. With no capital gain taxes, these positions would be identical. However, with capital gain taxes, the second strategy leads to more trading flexibility.

If the stock market increases over the next year, the investor's proportion of wealth in equity increases. Without taxes, the investor optimally sells some equity to rebalance back to his optimal total equity-wealth ratio. Under strategy 1, the investor has an embedded gain in both stock positions. To rebalance he pays capital gain taxes on the amount he sells. Strategy 2, however, gives the investor a way to rebalance while paying lower capital gain taxes than in strategy 1. By liquidating his losses in the short position in stock 2, the investor can offset the realized capital gain from rebalancing his stock 1 position. (This assumes that the capital loss from stock 2 is large enough to offset the realized gain in rebalancing stock 1, as occurs for the case considered.) Hence, strategy 2 creates a capital loss when it is most useful—when the investor's aggregate equity position has increased in value and he needs to sell to rebalance. If the stock market falls, the investor sells his long stock positions to create a tax loss under either strategy. However, under strategy 2 the short position in stock 2 now has an embedded gain. In the worst case scenario, he can also liquidate this position. His realized capital loss for the entire position under strategy 2 is identical to his realized capital loss under strategy 1. As a result, the investor is no worse off when the stock market falls under strategy 2 as compared to owning 9% of his wealth in each stock (strategy 1). Combining the two possible market outcomes, the investor is better off under strategy 2.

Returning to Fig. 4, when the investor's portfolio contains embedded gains, his strategy is to short the stock that is cheapest to trade. This *imperfect shorting-the-box strategy* leads to an overall exposure to equity that is smaller than that of a short-sale-constrained investor. For example, at age 20 with an initial position of 40% in stock 1 and 30% in stock 2 and a basis-price ratio of 0.3 for both stocks, the investor liquidates some of his stock 1 position which falls to 38.9% of wealth while shorting 15.7% of his wealth in stock 2. By doing so, the investor's net equity position is reduced to 23.2% as compared to 30.0% when short sales are not allowed. This imperfect shorting-the-box strategy is optimal even though the investor faces fundamental price risk given that the two stocks are not perfectly correlated.

3.3.2. Trading flexibility strategy

To explore the sensitivity of the trading flexibility strategy to different shorting costs and correlations, Fig. 5 examines the optimal stock allocation as a function of age for three different general collateral rates: 0 basis points (top panels), 30 basis points (middle panels), and 50 basis points (bottom panels) below the money market rate. In the three left plots, the investor's portfolio entering age t contains no embedded gains in either stock. The three right plots capture the scenario when the investor's total equity exposure entering age t is close to the no-tax optimal, but each stock contains embedded gains. Specifically, the investor's equity position entering age t is 10% of his wealth in each stock with a 0.5 basis-price ratio for each stock.

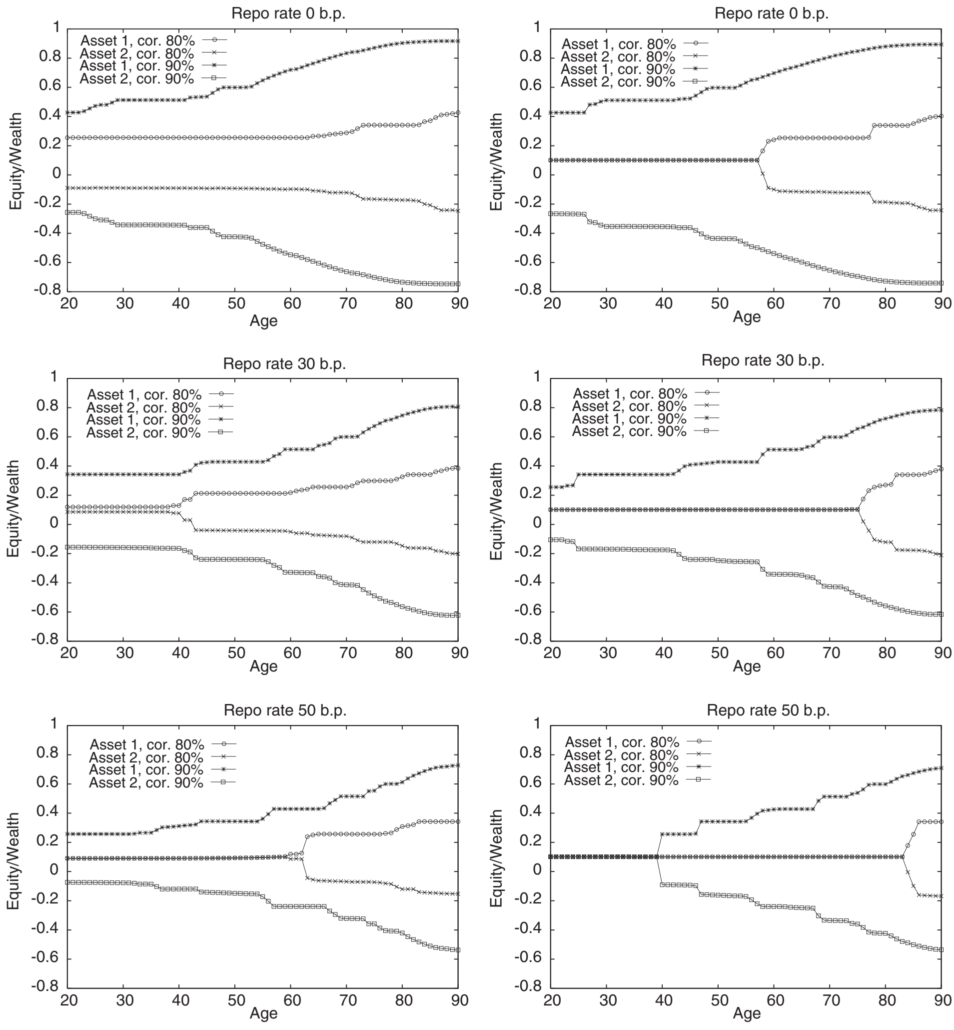


Fig. 5. *Trading flexibility strategy use versus age*. The plots examine the optimal allocation in the risky assets versus age as a fraction of wealth when the investor is engaged in the trading flexibility strategy and the correlation between the two stocks is $\rho = 0.8$ or $\rho = 0.9$. For the top, middle, and bottom panels, the investor can sell short subject to shorting costs of 0, 30, and 50 basis points. In the left panels, the investor enters the trading period with no embedded gains in either security. In the right panels, the investor enters the trading period with 10% of his wealth in each stock and a basis-price ratio of 0.5 for each stock.

Here the investor might desire to engage in the flexibility strategy to reduce future tax trading costs even though he will pay some capital gain taxes today to rebalance.

From Fig. 5, the use of the trading flexibility strategy is decreasing in the shorting cost and increasing in age and correlation. At a correlation of 90%, the flexibility strategy is used from age 20 for all shorting costs; however, the magnitude of the individual stock positions is reduced for higher shorting costs moving down the figure. Comparing the left and right panels, the use of the trading flexibility strategy is reduced when the investor's portfolio entering age t contains embedded gains (the right panels). This reduction manifests itself by the investor both engaging in the strategy later in life and holding smaller absolute positions in each stock. However, even when each stock has a basis-price ratio of 0.5, the investor is willing to realize capital gains at age 20 (age 40) with a correlation of 90% and a shorting cost of 0 or 30 (50) basis points to engage in the flexibility strategy. As the correlation between the two stocks falls, the usefulness of the trading flexibility strategy decreases because of the increasing probability that both the long and short legs of the portfolio simultaneously generate capital gains. This reduces the incentive to engage in such a tax management strategy. While the figure shows that utilizing the trading flexibility strategy is optimal when the correlation between the two stocks is $\rho = 80\%$, in numerical results that we do not report we found that it is optimal to utilize the trading flexibility strategy, given our parameters, if the correlation is at least $\rho = 65\%$.

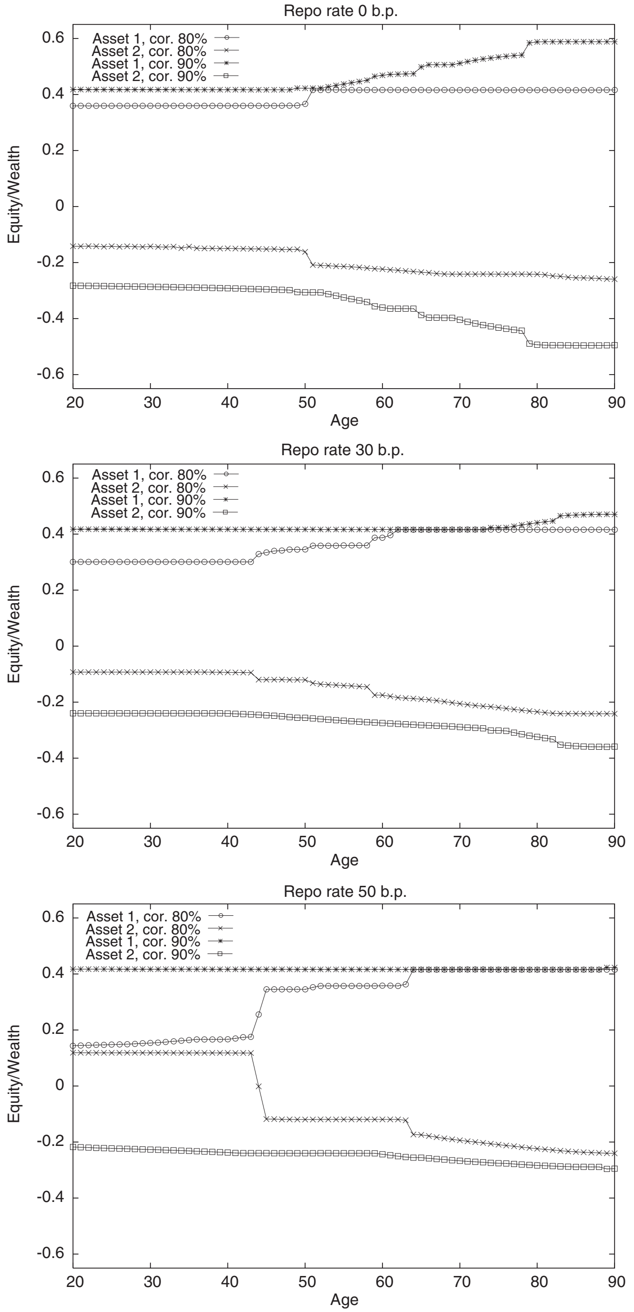
3.3.3. *Imperfect shorting-the-box strategy*

Fig. 6 explores the use of the imperfect shorting-the-box strategy for different correlations and shorting costs: 0 basis points (top panel), 30 basis points (middle panel), and 50 basis points (bottom panel) below the money market rate. The investor entering age t is grossly overinvested in equity with 40% (30%) of wealth in stock 1 (2). The investor also has large embedded gains in each stock—the basis-price ratio is 0.5 for each stock.

In contrast to the case with no short sales where the investor rebalances to a positive position in both stocks, an investor who can sell short is willing to fully realize his capital gain in stock 2 to establish an imperfect shorting-the-box strategy. By doing so, the investor can significantly reduce his total equity exposure. The desire to short stock 2 is driven by the cost of shorting as well as the price risk. When the general collateral rate is 0 or 30 basis points below the riskless money market rate, the investor shorts stock 2 at all ages. For the higher shorting cost of 50 basis points in the bottom panel, the investor does not short stock 2 until age 45 when the correlation is 80%. At a 90% correlation, the fundamental price risk has been reduced enough that the investor in all cases shorts stock immediately at age 20. In addition, the 90% correlation case shows that with less fundamental price risk the investor is willing to take larger absolute positions in the two securities that increase with age. Given that the correlation is higher, the probability of both the long and short positions realizing gains next period is smaller, leading to larger positions. The

→

Fig. 6. *Imperfect shorting-the-box use versus age*. The plots examine the optimal allocation in the risky assets versus age as a fraction of wealth when the investor is engaged in the imperfect shorting-the-box strategy and the correlation between the two stocks is $\rho = 0.8$ or 0.9. For the top, middle, and bottom panels, the investor can sell short subject to shorting costs of 0, 30, and 50 basis points respectively. The investor enters the age t trading period with 40% (30%) of his wealth invested in stock 1 (2) and a basis-price ratio across both stocks of 0.5.


Repo rate 0 b.p.
Asset 1, cor. 80%
Asset 2, cor. 80%
Asset 1, cor. 90%
Asset 2, cor. 90%
Equity/Wealth
Age
Repo rate 30 b.p.
Asset 1, cor. 80%
Asset 2, cor. 80%
Asset 1, cor. 90%
Asset 2, cor. 90%
Equity/Wealth
Age
Repo rate 50 b.p.
Asset 1, cor. 80%
Asset 2, cor. 80%
Asset 1, cor. 90%
Asset 2, cor. 90%
Equity/Wealth
Age

imperfect shorting-the-box strategy is also employed when the return correlation between the two stocks is considerably lower than 80%. In numerical simulations, we find that it is optimal to utilize the imperfect shorting-the-box strategy, given our parameters, if the correlation is at least $\rho = 40\%$ at later ages when the investor is overexposed to one stock with a large embedded gain.

3.3.4. *Implications for stock allocation relative to no short sales*

The use of the trading flexibility strategy highlights an interesting feature of short selling.[9] With no embedded gains, a short-sale-constrained investor holds a positive position in each stock, but if the constraint is relaxed, he optimally shorts one of the stocks at a low enough shorting cost. Figs. 7 and 8 explore this feature across all basis distributions by considering a risky stock portfolio where the investor entering age 20 or 80 owns equal weights in each security. Each panel explores the characteristics of optimal portfolio choice across the basis-price ratios of the two stocks. In the region marked "1," a constrained investor trades to positive holdings in both stocks, while a shorting investor trades to a negative position in one stock. If both investor types hold positive positions in each stock, the region is marked "2." Finally, the region marked "3" corresponds to the short-sale-constrained investor holding a zero position in one stock and the unconstrained investor shorting one of the stocks. The left (right) panels characterize trading behavior at age 20 (80). The total portion of wealth in equity entering the trading period increases moving down the panels. We present results when the equity exposure in each stock is 10% (the results are identical for an equity exposure in each stock less than or equal to 10%) and 20% of wealth initially. The correlation between the two stocks is 80% for Fig. 7 and 90% for Fig. 8; the general collateral rate when short selling is 30 basis points below the money market rate.

From both figures, it is very common for an unconstrained investor to short, while a constrained investor holds positive weights in each stock (region "1"). Such behavior is most prominent at age 80. For this age, the only time the short-sale-constrained investor holds a zero position in one of the stocks is in the two regions marked "3" in the lower right panel of the 90% correlation figure. Given that in this panel the investor holds 40% of his wealth in equity entering the trading period, an unconstrained investor always shorts one of the stocks to engage in the trading flexibility or the imperfect shorting-the-box strategy. The constrained investor entirely liquidates one of his equity positions only when it has small embedded gains relative to the other stock. At age 20, this behavior is more dependent on the return correlation between the two stocks. In the top left panel of Fig. 7 where the investors are not overexposed to equity, both types of investors hold strictly positive positions in both stocks. This behavior is driven by the flexibility strategy not being optimal for the unconstrained investor at age 20 and a return correlation of 80%. However, when the return correlation increases to 90% (top left panel of Fig. 8), the unconstrained investor at age 20 again shorts while the constrained investor holds positive weights in each stock.

[9]We thank the referee for drawing this feature to our attention.

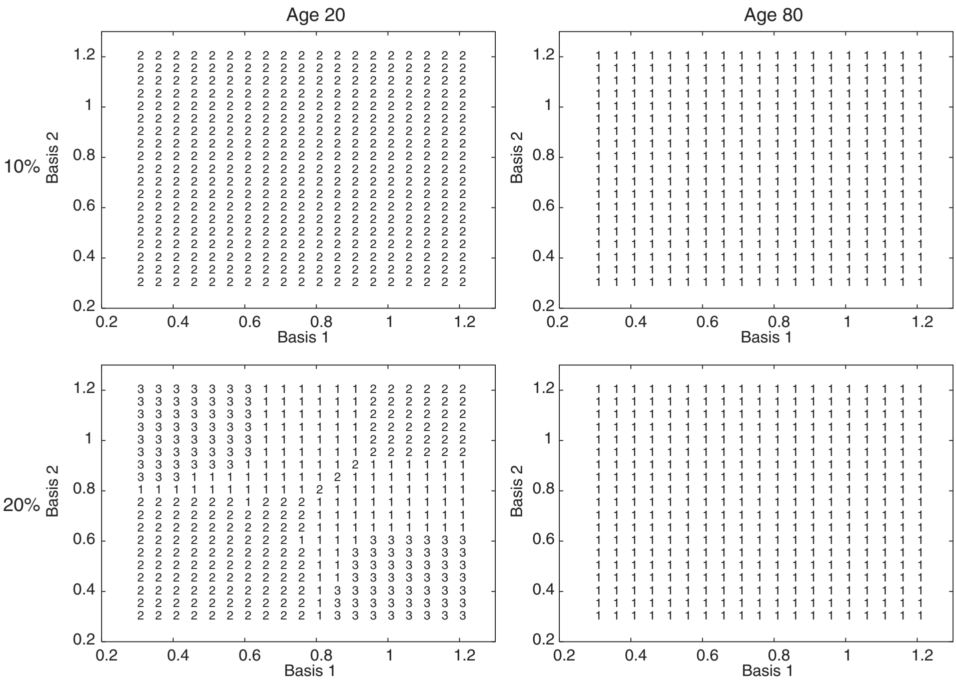


Fig. 7. *Regions of shorting behavior—80% correlation*. The plots examine the differences in asset holdings with and without a short-sale constraint across different basis-price ratios for the two stocks. The left (right) panels are for age 20 (80). In the top and bottom panels, the investor's fraction of wealth invested in each stock entering the trading period is 10% and 20%, respectively. In the region marked "1," the short-sale-constrained investor trades to positive holdings in each stock, while an unconstraint investor trades to a position that shorts one of the stocks. In the region marked "2," both types of investors hold long positions in each security. In the region marked "3," the short-sale-constrained investor trades to a zero position in one stock, while an unconstrained investor optimally shorts one of the stocks. The correlation between the two stocks is $\rho = 0.8$. The shorting cost is 30 basis points.

This discontinuity in the optimal asset allocation occurs because the net tax benefit of selling shares is not monotonic in the trading strategy when shorting is allowed. When the initial portfolio contains no embedded gains, an unconstrained investor reaps the benefits of the trading flexibility strategy by shorting one of the stocks to generate a capital loss in the portfolio when the total equity position must be reduced in the future. Such a trade only has a benefit to the investor in the short selling region. Constrained investors cannot capture the benefit given that they can only trade to a zero position that will not generate a capital loss when the overall equity position needs to be scaled back next period. Likewise, when an investor is initially overexposed to equity, he liquidates the stock with the smallest marginal tax trading cost or the highest basis-price ratio. For a constrained investor, it might be too costly to fully liquidate to a zero position in this stock. However, it may be optimal for an unconstrained investor to short this security. Once he has closed the long position, today's marginal tax cost of shorting is zero, but the marginal benefit through

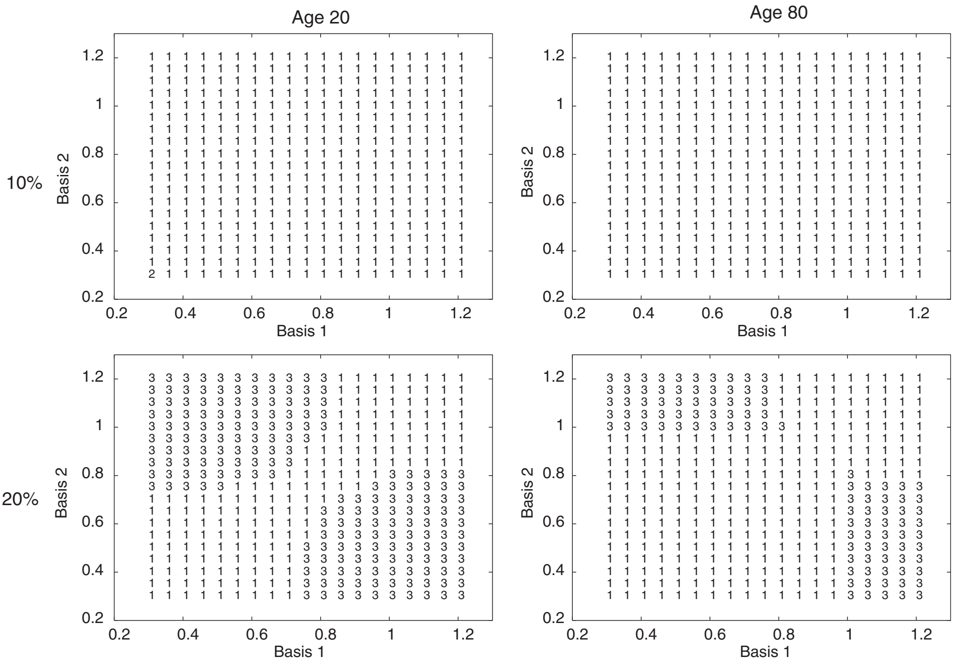


Fig. 8. *Regions of shorting behavior—90% correlation*. The plots examine the differences in asset holdings with and without a short-sale constraint across different basis-price ratios for the two stocks. The left (right) panels are for age 20 (80). In the top and bottom panels, the investor's fraction of wealth invested in each stock entering the trading period is 10% and 20%, respectively. In the region marked "1," the short-sale-constrained investor trades to positive holdings in each stock, while an unconstrained investor trades to a position that shorts one of the stocks. In the region marked "2," both types of investors hold long positions in each security. In the region marked "3," the short-sale-constrained investor trades to a zero position in one stock, while an unconstrained investor optimally shorts one of the stocks. The correlation between the two stocks is $\rho = 0.9$. The shorting cost is 30 basis points.

reducing the total equity exposure and using the trading flexibility strategy is still positive.

3.3.5. Simulation results

To determine whether short selling translates into significantly different lifetime unconditional portfolio allocations relative to the constrained case, we again turn to our simulation analysis summarized in Table 1. In addition to the allocation distributions examined in the case with no short selling, the last column of the table labeled "Time Shorting" reports the fraction of simulation paths at each age when the investor is shorting one of the stocks. The simulations reveal that unconditional portfolio allocations are greatly influenced by the ability to sell short. First, shorting behavior is prominent. At age 40 and a shorting cost of 30 basis points in Panel A, the investor is shorting 55.3% of the time. By age 60 (80) in Panel B (C), even an investor who faces a shorting cost of 100 basis points is shorting 29.3% (46.5%) of the time. This shorting behavior leads to the investor better managing his net equity exposure relative to the money market account. For example, at age 80 in the column labeled "Total Equity Mean" in Panel C, an unconstrained investor paying shorting costs of 30 basis points on average holds 21% of his wealth in equity — a 25% reduction as compared to a constrained investor. Additionally, this allocation is only slightly larger than the optimal total equity allocation of 20% with no capital gain tax. Shorting also leads to higher embedded gains in the portfolio. At age 80, an unconstrained investor paying shorting costs of 30 basis points has embedded gains of 34% of his financial wealth on average, as compared to 17% for a constrained investor. These embedded gains are concentrated in one stock as can be seen from the reported average bases of the two stocks. Essentially, the investor's portfolio generates increased embedded gains in one stock, but the investor is able to maintain an equity exposure closer to the no-tax benchmark by shorting the other stock. This leads to the composition of the short seller's equity portfolio becoming less diversified with age since the sizes of the long and short equity positions grow, as can be seen at ages 60 and 80.

This increased embedded gain relative to the case with no short sales highlights the ability of the trading flexibility strategy to defer gains. If instead we assume that capital gains are not forgiven at death as in the Canadian tax code, the investor no longer uses the trading flexibility strategy due to the large embedded gain it generates that will not be forgiven. Even under the Canadian tax code, however, the investor will still short one stock if he is currently overinvested in the other stock with a large embedded gain.[10]

4. Economic significance and comparative statics of optimal portfolios

Our analysis of the structure of optimal portfolios in Section 3 highlights that when short-sale constrained, an investor's portfolio experiences limited cross-holding effects due to taxation even when the correlation is high. In contrast, allowing costly short selling leads to optimal portfolios that are greatly influenced by tax trading costs. By employing the trading flexibility and imperfect shorting-the-box strategies, it is common for the investor to be short one stock and long the other when an otherwise identical short-sale-constrained investor holds positive positions in both stocks. However, we have yet to address how economically beneficial it is to trade the components of the stock index relative to the index itself. Table 2 reports the age 20 lifetime utility benefits of investing in multiple risky stocks as well as the no-embedded-gains stock allocations at ages 20, 40, and 60. Panel A considers the base case model parameters used in the previous section, while Panel B considers a variety of different parameterizations.

The wealth benefit ratios measure the fraction of wealth an age 20 index investor would pay to be indifferent between his current index fund investment and a two-stock equity portfolio with or without costly shorting. In contrast to existing

[10]These results are available from the authors.

Table 2
Comparative statics and wealth benefits

This table summarizes the differences in trading strategies across a variety of different scenarios as well as the wealth benefit of each scenario as compared to investing in a single index fund. The wealth benefit is the fraction of wealth an investor at age 20 who currently invests in a single index fund would pay to be indifferent to switching to each scenario. Panel A summarizes the results when the investor has a relative risk aversion coefficient of 3. Panel B considers several different scenarios where risk aversion, volatility, margin requirements, and bequest motives are changed from the base case. The "No Embedded Gains Stock Allocations" columns summarize the optimal portfolio choice for an investor with no embedded gains across ages 20, 40, and 60. The "Wealth Benefit Ratio at Age 20" columns summarize the wealth benefit ratios for no embedded gains and for three different embedded gains scenarios when the investor holds 40% of his financial wealth in an equally weighted stock portfolio.

		No embedded gains stock allocations						Wealth benefit ratio at age 20			
		Age 20		Age 40		Age 60		No embedded gains	Total equity allocation 40%		
		Stock 1 (%)	Stock 2 (%)	Stock 1 (%)	Stock 2 (%)	Stock 1 (%)	Stock 2 (%)	$b_1 = 1.0$, $b_2 = 1.0$(%)	$b_1 = 0.5$, $b_2 = 1.0$(%)	$b_1 = 0.5$, $b_2 = 0.75$(%)	$b_1 = 0.5$, $b_2 = 0.5$(%)
Panel A: base case–risk aversion 3											
Correlation 40%											
No short sales		9.2	9.2	9.3	9.3	9.8	9.8	0.2	0.4	0.3	0.2
Short sales	−0 b.p.	9.3	9.3	9.3	9.3	9.8	9.8	0.2	0.5	0.4	0.2
Correlation 80 %											
No short sales		9.0	9.0	9.0	9.0	9.2	9.2	0.3	0.7	0.5	0.3
Short sales	−0 b.p.	21.2	−6.0	21.5	−6.0	25.5	−9.6	2.0	2.6	2.4	2.0
	−30 b.p.	11.9	8.5	12.9	7.7	21.7	−5.0	1.5	1.9	1.8	1.5
	−50 b.p.	9.0	9.0	9.1	9.1	12.0	8.6	0.8	1.3	1.1	0.8
	−100 b.p.	9.0	9.0	9.0	9.0	9.3	9.3	0.4	0.9	0.7	0.4
	−600 b.p.	9.0	9.0	9.0	9.0	9.2	9.2	0.3	0.7	0.5	0.3
Correlation 90 %											
No short sales		9.1	9.1	9.1	9.1	10.1	8.6	0.4	0.8	0.6	0.4
Short sales	−0 b.p.	42.8	−25.6	51.3	−34.4	72.0	−54.7	10.1	10.7	10.4	10.1
	−30 b.p.	34.3	−15.5	34.3	−16.4	51.4	−32.8	5.8	6.4	6.1	5.8
	−50 b.p.	25.7	−7.4	31.2	−11.9	42.9	−23.9	3.8	4.3	4.1	3.8
	−100 b.p.	9.1	9.1	17.2	2.1	25.9	−4.6	1.0	1.6	1.3	1.0
	−600 b.p.	9.1	9.1	9.1	9.1	10.1	8.6	0.4	0.8	0.6	0.4

Panel B: *Comparative statics*

Correlation 80 %											
Risk aversion 4											
No short sales		7.5	7.5	7.6	7.6	8.2	8.2	0.2	0.5	0.3	0.2
Short sales	−30 b.p.	7.5	7.5	8.3	8.3	16.9	−3.7	0.5	0.9	0.7	0.5
Index volatility 30 %											
No short sales		4.0	4.0	4.0	4.0	4.0	4.0	0.1	0.1	0.1	0.1
Short sales	−30 b.p.	15.5	−9.6	15.5	−9.6	15.5	−9.8	2.1	2.3	2.3	2.2
Margin requirement											
Short sales – long margin increased to 75 %		11.9	8.5	12.9	7.7	21.7	−5.0	0.6	1.0	0.8	0.6
Short sales – short margin increased to 100 %		11.9	8.5	12.9	7.7	21.7	−5.0	0.6	1.0	0.8	0.6
No bequest motive											
No short sales		9.0	9.0	9.0	9.0	9.0	9.0	0.0	0.1	0.0	0.0
Short sales	−30 b.p.	9.0	9.0	9.0	9.0	9.0	9.0	0.0	0.2	0.1	0.0
Correlation 90 %											
Risk aversion 4											
No short sales		8.5	6.2	8.5	6.2	8.5	7.2	0.2	0.6	0.5	0.3
Short sales	−30 b.p.	27.7	−19.6	27.7	−19.6	42.3	−29.4	3.9	4.5	4.3	4.1
Index volatility 30 %											
No short sales		4.1	4.0	4.1	4.0	4.3	4.0	0.1	0.2	0.2	0.1
Short sales	−30 b.p.	23.3	−17.0	23.3	−17.0	26.1	−20.1	6.7	7.0	7.0	6.9
Margin requirement											
Short sales – long margin increased to 75 %		34.3	−15.5	34.3	−16.4	51.4	−32.8	4.1	4.5	4.4	4.2
Short sales – short margin increased to 200 %		34.3	−15.5	34.3	−16.4	51.4	−32.8	3.9	4.4	4.2	3.9
No bequest motive											
No short sales		9.0	9.0	9.0	9.0	9.0	9.0	0.0	0.2	0.1	0.0
Short sales	−30 b.p.	25.6	−8.5	25.6	−8.5	25.6	−8.5	1.3	1.6	1.4	1.3

literature (Constantinides, 1983; Dammon et al., 2001b; Garlappi et al., 2001), we do not measure the wealth benefits relative to an accrual-based capital gain taxation system where all gains and losses are marked to market annually. Instead, our wealth benefit ratio is meant to capture the marginal contribution to investor welfare of switching from a single index fund to a portfolio of risky stocks—measuring the wealth benefit relative to an accrual-based tax would lead to significantly higher wealth benefits. We present results when the initial portfolio has no embedded gains as well as when the investor is initially overexposed to equity with embedded gains. Specifically, we examine the wealth benefit for an investor who holds 40% of his financial wealth in an equally weighted equity portfolio under three different embedded gain scenarios.

When short sales are prohibited, the wealth benefits from Panels A and B show a striking feature—the benefits of trading the components of the index are very small. From Panel A, the wealth benefit ratio of trading each stock separately with a short sale constraint is 0.2%, 0.3%, and 0.4% for stock return correlations of 40%, 80%, and 90%, respectively, with no embedded gains. Additionally, the wealth benefits are small across the comparative static exercises of Panel B. Given the small differences in the simulation results in Table 1 between the one- and two-stock cases with no short sales, the small wealth benefits are not necessarily surprising.

When short sales are allowed, the wealth benefits increase relative to the case with no short sales at age 20, but are sensitive to the correlation between the two stocks. At 90% correlation, wealth benefit ratios are 10.1%, 5.8%, and 3.8% with no embedded gains and shorting costs of 0, 30, and 50 basis points, respectively. When more price risk is introduced when engaging in a short position by dropping the stock correlation to 80%, the wealth benefit drops to 2.0%, 1.5%, and 0.8% for respective shorting costs of 0, 30, and 50 basis points. Although these are smaller than the 90% correlation setting, they are significantly larger than the wealth benefit ratios with no short sales. When the shorting costs increase to a rate faced by a small investor (600 basis points), the welfare benefits of shorting disappear, leading to the same welfare benefits as the constrained case. In summary, the bulk of the economic benefits of splitting the index into its components are through short selling when shorting costs are not too high—short-sale-constrained portfolios only lead to small wealth benefit increases. The optimal consumption strategies (not reported) under the different stock trading scenarios mirror the welfare results in that allowing short selling with low shorting costs leads to higher consumption rates. In addition, when embedded gains are large, the investor reduces his consumption slightly due to an effectively lower wealth today given future tax liabilities.

To understand the sensitivity of the results to different model parameters, Panel B of Table 2 considers several variations on the base case set of parameters when the return correlation between the two stocks is 80% or 90%. The cases of no short sales and short sales with a shorting cost of 30 basis points are again considered. The optimal stock allocations with no embedded gains are presented for ages 20, 40, and 60 along with the wealth benefit ratios for age 20.

First, the relative risk aversion is increased to 4 from the base case of 3. Not surprisingly, the total equity positions are decreasing in risk aversion. Given smaller equity positions in the portfolio for higher risk aversions, the wealth benefit ratios are slightly lower than the base case. Second, increasing the index volatility to 30% leads to lower pre-tax Sharpe ratios and hence lower equity investments. However, in contrast to when the risk aversion is increased, the wealth benefit ratios actually increase when the index volatility is 30%. This occurs because the value of the tax loss selling option in the portfolio increases in volatility as in Constantinides (1983).

We also consider tightening either the long margin requirement to 75% from 50% or the short margin requirement to 200% from 150%. From Panel B of Table 2, the optimal equity positions with no embedded gains are not influenced up to age 60 by tightening the margin constraints. However, the trading strategy is influenced at later ages. To gain insight on how frequently the margin constraints bind, we use our simulation analysis for the 80% return correlation case where we tabulate the frequency of binding margin constraints in our base case parameters with shorting costs of 30 basis points. From the simulation starting at age 20 with no embedded gains, the margin constraints never bind before age 70. Along a path starting at age 20, the unconditional probability that the constraint eventually binds is 4.3%. If the long margin requirement increases to 75% from 50%, this probability increases to 8.0%. If instead the short margin requirement increases to 200% from 150%, the unconditional probability that the margin constraint eventually binds is 15.2%. This translates into lower wealth benefit ratios as can be seen in the table.

Finally, we examine the role of the bequest motive on optimal behavior and on the wealth benefits. The last entry in Panel B of Table 2 considers the case when the investor has no bequest motive. At a stock return correlation of 80%, the stock allocations with no embedded gains are identical whether or not short sales are allowed. With no bequest motive, wealth benefits are also significantly reduced. At the 80% correlation, no wealth benefit is larger than 0.2%. However, this is still larger than the case with no short sales due to the investor optimally employing the imperfect shorting-the-box strategy in some scenarios when overexposed to one stock with a large embedded gain. When the return correlation increases to 90%, the investor again engages in the trading flexibility strategy when short selling is allowed. By doing so, the investor can defer realizing capital gains leading to wealth benefits no smaller than 1.3% at age 20.

5. Tax management through derivative trading

While we have established that large investors who pay low shorting costs are better off trading components of an index than the index itself, our results also indicate that small investors with large shorting costs are effectively precluded from following the tax management strategies we have described. In this section we present a different strategy that most investors would be able to implement, restoring the ability of investors to balance gains in one stock with losses in another to minimize realized capital gains when portfolio rebalancing.

This strategy is related to the one discussed in Constantinides and Scholes (1980), where an investor uses derivatives to defer capital gain taxes. We investigate the case where the investor trades in one stock while holding a one-year, at-the-money put option on another highly correlated stock. We focus on trading a put option on a stock with high correlation but not actually held by the investor since buying put options against positions owned by an investor can be viewed as a sale for tax purposes. Holding a put is similar to the tax management strategies we described for the wealthy investor who can sell short at a low cost since a put can be decomposed into a short position in the stock and an investment in the money market account. The put generates gains (losses) when the stock tends to generate losses (gains). Since the put option is reset every trading period, this setting is especially tractable since the only state variables needed to characterize the investor's portfolio choice at a particular age are the stock's basis and position entering the trading period. The underlying stocks have the same characteristics as in the cases we have considered so far. The parameters used are the same as described in Section 2.4. We also assume that the investor is precluded from short positions in any security.

Pricing the put option by no arbitrage is complicated by the taxes in our economy. When the participants in the option market are differentially taxed on the option's cash flows, it is not immediately clear who is the marginal pricer of the option given that different valuations can arise for different tax rates. These issues are discussed in detail in Schaefer (1982), Dybvig and Ross (1986), Dammon and Green (1987), and Ross (1987). For our portfolio choice problem with a put option, we assume that the price of the option is set by an investor who faces realized capital gain taxation and can costlessly short. The option price is determined by using standard replication arguments on the option's after-tax payoff. The after-tax valuation actually leads to a higher price than if the pre-tax cash flows are used to value the option. This is because the put option generates a positive after-tax cash flow when the stock price increases due to the capital loss write-off from the put option. By using this higher put price, our calculated wealth benefits of trading the put are understated as compared to a put option valued using pre-tax cash flows. The modifications of our earlier model to incorporate investing in a put option and computing its price are provided in Appendix B.

Table 3 describes the optimal strategy as well as the welfare benefits for different correlations for a 20-year-old investor. Panel A presents the optimal trading strategy when the only risky security is the stock index combining the two individual stocks in an equally weighted portfolio. Panel B documents optimal portfolio choice across different correlations between the two stocks when the risky investment opportunity set consists of stock 1 and the put written against stock 2. For each correlation between the two stocks, the top line gives the optimal portfolio when the portfolio initially contains no embedded gains. The other four lines describe optimal portfolio choice when the investor initially holds 20% or 40% of his wealth in stock 1 with basis-price ratio of either 0.8 or 0.4.

By trading in the put, the investor is able to engage in both the trading flexibility and imperfect shorting-the-box strategies from the previous section. From the table, the investor uses the put to construct a trading flexibility strategy when he initially holds no embedded gains in his portfolio. When the stock return correlation is 90% with no embedded gains in the portfolio, the investor places 0.6% of his wealth in the

Table 3
Optimal stock and put option trade

This table presents the optimal investment in stock 1 and an at-the-money put written on stock 2, for different correlations, initial allocations in stock 1, and initial basis-price ratios for an age 20 investor. The optimal allocations are expressed as a percentage of wealth. The effective total stock allocation is computed by translating the put position to an equivalent short position in stock 2 by multiplying the put position by the put's delta and adding it to the stock 1 position. The wealth-benefit ratio is computed as the fraction of wealth that an age 20 investor who trades in an equally weighted index of the two stocks would pay to be indifferent between trading his current equity portfolio and an equity portfolio consisting of stock 1 and a put option written on stock 2. Panel A corresponds to investing in an equally weighted index. Panel B corresponds to investing in one of the stocks that make up the index and a put option written on the other stock.

	Initial stock allocation	Basis-price ratio	Optimal stock 1 allocation %	Optimal put allocation %	Stock 2 exposure from put allocation %	Effective total stock allocation %	Wealth benefit ratio at age 20 %
Panel A: *stock index only*	0.0	1.0	18.2	0.0	0.0	18.2	
	0.2	0.8	20.2	0.0	0.0	20.2	
	0.2	0.4	20.2	0.0	0.0	20.2	
	0.4	0.8	21.7	0.0	0.0	21.7	
	0.4	0.4	34.7	0.0	0.0	34.7	
Panel B: *stock 1 and put option*							
Correlation 85%							
	0.0	1.0	17.6	0.0	0.0	17.6	0.53
	0.2	0.8	20.2	0.4	−1.7	18.5	0.55
	0.2	0.4	20.3	0.6	−2.8	17.5	0.61
	0.4	0.8	21.5	0.6	−2.7	18.8	0.55
	0.4	0.4	35.7	3.5	−16.7	19.1	1.71
Correlation 90%							
	0.0	1.0	21.5	0.6	−2.9	18.6	1.84
	0.2	0.8	21.4	0.9	−4.2	17.2	1.88
	0.2	0.4	20.3	0.9	−4.2	16.1	1.93
	0.4	0.8	30.2	2.6	−12.4	17.8	1.94
	0.4	0.4	40.4	4.8	−23.0	17.5	3.46
Correlation 95%							
	0.0	1.0	34.7	3.6	−17.3	17.4	4.55
	0.2	0.8	34.6	3.8	−18.4	16.2	4.57
	0.2	0.4	26.3	2.4	−11.7	14.6	4.60
	0.4	0.8	40.5	5.2	−25.0	15.5	4.79
	0.4	0.4	40.5	5.6	−26.8	13.7	6.48

put. When the correlation increases to 95%, the put investment increases to 3.6% of wealth. From the put's delta, the column labeled "Stock 2 Exposure from Put Allocation" gives the synthetic stock 2 portfolio position as a fraction of wealth. For the case with no embedded gains, the investor's synthetic stock position is −2.9% of wealth when the stock return correlation is 90% and −17.3% of wealth when the stock return correlation is 95%.

The investor also uses the put to construct an imperfect form of the shorting-the-box strategy. For example, when the return correlation between the two stocks is 90%, the investor places 4.8% of his wealth in the put (a synthetic position of −23% of his wealth in stock 2) and does not trade his stock 1 position when he initially holds 40% of his wealth in equity with a basis-price ratio of 0.4. By so doing, the investor's effective total equity exposure falls to 17.5% of wealth.[11] If instead the investor holds only the index (Panel A), his equity exposure only falls to 34.7% of wealth.

We note that the welfare benefits at age 20 of investing in a put are sensitive to the return correlation as well as to the pricing of the put.[12] However, the put's welfare benefits are comparable to the welfare benefits of an investor who can sell short subject to low shorting costs even using the conservative valuation of the put. For example, with no embedded gains, the wealth benefit of investing in the put is 0.53%, 1.84%, and 4.55% across return correlations of 85%, 90%, and 95%, respectively. If instead we assume the put is priced by no arbitrage by a tax-exempt investor leading to a lower initial price, the respective wealth benefits (not reported in the table) across the three previous correlations are significantly higher: 4.6%, 7.6%, and 12.4% of wealth. This analysis indicates that investing in puts is a viable alternative for the tax management of capital gains for investors who face large shorting costs.

6. Conclusion

In selecting an optimal consumption-investment strategy, investors need to balance two considerations: present and future tax selling costs and diversification both within the equity portfolio and between bonds and equity. In contrast to Constantinides (1983) where shorting the box is allowed, the consumption-portfolio problem in our paper is not independent of tax trading concerns.

We have shown that with shorting prohibited, the investor's optimal strategy is similar to the case of one risky stock with the notable exception that tax trading costs can lead to the investor holding an undiversified equity position. With no embedded gains, however, the welfare benefit of being able to trade the two stocks separately as opposed to holding an index fund that combines the two is small.

Allowing short sales leads to dramatically different behavior. The investor can use short selling both as an ex ante and as an ex post tax-efficient mechanism to reduce overall equity exposure and tax-related trading costs. The combination of lower trading costs and the ability to scale back overexposure to equity dominates the loss in diversification within the equity portion of the portfolio. Specifically, we describe a

[11]To compute the effective total stock allocation, we translate the position in the put to an equivalent short position in the stock 2 by multiplying the put position by the put's delta, i.e., its sensitivity to the underlying stock price, and adding it to the stock 1 position.

[12]The welfare benefits of investing in a put option are captured by the wealth benefit ratio calculated in the same fashion as in the two-stock case, namely, the fraction of wealth an age 20 index investor would pay to be indifferent between his current index fund investment or a risky portfolio consisting of stock 1 and a put option written on stock 2.

novel ex ante strategy that an investor can use, the *trading flexibility strategy*, where he shorts one of the stocks even when no stock has an embedded capital gain. Ex post, he can use an imperfect form of shorting the box. Interestingly, this shorting behavior commonly occurs when an otherwise identical but short-sale-constrained investor would hold strictly positive investments in all stocks. This behavior arises because the investor's net tax benefit of selling shares is not monotonic in the trading strategy when shorting is allowed.

Not surprisingly, the tax benefit from shorting depends on the shorting cost that the investor faces. For a large investor, the benefit of shorting outweighs the cost. For small investors who get a low rebate rate on their short sale proceeds, the benefit is reduced. For a high enough shorting cost, a small investor will no longer use the trading flexibility strategy. We have proposed buying puts as an alternative mechanism for utilizing the trading flexibility strategy. The welfare benefit associated with using puts is similar to the one obtained with equity short sales at a low shorting cost, making put-based strategies a viable tax management tool even for smaller investors.

While we have restricted the analysis to two risky stocks and a money market account for tractability, we expect similar results to hold when the investor is allowed to hold more than two stocks. To manage tax trading costs, the investor could use the trading flexibility and imperfect shorting-the-box strategies on highly correlated sets of stocks within the portfolio.

Appendix A. Investor consumption-portfolio problem with two stocks

The mathematical description of the framework described in Section 2 is now presented. Our two risky stock model is an extension of the single-stock setting of Dammon et al. (2001b) that accommodates costly short selling with margin constraints. We consider a discrete-time economy with trading dates $t = 0, \ldots, T$ in which an investor endowed with initial wealth in the assets chooses an optimal consumption and investment policy in the presence of realized capital gain taxation. We assume that the investor lives for at most T periods and faces a positive probability of death each period. The probability that an investor lives up to period t, $0 < t < T$, is given by the survival function $H(t) = \exp(-\sum_{n=0}^{t} \lambda_n)$ where λ_n is the single-period hazard rate for period n where we assume $\lambda_n > 0,\ \forall n$, and $\lambda_T = \infty$. This implies $0 \leqslant H(t) < 1,\ \forall\, 0 \leqslant t < T$. At T, the investor exits the economy, implying $H(T) = 0$. We assume that the investor makes annual decisions starting at age 20 corresponding to $t = 0$ with certain exit from the economy at age 100 implying $T = 80$. The hazard rates λ_n are calibrated to the 1990 U.S. Life Table compiled by the National Center for Health Statistics to compute the survival function $H(t)$ from ages 20 ($t = 0$) to 99 ($t = 79$).

A.1. Security market

The market consists of three assets: a riskless money market account and two risky dividend-paying stocks. The riskless money market account with time t price $S_0(t)$

pays a continuously compounded pre-tax interest rate r. The price dynamics over the time interval Δ_t are

$$\frac{S_0(t+\Delta_t)}{S_0(t)} = \exp(r\Delta_t). \tag{A.1}$$

The two risky stocks with time t ex-dividend prices $S_i(t)$, $i \in \{1,2\}$, pay pre-tax dividends of $\delta_i S_i(t)$ at time $t+\Delta_t$ where δ_i is each stock's dividend yield. The pre-tax ex-dividend stock prices follow binomial processes with price dynamics over the time interval Δ_t given by

$$\frac{S_i(t+\Delta_t)}{S_i(t)} = \exp\left(\left(\mu_i - \frac{1}{2}\sigma_i^2\right)\Delta_t + \sigma_i\sqrt{\Delta_t}\tilde{z}_i\right), \quad i \in \{1,2\}, \tag{A.2}$$

where $\tilde{z}_i$ is a binomial random variable taking the two values -1 and 1 with a joint probability distribution specified as $p(\tilde{z}_1 = 1, \tilde{z}_2 = 1) = p(\tilde{z}_1 = -1, \tilde{z}_2 = -1) = \frac{1}{4}(1+\rho)$ and $p(\tilde{z}_1 = 1, \tilde{z}_2 = -1) = p(\tilde{z}_1 = -1, \tilde{z}_2 = 1) = \frac{1}{4}(1-\rho)$ where ρ is the correlation between the two stock prices. For each stock, the quantity μ_i is the instantaneous capital gain expected growth rate and σ_i is the instantaneous volatility. We also assume that $\Delta_t = 1$ matching the trading interval of the investor in the economy with time units measured in years. A trading strategy from time t to $t+1$ in the money market account and two stocks is given by $(\alpha(t), \theta_1(t), \theta_2(t))$ where $\alpha(t)$ denotes the shares of the money market held and $\theta_i(t)$ denotes the shares of stock i held.

A.2. Capital gain taxation

Realized capital gains and losses are subject to a constant capital gain tax rate of τ_C where we assume that the full proceeds of capital losses can be used. Given that our investor trades at an annual frequency, no distinction is made between long- and short-term capital gains.

The tax basis for computing capital gains or losses is calculated as a weighted-average purchase price for tractability. Let $B_i(t+1)$ denote the nominal tax basis of stock i after trading at time $t+1$. If long stock at date $t+1$, $\theta_i(t+1) > 0$, the stock basis evolves as

$$B_i(t+1) = \begin{cases} S_i(t+1) & \text{if } \theta_i(t) \leqslant 0 \text{ or } \frac{B_i(t)}{S_i(t+1)} > 1, \\ \dfrac{B_i(t)\theta_i(t) + (\theta_i(t+1) - \theta_i(t))^+ S_i(t+1)}{\theta_i(t) + (\theta_i(t+1) - \theta_i(t))^+} & \text{otherwise,} \end{cases} \tag{A.3}$$

where $x^+ \triangleq \max(x, 0)$. If short stock at date $t+1$, $\theta_i(t+1) < 0$, the stock basis evolves as

$$B_i(t+1) = \begin{cases} S_i(t+1) & \text{if } \theta_i(t) \geqslant 0 \text{ or } \frac{B_i(t)}{S_i(t+1)} < 1, \\ \dfrac{B_i(t)|\theta_i(t)| + (|\theta_i(t+1)| - |\theta_i(t)|)^+ S_i(t+1)}{|\theta_i(t)| + (|\theta_i(t+1)| - |\theta_i(t)|)^+} & \text{otherwise.} \end{cases} \tag{A.4}$$

If $\theta_i(t+1)=0$, the basis resets to the current stock price, $B_i(t+1)=S_i(t+1)$. In summary, stock i's basis resets to the current stock price when the position changes sign from the previous period or when the investor is tax-loss selling. For stock i trades that increase the magnitude of the current position, the basis evolves as a share-weighted average of the previous basis and the current share price given by the "otherwise" case in (A.3) and (A.4). Note that this case also covers the scenario when the investor reduces the magnitude of his stock i position without changing its sign, implying an unchanged basis.

Any realized capital gains or losses are subject to capital gain taxation. Several different types of trades trigger realized capital gains or losses. First, tax-loss selling (denoted TL) induces realized capital losses. The tax-loss selling contribution to time t capital gain taxation is denoted as $\Phi_{CG}^{TL}(t)$ and given by

$$\begin{aligned}\Phi_{CG}^{TL}(t) = \tau_C \sum_{i=1}^{2} &(S_i(t) - B_i(t-1))\theta_i(t-1) \\ &\times(\mathbf{1}_{\{\theta_i(t-1)>0,\, B_i(t-1)>S_i(t)\}} + \mathbf{1}_{\{\theta_i(t-1)<0,\, B_i(t-1)<S_i(t)\}}),\end{aligned} \tag{A.5}$$

where the first (second) indicator function denotes tax-loss selling when holding a long (short) position. Second, since the investor is precluded from shorting the box, closing out a time $t-1$ position and investing at time t in the position of opposite sign for that stock triggers a taxable event. Such a trade is denoted CP for "close position." The time t realized capital gain taxes, $\Phi_{CG}^{CP}(t)$, from such a trade are

$$\begin{aligned}\Phi_{CG}^{CP}(t) = \tau_C \sum_{i=1}^{2} &(S_i(t) - B_i(t-1))\theta_i(t-1) \\ &\times(\mathbf{1}_{\{\theta_i(t)\leqslant 0,\, \theta_i(t-1)>0,\, B_i(t-1)\leqslant S_i(t)\}} + \mathbf{1}_{\{\theta_i(t)\geqslant 0,\, \theta_i(t-1)<0,\, B_i(t-1)\geqslant S_i(t)\}}).\end{aligned} \tag{A.6}$$

The first (second) indicator function in (A.6) captures moving from a long (short) position to a short (long) position in stock i when the $t-1$ position contains a capital gain. The capital loss case of closing out a $t-1$ position is captured by (A.5). Finally, when the investor reduces the size of his $t-1$ position at time t denoted RP, capital gain taxes are assessed. Their contribution to time t capital gain taxes, $\Phi_{CG}^{RP}(t)$, is

$$\begin{aligned}\Phi_{CG}^{RP}(t) = \tau_C \sum_{i=1}^{2} &(S_i(t) - B_i(t-1)) \\ &\times(\mathbf{1}_{\{\theta_i(t)>0,\, \theta_i(t-1)\geqslant 0,\, B_i(t-1)\leqslant S_i(t)\}}(\theta_i(t-1) - \theta_i(t))^+ \\ &- \mathbf{1}_{\{\theta_i(t)<0,\, \theta_i(t-1)\leqslant 0,\, B_i(t-1)\geqslant S_i(t)\}}(\theta_i(t) - \theta_i(t-1))^+),\end{aligned} \tag{A.7}$$

where the first (second) indicator function captures reducing a long (short) position in stock i. In summary, the total capital gain taxes paid at time t, $\Phi^{CG}(t)$, consisting of realized capital gains/losses from tax-loss selling (TL), closing positions (CP), and reducing positions (RP), are

$$\Phi_{CG}(t) = \Phi_{CG}^{TL}(t) + \Phi_{CG}^{CP}(t) + \Phi_{CG}^{RP}(t). \tag{A.8}$$

If death occurs at some time t', all capital gain taxes are forgiven implying $\Phi_{CG}(t')=0$.

A.3. Admissible trading strategies

We now define the set of admissible trading strategies when the investor can short stock in a margin account. To model the margin requirements, we assume that there is no difference between initial and maintenance margins. Explicitly incorporating maintenance margins into our optimization problem would lead to the problem being numerically intractable. The quantity $1+\kappa_- \geqslant 1$ denotes the proportional margin requirement on short stock positions satisfied by cash or other securities and $1-\kappa_+ \in [0,1]$ denotes the fraction of the value of a long position that is marginable. From Regulation T on initial margin requirements, this implies $\kappa_- = \kappa_+ = 0.5$. Denoting $\alpha_C(t)$ as the amount of money in cash collateral at time t, short selling bounds this amount from below by

$$\alpha_C(t) \geqslant 1.02 \sum_{i=1}^{2} \theta_i(t)^- S_i(t), \tag{A.9}$$

where $x^- \triangleq \max(-x,0)$ and we have assumed that the required cash collateral is 102% of the value of the short position. The margin requirements impose a restriction that the investor can only borrow using as collateral cash or stock in excess of the required margin. This implies a lower bound on the dollar amount invested in the money market account,

$$\alpha(t)\, S_0(t) \geqslant -\left(\alpha_C - (1+\kappa_-)\sum_{i=1}^{2} \theta_i(t)^- S_i(t) \right.$$
$$\left. + (1-\kappa_+)\sum_{i=1}^{2} \theta_i(t)^+ S_i(t)\right). \tag{A.10}$$

The term $\alpha_C - (1+\kappa_-)\sum_{i=1}^{2} \theta_i(t)^- S_i(t)$ is the additional margin needed in other securities after taking into account the cash collateral when short selling. The term $(1-\kappa_+)\sum_{i=1}^{2} \theta_i(t)^+ S_i(t)$ is the marginable value of all long stock positions.

Given the money market account's return always dominates the return received on cash collateral, Eq. (A.9) always holds with equality, so cash collateral is completely determined by the size of the investor's short position. The bound on the money market account given by (A.10) can also be simplified by appealing to the time t self-financing condition. Let $W(t)$ denote the time t wealth before portfolio rebalancing and before any capital gain taxes are paid, but after dividend and interest taxes are paid. Given that no resources are lost when rebalancing the portfolio at time t,

$$W(t) = \alpha_C + \alpha(t)\, S_0(t) + \sum_{i=1}^{2} \theta_i(t) S_i(t) + C(t) + \Phi_{CG}(t), \tag{A.11}$$

where $C(t) > 0$ is the time t consumption. Solving the above expression for money market holdings and using $\sum_{i=1}^{2} \theta_i(t) S_i(t) = \sum_{i=1}^{2} \theta_i(t)^+ S_i(t) - \sum_{i=1}^{2} \theta_i(t)^- S_i(t)$, Eq. (A.10) simplifies to

$$W(t) - C(t) - \Phi_{CG}(t) \geqslant \kappa_- \sum_{i=1}^{2} \theta_i(t)^- S_i(t) + \kappa_+ \sum_{i=1}^{2} \theta_i(t)^+ S_i(t). \tag{A.12}$$

Dividend and interest income are taxed as ordinary income at the constant rate τ_D. Each period, the investor pays taxes on his dividend and interest income. The total taxes paid on this income at time t is

$$\begin{aligned}\Phi_D(t) = {} & \tau_D \alpha(t-1)\, S_0(t-1)\,(\exp(r) - 1) \\ & + \tau_D(1.02) \sum_{i=1}^{2} \theta_i(t-1)^- S_i(t-1)(\exp(r_c) - 1) \\ & + \tau_D \sum_{i=1}^{2} \theta_i(t-1) S_i(t-1) \delta_i. \end{aligned} \tag{A.13}$$

The first term in (A.13) is the income tax due to money market interest. The second term captures the taxes on any interest on cash collateral while the last term accounts for the dividend taxes. If the investor dies at time t, interest and dividend taxes are still paid.

Again letting $W(t+1)$ denote the time $t+1$ wealth before portfolio rebalancing and any capital gain taxes are paid, but after dividend and interest taxes are paid, $W(t+1)$ is given by

$$\begin{aligned} W(t+1) = {} & \alpha(t) S_0(t)((1-\tau_D)\exp(r) + \tau_D) \\ & + (1.02) \sum_{i=1}^{2} \theta_i(t)^- S_i(t)((1-\tau_D)\exp(r_c) + \tau_D) \\ & + \sum_{i=1}^{2} \theta_i(t)(S_i(t+1) + S_i(t)\delta_i(1-\tau_D)), \end{aligned} \tag{A.14}$$

where (A.13) has been substituted. Substituting (A.11) into (A.14) gives the dynamic after-tax wealth evolution of the investor,

$$\begin{aligned} W(t+1) = {} & \left(W(t) - \sum_{i=1}^{2} \theta_i(t) S_i(t) - C(t) - \Phi^{CG}(t) \right. \\ & \left. - (1.02) \sum_{i=1}^{2} \theta_i(t)^- S_i(t) \right) ((1-\tau_D)\exp(r) + \tau_D) \\ & + (1.02) \sum_{i=1}^{2} \theta_i(t)^- S_i(t)((1-\tau_D)\exp(r_c) + \tau_D) \\ & + \sum_{i=1}^{2} \theta_i(t)(S_i(t+1) + S_i(t)\delta_i(1-\tau_D)). \end{aligned} \tag{A.15}$$

An *admissible* trading strategy is a consumption and security trading policy $(C, \alpha, \theta_1, \theta_2)$ such that for all t, $C(t) \geqslant 0$, $W(t) \geqslant 0$, and Eqs. (A.9)–(A.15) are satisfied. The set of admissible trading strategies is denoted $\mathscr{A}$.

A.4. Investor's objective

The investor's objective is to maximize his discounted expected utility of real lifetime consumption and final-period wealth at the time of death by choosing an admissible trading strategy given an initial endowment. If death occurs on date t, the investor's assets totalling $W(t)$ are liquidated and used to purchase a perpetuity that pays to his heirs a constant real after-tax cash flow of $R^* W(t)$ each period starting on date $t+1$. The quantity R^* is the one-period after-tax real riskless interest rate computed using simple compounding. In terms of the instantaneous nominal riskless money market rate r and the instantaneous inflation rate i, R^* is defined by

$$R^* = ((1-\tau_D)\exp(r) + \tau_D)\exp(-\mathrm{i}) - 1.$$

Under the assumption that the investor and his heirs have identical preferences of the constant relative risk aversion (CRRA) form with a coefficient of relative risk aversion of γ and a common time preference parameter β, the investor's optimization problem is given by

$$\max_{(C,\alpha,\theta_1,\theta_2)\in\mathscr{A}} E\left[\sum_{t=0}^{T} \beta^t \left\{ \frac{H(t)}{1-\gamma}(\exp(-\mathrm{i}t)C(t))^{1-\gamma} + \frac{H(t-1)-H(t)}{1-\gamma} \sum_{s=t+1}^{\infty} \beta^{s-t}(\exp(-\mathrm{i}t)R^* W(t))^{1-\gamma} \right\}\right] \tag{A.16}$$

subject to (A.9)–(A.15) where β is the time-preference parameter. The objective function captures the expected utility of future real consumption as well as the bequest motive to the investor's heirs. In Section 4, we consider one scenario where the investor has no bequest motive. In this setting, the last term is eliminated from the above objective function.

Since $\sum_{s=t+1}^{\infty} \beta^{s-t} = \frac{\beta}{1-\beta}$, the investor's objective function simplifies, leading to the optimization problem

$$\max_{(C,\alpha,\theta_1,\theta_2)\in\mathscr{A}} E\left[\sum_{t=0}^{T} \beta^t \left\{ \frac{H(t)}{1-\gamma}(\exp(-\mathrm{i}t)C(t))^{1-\gamma} + \frac{H(t-1)-H(t)}{1-\gamma} \frac{\beta}{1-\beta}(\exp(-\mathrm{i}t)R^* W(t))^{1-\gamma} \right\}\right] \tag{A.17}$$

subject to (A.9) through (A.15).

A.5. A convenient change of variables

As in a no-tax portfolio choice problem with CRRA preferences, the optimization problem (A.17) is homogeneous in wealth, and thus independent of the investor's initial wealth. To show that wealth is not needed as a state variable when solving (A.17), we express the optimization problem's controls as being proportional to time t wealth $W(t)$ before trade has occurred but after dividend and interest taxes have been paid. We define

$$\underline{\pi}_i(t) \triangleq \frac{S_i(t)\theta_i(t-1)}{W(t)}, \qquad \overline{\pi}_i(t) \triangleq \frac{S_i(t)\theta_i(t)}{W(t)}, \tag{A.18}$$

where $\underline{\pi}_i(t)$ and $\overline{\pi}_i(t)$ are the proportions of stock i owned entering and leaving period t, with respect to time t wealth $W(t)$. Since we allow for short sales in our formulation, it is possible that financial wealth $W(t)$ no longer remains positive implying that it is not possible to define portfolio weights. However, the investor will never choose a trading strategy that leads to a non-positive wealth at any time given our utility function choice, the bequest motive, and the positive probability of death over each period.

Using (A.18), it is useful to express the basis $B_i(t)$ as a basis-price ratio $b_i(t+1) \triangleq B_i(t)/S_i(t+1)$. Using (A.3), if long stock at date t, $\overline{\pi}_i(t) > 0$, the basis-price ratio evolves as

$$b_i(t+1) = \begin{cases} \dfrac{S_i(t)}{S_i(t+1)} & \text{if } \underline{\pi}_i(t) \leqslant 0 \text{ or } b_i(t) > 1, \\ \dfrac{b_i(t)\underline{\pi}_i(t) + (\overline{\pi}_i(t) - \underline{\pi}_i(t))^+}{\frac{S_i(t+1)}{S_i(t)}(\underline{\pi}_i(t) + (\overline{\pi}_i(t) - \underline{\pi}_i(t))^+)} & \text{otherwise.} \end{cases} \tag{A.19}$$

If short stock at date t, $\overline{\pi}_i(t) < 0$, the stock basis-price ratio evolves as

$$b_i(t+1) = \begin{cases} \dfrac{S_i(t)}{S_i(t+1)} & \text{if } \underline{\pi}_i(t) \geqslant 0 \text{ or } b_i(t) < 1, \\ \dfrac{b_i(t)|\underline{\pi}_i(t)| + (|\overline{\pi}_i(t)| - |\underline{\pi}_i(t)|)^+}{\frac{S_i(t+1)}{S_i(t)}(|\underline{\pi}_i(t)| + (|\overline{\pi}_i(t)| - |\underline{\pi}_i(t)|)^+)} & \text{otherwise.} \end{cases} \tag{A.20}$$

If $\overline{\pi}_i(t) = 0$, the basis-price ratio $b_i(t+1)$ resets to the ratio of the time t and $t+1$ stock price, $b_i(t+1) = S_i(t)/S_i(t+1)$. The basis-price ratio at time $t+1$ can be expressed as a function of the capital gain of stock i over one period $S_i(t+1)/S_i(t)$, the previous period's basis-price ratio $b_i(t)$, and the equity proportions $\overline{\pi}_i(t)$ and $\underline{\pi}_i(t)$.

Using the equity proportions and the basis-price ratios, the total capital gain taxes paid at time t, $\Phi_{CG}(t) = \Phi_{CG}^{TL}(t) + \Phi_{CG}^{CP}(t) + \Phi_{CG}^{RP}(t)$, can be written proportional to $W(t)$, $\Phi_{CG}(t) = W(t)\,\phi_{CG}(t)$ where $\phi_{CG}(t)$ is independent of $W(t)$, since each component of the total capital gain taxes paid can be written proportional to $W(t)$:

$$\Phi_{CG}^{TL}(t) = \tau_C W(t) \sum_{i=1}^{2} (1 - b_i(t))\underline{\pi}_i(t) \times (\mathbf{1}_{\{\underline{\pi}_i(t) > 0,\, b_i(t) > 1\}} + \mathbf{1}_{\{\underline{\pi}_i(t) < 0,\, b_i(t) < 1\}}), \tag{A.21}$$

$$\Phi^{CP}_{CG}(t) = \tau_C W(t) \sum_{i=1}^{2} (1 - b_i(t)) \underline{\pi}_i(t) \times (\mathbf{1}_{\{\overline{\pi}_i(t) \leqslant 0,\, \underline{\pi}_i(t) > 0,\, b_i(t) \leqslant 1\}} + \mathbf{1}_{\{\overline{\pi}_i(t) \geqslant 0,\, \underline{\pi}_i(t) < 0,\, b_i(t) \geqslant 1\}}), \tag{A.22}$$

$$\Phi^{RP}_{CG}(t) = \tau_C W(t) \sum_{i=1}^{2} (1 - b_i(t))(\mathbf{1}_{\{\theta_i(t) > 0,\, \underline{\pi}_i(t) \geqslant 0,\, b_i(t) \leqslant 1\}} (\underline{\pi}_i(t) - \overline{\pi}_i(t))^+ - \mathbf{1}_{\{\overline{\pi}_i(t) < 0,\, \underline{\pi}_i(t) \leqslant 0,\, b_i(t) \geqslant 1\}} (\overline{\pi}_i(t) - \underline{\pi}_i(t))^+). \tag{A.23}$$

Given that no resources are lost when portfolio rebalancing and paying taxes, Eq. (A.11) implies that the money market investment $\alpha(t)\,S_0(t)$ can be written proportional to $W(t)$:

$$\alpha(t) S_0(t) = W(t) \left(1 - (1.02) \sum_{i=1}^{2} \overline{\pi}_i(t)^- - \sum_{i=1}^{2} \overline{\pi}_i(t) - c(t) - \phi_{CG}(t) \right), \tag{A.24}$$

where $c(t) \triangleq C(t)/W(t)$. Using (A.24), the margin requirement constraint given by (A.10) can be written independent of wealth:

$$1 - c(t) - \phi_{CG}(t) \geqslant \kappa_- \sum_{i=1}^{2} \overline{\pi}_i(t)^- + \kappa_+ \sum_{i=1}^{2} \overline{\pi}_i(t)^+. \tag{A.25}$$

The wealth evolution (A.15) can also be written proportional to $W(t)$ implying

$$\frac{W(t+1)}{W(t)} = \frac{1}{1 - \sum_{i=1}^{2} \underline{\pi}_i(t+1)} \Bigg[((1 - \tau_D) \exp(r) + \tau_D) \times \left(1 - \sum_{i=1}^{2} \overline{\pi}_i(t) - c(t) - \phi^{CG}(t) - (1.02) \sum_{i=1}^{2} \overline{\pi}_i(t)^- \right) + (1.02) \sum_{i=1}^{2} \overline{\pi}_i(t)^- ((1 - \tau_D) \exp(r_c) + \tau_D) + \sum_{i=1}^{2} \overline{\pi}_i(t) \delta_i (1 - \tau_D) \Bigg]. \tag{A.26}$$

Additionally, the stock proportion evolution is given by

$$\underline{\pi}_i(t+1) = \frac{S_i(t+1)/S_i(t)\overline{\pi}_i(t)}{W(t+1)/W(t)}, \tag{A.27}$$

a quantity that is independent of time t wealth. This evolution is needed in the dynamic programming formulation of the investor's problem where $\underline{\pi}_i$ is a state variable and $\overline{\pi}_i$ is a control variable.

Using the principle of dynamic programming and substituting out $W(t)$, the Bellman equation for the investor's optimization problem (A.17) is given by

$$V(t, \underline{\pi}_1(t), \underline{\pi}_2(t), b_1(t), b_2(t)) = \max_{c(t),\overline{\pi}_1(t),\overline{\pi}_2(t)} \frac{e^{-\lambda_t} c(t)^{1-\gamma}}{1-\gamma} + \frac{(1-e^{-\lambda_t})\beta(R^*)^{1-\gamma}}{(1-\beta)(1-\gamma)} + e^{-\lambda_t}\beta E_t\left[\left(\frac{e^{-i} W(t+1)}{W(t)}\right)^{(1-\gamma)} \times V(t+1, \underline{\pi}_1(t+1), \underline{\pi}_2(t+1), b_1(t+1), b_2(t+1))\right], \tag{A.28}$$

for $t = 0, 1, \ldots, T-1$ subject to the wealth evolution (A.26), the margin requirement constraint (A.25), and the stock proportion dynamics (A.27).

Appendix B. Investor consumption-portfolio problem with one stock and a put option

The modifications to the portfolio problem to incorporate investment in a put option as described in Section 5 are now presented. The setup is the same as described in Appendix A with the following notable exceptions.

B.1. Security market

The market consists of four assets every trading period: a riskless money market account, two risky dividend-paying stocks, and a European put option that expires in one trading period (one year in our analysis). The price dynamics of the riskless money market account and the two risky stocks are given by (A.1) and (A.2), respectively.

The investor trades in the riskless money market, stock 1, and the put option. He does not trade in stock 2; rather, the stock 2's only role is to be the put option's underlying asset. The put option's strike price is the current price of stock 2, $S_2(t)$. The put option's price $p(t)$ is presented below after discussing capital gain taxation.

B.2. Capital gain taxation

As in the two-stock portfolio problem, realized capital gains and losses are subject to a constant capital gain tax rate of τ_C where we assume that the full proceeds of capital losses can be used.

The calculation of the tax basis for the investment in stock 1 is simplified in that we assume the investor cannot short any security. The tax basis for stock 1 for computing capital gains or losses is calculated as a weighted-average purchase price for tractability. Let $B_1(t+1)$ denote the nominal tax basis of stock 1 after trading at time $t+1$. The stock basis evolves as

$$B_1(t+1) = \begin{cases} S_1(t+1) & \text{if } \theta_1(t) = 0 \text{ or } \dfrac{B_1(t)}{S_1(t+1)} > 1, \\ \dfrac{B_1(t)\theta_1(t) + (\theta_1(t+1) - \theta_1(t))^+ S_1(t+1)}{\theta_1(t) + (\theta_1(t+1) - \theta_1(t))^+} & \text{otherwise,} \end{cases} \tag{B.1}$$

where $x^+ \triangleq \max(x, 0)$. If $\theta_1(t+1) = 0$, the basis resets to the current stock price, $B_1(t+1) = S_1(t+1)$.

The put option is treated as a cash-settled contract, so capital gain taxes are assessed on the realized gain or loss on the put investment. By assuming that the put option expires after one trading period, computing the capital gain on the put position is simplified since the tax basis is always the purchase price of the put. If $\theta_p(t-1)$ is the number of units of the put owned from time $t-1$ to t, the dollar capital gain or loss from the put position at time t is $\theta_p(t-1)(\max\{S_2(t-1) - S_2(t), 0\} - p(t-1))$ since the strike price of the put is $S_2(t-1)$.

Any realized capital gains or losses are subject to capital gain taxation. The capital gain taxes $\Phi_{CG}(t)$ at time t are

$$\begin{aligned} \Phi_{CG}(t) = \tau_C((S_1(t) - B_1(t-1))\theta_1(t-1)\mathbf{1}_{\{B_1(t-1) > S_1(t)\}} \\ + (S_1(t) - B_1(t-1))\mathbf{1}_{\{B_1(t-1) \leqslant S_1(t)\}}(\theta_1(t-1) - \theta_1(t))^+ \\ + \theta_p(t-1)(\max\{S_2(t-1) - S_2(t), 0\} - p(t-1))), \end{aligned} \tag{B.2}$$

where the first line calculates capital losses from tax-loss selling stock 1, the second line calculates taxes from selling stock with a capital gain, and the last line calculates capital gains or losses from the put option holdings. If death occurs at some time t', all capital gain taxes are forgiven implying $\Phi_{CG}(t') = 0$.

B.3. Valuation of the put option

The put option is priced by no arbitrage, but this is complicated by the taxes in our economy. When the participants in the option market are differentially taxed on the option's cash flows, it is not immediately clear who is the marginal pricer of the option given different valuations can arise for different tax rates. These pricing with differential taxation issues are discussed in detail in Schaefer (1982), Dybvig and Ross (1986), Dammon and Green (1987), and Ross (1987). For our portfolio choice exercise with a put option, we assume that the price of the option is set by an investor who faces realized capital gain taxation and can costlessly short.

This price is determined by using standard replication arguments on the after-tax payoffs of the option. Let α and θ be the respective portfolio positions in the riskless money market and stock 2 used to replicate the put option's after-tax payoff. If stock 2 increases (denoted $S_2(t+1, U)$) or decreases (denoted $S_2(t+1, D)$) next period, the replicating portfolio's after-tax payoff must match the put's after-tax payoff, implying

$$
\begin{aligned}
&\tau_c p(t) \\
&\quad = ((1-\tau_c)S_2(t+1,\ U) + \tau_c S_2(t) + (1-\tau_D)S_2(t)\delta_2)\theta \\
&\qquad + \ S_0(t)((1-\tau_D)\exp(r) + \tau_D)\alpha,
\end{aligned} \tag{B.3}
$$

$$
\begin{aligned}
&(1-\tau_c)(S_2(t) - S_2(t+1,\ D)) + \tau_c p(t) \\
&\quad = ((1-\tau_c)S_2(t+1,\ D) + \tau_c S_2(t) + (1-\tau_D)S_2(t)\delta_2)\theta \\
&\qquad + \ S_0(t)((1-\tau_D)\exp(r) + \tau_D)\alpha.
\end{aligned} \tag{B.4}
$$

By no arbitrage, the price of the put $p(t)$ must equal the initial value of the replicating portfolio,

$$p(t) = S_2(t)\theta + S_0(t)\alpha. \tag{B.5}$$

Given the parameters used for the price system, the put is only in the money if stock 2 decreases over the next period. Solving (B.3)–(B.5) yields the put price $p(t)$ as well as the replicating strategy.

This valuation of the option actually leads to a higher price than if the pre-capital gain tax cash flows are used to value the option by setting $\tau_c = 0$. In contrast to a valuation with no capital gain tax, the put option has a positive after-tax payoff of $\tau_c p(t)$ if the option expires out of the money. This payoff is the capital loss write-off from the put option. By using this higher put price, our calculated wealth benefits of trading the put are understated as compared to a put option valued using pre-tax cash flows.

B.4. Admissible trading strategies

We now define the set of admissible trading strategies when the investor can invest in stock 1, the put option, and the riskless money market account. Again, we assume that the investor is prohibited from shorting any security. Dividend and interest income are taxed as ordinary income at the constant rate τ_D. Each period, the investor pays taxes on his dividend and interest income. The total taxes paid on this income at time t is

$$\Phi_D(t) = \tau_D\alpha(t-1)S_0(t-1)(\exp(r)-1) + \tau_D\theta_1(t-1)S_1(t-1)\delta_1. \tag{B.6}$$

If the investor dies at time t, interest and dividend taxes are still paid.

Letting $W(t+1)$ denote the time $t+1$ wealth before portfolio rebalancing and any capital gain taxes are paid, but after dividend and interest taxes are paid, $W(t+1)$ is given by

$$
\begin{aligned}
W(t+1) = {} & \alpha(t)S_0(t)((1-\tau_D)\exp(r) + \tau_D) \\
& + \ \theta_1(t)(S_1(t+1) + S_1(t)\delta_1(1-\tau_D)) \\
& + \ \theta_p(t)\max\{S_2(t) - S_2(t+1),\ 0\},
\end{aligned} \tag{B.7}
$$

where (B.6) has been substituted. Given that no resources are lost when rebalancing the portfolio at time t, $W(t)$ is given by

$$W(t) = \alpha(t)S_0(t) + \theta_1(t)S_1(t) + \theta_p(t)p(t) + C(t) + \Phi_{CG}(t), \tag{B.8}$$

where $C(t) > 0$ is the time t consumption.

Substituting (B.8) into (B.7) gives the dynamic after-tax wealth evolution of the investor,

$$\begin{aligned} W(t+1) &= (W(t) - \theta_1(t)S_1(t) - \theta_p(t)p(t) - C(t) - \Phi^{CG}(t))((1-\tau_D)\exp(r) + \tau_D) \\ &\quad + \theta_1(t)(S_1(t+1) + S_1(t)\delta_1(1-\tau_D)) + \theta_p(t)\max\{S_2(t) - S_2(t+1), 0\}. \end{aligned} \tag{B.9}$$

An *admissible* trading strategy is a consumption and a security trading policy $(C, \alpha, \theta_1, \theta_p)$ such that for all t, $C(t) \geqslant 0$, $W(t) \geqslant 0$, $\alpha(t) \geqslant 0$, $\theta_1(t) \geqslant 0$, $\theta_p(t) \geqslant 0$, and (B.9) is satisfied. The set of admissible trading strategies is denoted $\mathscr{A}$.

B.5. Investor's objective

The investor's optimization problem is given by (A.17) where the investment in stock 2 is replaced by the investment in the put option. As in Appendix A, the optimization problem can be written independent of wealth leading to a dynamic programming problem with one less state variable given that the capital gain tax basis of the put option is no longer needed since the contract expires every period.

Appendix C. Numerical setup and optimization

The value function (A.28) is solved numerically by backward recursion starting at time T. The optimization problem is 80 years long with annual time steps, four state variables, and three choice variables. The state variables are the beginning-of-period wealth allocation in each risky asset and the beginning-of-period tax basis for each risky asset. In the notation introduced earlier, the state variables correspond to $(\underline{\pi}_1(t), \underline{\pi}_2(t), b_1(t), b_2(t))$.

On each point on the grid of state variables we solve an optimization problem for three choice variables: the optimal consumption level and the wealth allocation in each risky asset $(c(t), \overline{\pi}_1(t), \overline{\pi}_2(t))$. We limit the search over a set of values for which the margin constraint is satisfied, and we search this set using a grid, in a recursive manner (coarse grid to fine grid). Specifically, for each period and each point in the state variable grid, an initial coarse grid in consumption levels and wealth allocation in equity levels is searched for the choice variable values that maximize the value function. Subsequently the search grid is refined around the choice variable points with the highest level of the value function. The procedure is repeated until the size of the dimensions of the grid are smaller than some required level of accuracy. The accuracy used in the optimizations reported in the paper was 10 basis points in the optimal wealth allocations in equity and 1 basis point in the consumption level. The state variables for each period are discretized with a discretization step of 5% both for the tax basis and the wealth allocation in each asset. Due to the discrete nature of the state variable grid, and the arbitrary positions taken when exploring the set of allowed values in the choice variables, we interpolate the value function on the state variable grid. The interpolation algorithm we use relies on first scaling the value function by the appropriate wealth-related factor, and on linear interpolation of the resulting deflated value function.

C.1. Numerical accuracy

Our numerical scheme potentially suffers from inaccuracies due to the discretization of the state variable grid, and the subsequent interpolation algorithm. To estimate the level of accuracy we use grids with different step sizes for the state variable and checked the calculated results for consistency. We vary the grid step sizes in both the basis and initial asset allocation directions between 2.5% and 10%. The results are uniformly consistent with those reported in the paper, with the biggest differences across different grid sizes in wealth benefits being 0.2%, and in optimal asset allocations 2%.

C.2. Parallelization

When solving a dynamic programming problem with many state variables, one runs into the well-known curse of dimensionality. For the problem we consider, it would take more than two weeks on a single 2 GHz computer to compute the optimal allocations for one set of parameters. To partially overcome the curse, we have parallelized many of the computations. Specifically, each optimization problem that is solved on the grid for each time slice is independent of all other maximization problems on the same time slice. To parallelize the computation, we have divided the grid for each time slice into subgrids which are allocated to a specific computer in a cluster. Once all computers have solved the optimization problem for their subgrid, they communicate the value function to all other computers in the cluster and the optimization problems for all the nodes in the grid of the previous time slice is undertaken. This parallelization procedure is very general and can be successfully applied to other dynamic programming problems.

We used two different Cray machines, available at the Texas Advanced Center for Computing (TACC). The first machine had 88 nodes operating at 300Mhz each, but we were never allocated more than 32 nodes at a time. The second machine had 270 nodes operating at 300 MHz each, and we were allocated either 64 or 128 nodes. In addition, at TACC, we had access to an IBM Power4 system with 224 processors operating at 1.3 GHz each (we were allowed to use up to 64 processors at a time), and an IBM Pentium III Linux cluster with 64 processors operating at 1 GHz each (we were allowed to use up to 32 processors at a time). Other computing resources used included two computer clusters at the McCombs Business School with individual machine speeds from 300 to 400 MHz. By parallelizing the dynamic programming problem, we were able to obtain computational speeds that corresponded approximately to an 83 GHz stand-alone machine.

References

Akian, M., Menaldi, J.L., Sulem, A., 1996. On an investment-consumption model with transaction costs. SIAM Journal of Control and Optimization 34 (1), 329–364.

Auerbach, A., Siegel, J., May 2000. Capital gains realizations of the rich and sophisticated. American Economic Review 90, 276–282.

Balcer, Y., Judd, K., 1987. Effects of capital gains taxation on life-cycle investment and portfolio management. Journal of Finance 42, 743–761.

Balduzzi, P., Lynch, A., 1999. Transaction costs and predictability: some utility cost calculations. Journal of Financial Economics 52, 47–78.

Balduzzi, P., Lynch, A., 2000. Predictability and transaction costs: the impact on rebalancing rules and behavior. Journal of Finance 55, 2285–2310.

Cadenillas, A., Pliska, S.R., February 1999. Optimal trading of a security when there are taxes and transaction costs. Finance and Stochastics 3 (2), 137–165.

Constantinides, G., 1983. Capital market equilibrium with personal tax. Econometrica 51, 611–636.

Constantinides, G., 1984. Optimal stock trading with personal taxes: implications for prices and the abnormal January returns. Journal of Financial Economics 13, 65–89.

Constantinides, G., 1986. Capital market equilibrium with transaction costs. Journal of Political Economy 94, 842–862.

Constantinides, G., Scholes, M.S., 1980. Optimal liquidation of assets in the presence of personal taxes: implications for asset pricing. Journal of Finance 35, 439–449.

Dammon, R., Green, R., 1987. Tax arbitrage and the existence of equilibrium prices for financial assets. Journal of Finance 42, 1143–1166.

Dammon, R., Spatt, C., Zhang, H., November 2001a. Diversification and Capital Gains Taxes with Multiple Risky Assets. Carnegie Mellon University.

Dammon, R., Spatt, C., Zhang, H., 2001b. Optimal consumption and investment with capital gains tax. Review of Financial Studies 14 (3), 583–616.

Dammon, R., Spatt, C., Zhang, H., 2004. Optimal asset location and allocation with taxable and tax-deferred investing. Journal of Finance 59 (3), 999–1037.

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

Davis, M., Panas, V., Zariphopoulou, T., 1993. European option pricing with transaction costs. SIAM Journal of Control and Optimization 31, 470–493.

DeMiguel, A.-V., Uppal, R., 2003. Portfolio Investment with the Exact Tax Basis via Nonlinear Programming. London Business School.

Duffie, D., Gârleanu, N., Pedersen, L., 2002. Securities lending, shorting, and pricing. Journal of Financial Economics 66, 307–339.

Dybvig, P., Koo, H.K., 1996. Investment with taxes. Washington University.

Dybvig, P., Ross, S., 1986. Tax clienteles and asset pricing. Journal of Finance 41, 751–762.

Elton, E.J., Gruber, M.J., 1978. Taxes and portfolio composition. Journal of Financial Economics 6, 399–410.

Fortune, P., September/October 2000. Margin requirements, margin loans, and margin rates: practice and principles. New England Economic Review 23, 19–44.

Gallmeyer, M., Srivastava, S., 2003. Arbitrage and the Tax Code. Carnegie Mellon University.

Garlappi, L., Naik, L.V., Slive, J., 2001. Portfolio Selection with Multiple Assets and Capital Gains Taxes. University of British Columbia, New York.

Geczy, C.C., Musto, D.K., Reed, A.V., 2002. Stocks are special too: an analysis of the equity lending market. Journal of Financial Economics 66 (2–3), 241–269.

Huang, J., 2001. Taxable or tax-deferred account? Portfolio Decisions with Multiple Investment Goals. MIT Sloan School of Management.

Hur, S.-K., 2001. Optimal Portfolio Selection with Personal Tax. University of Chicago, New York.

Jouini, E., Koehl, P.-F., Touzi, N., May 2000. Optimal investment with taxes: an existence result. Journal of Mathematical Economics 33 (4), 373–388.

Leland, H., 2001. Optimal Portfolio Management with Transaction Costs and Capital Gains Taxes. University of California-Berkeley.

Poterba, J.M., 2001. Taxation and portfolio structure: issues and implications. NBER Working Paper 8223.

Poterba, J.M., Shoven, J.B., 2002. Exchange traded funds: a new investment option for taxable investors. NBER Working Paper 8781.

Ross, S., 1987. Arbitrage and martingales with taxation. Journal of Political Economy 95, 371–393.

Schaefer, S.M., 1982. Taxes and security market equilibrium. In: Sharpe, W., Cootner, C. (Eds.), Financial Economics: Essays in Honor of Paul H. Cootner. Prentice-Hall, Englewood Cliffs, NJ, pp. 159–178.

Shreve, S., Soner, H.M., 1994. Optimal investment and consumption with transaction costs. Annals of Applied Probability 4 (3), 609–692.