

Portfolio Choice with Capital Gain Taxation and the Limited Use of Losses *

Paul Ehling[†]
BI Norwegian Business School

Michael Gallmeyer[‡]
McIntire School of Commerce
University of Virginia

Sanjay Srivastava[§]
OS Financial Trading Systems

Stathis Tompaidis[¶]
McCombs School of Business
University of Texas at Austin

Chunyu Yang^{||}
BI Norwegian Business School

Current Draft: December 2014

*We thank Wolfgang Buehler, Victor DeMiguel, Pascal François, Lorenzo Garlappi, Bruce Grundy, Christian Heyerdahl-Larsen, Philipp Illeditsch, Urban Jermann, Spencer Martin, Jeffery Pontiff, Rémy Praz, Neal Stoughton, Neng Wang, Alan White, Avi Wohl, Amir Yaron, Harold Zhang, and seminar participants at Australia National University, the University of Melbourne, the University of New South Wales, the Wharton School of the University of Pennsylvania, the Young Scholars Nordic Finance Workshop, the University of Piraeus, IDC Herzliya, and the Office of Financial Research. An earlier version of the paper that only incorporated a single stock analysis was presented, among other places, at the Western Finance Association meeting, the European Finance Association Meeting, and the UBC Summer Conference. We acknowledge funding support from the Research Council of Norway (Ehling), the Centre for Asset Pricing Research (CAPR) at BI (Ehling and Yang), and the DeMong-Pettit Research Fund at the McIntire School of Commerce (Gallmeyer). We also thank the Texas Advanced Computing Center and the Norwegian Metacenter for Computational Science (Notur) for providing computing resources.

[†]paul.ehling@bi.no

[‡]mgallmeyer@virginia.edu

[§]srivastavafts@gmail.com

[¶]stathis.tompaidis@mcombs.utexas.edu

^{||}chunyu.yang@bi.no

Abstract

Portfolio Choice with Capital Gain Taxation and the Limited Use of Losses

We study portfolio choice with multiple stocks and capital gain taxation assuming that capital losses can only be used to offset current or future realized capital gains. We show through backtesting, using the time series and empirical distribution of the S&P 500 Index, that on average optimal equity holdings are over an extended period of time significantly lower compared to the case typically studied in the literature where the use of capital losses is unrestricted. Using Value and Growth and Small and Large portfolios, the backtests show that allocations with multiple stocks remain persistently under-diversified.

Keywords: portfolio choice, capital gain taxation, limited use of capital losses, carry-over losses

JEL Classification: G11, H20

1 Introduction

According to the tax code, realized capital losses on equity can only be used to offset current or future realized capital gains. However, to simplify the dimensionality of the numerical problem studies of portfolio choice with capital gain taxation typically assume that the use of capital losses is unrestricted. The goal of this paper is to understand how assumptions on the use of capital losses drive portfolio choice with capital gain taxation. In this regard, using the time series and empirical distribution of the S&P 500 Index, we show that optimal equity holdings with the *limited use of losses* (LUL) remain over an extended period of time significantly lower than the optimal equity holdings with the *full use of losses* (FUL). More specifically, we find that on average it takes 30 years before the optimal equity holdings equate.

Due to the dimensionality and singularity in the dynamics of the portfolio choice problem with capital gain taxes, in the extant academic literature, it is commonly assumed that the use of capital losses is unrestricted. If capital losses are larger than capital gains in a period, the investor receives a tax rebate that cushions the downside of holding equity. This simplification is helpful as then an unused carry-over loss in the portfolio can never occur; hence, it does not need to be tracked over time. While it is convenient to reduce the number of state variables, the simplification comes with a cost. Namely, tax rebates boost the demand for equity relative to the LUL case. What is surprising is the magnitude of the difference and that an FUL investor holds even more equity than an untaxed investor (NCGT) when the portfolio contains no capital gain. We also document that even when both the FUL and the LUL investor are locked-in with their equity holdings, i.e., the tax-induced trading cost exceeds the cost from holding equity above the untaxed benchmark, we see that expected tax rebates can elevate the equity-to-wealth ratio of the FUL investor relative to the LUL investor. Only when the lock-in is so severe that expected rebates are insignificant do we see that the equity-to-wealth ratios of the FUL and LUL investors equate and move in lockstep.

Before we further delve into the main findings of the paper, we illustrate the impact of alternative assumptions on the tax treatment of capital losses in a simple two-period portfolio choice problem with one stock and a bond. Consider that up (down) moves in a binomial tree denote stock price increases (decreases). There are two trading dates and a final date where the portfolio is liquidated. The investor maximizes after-tax final period wealth with constant relative risk aversion utility and an initial endowment of \$100 with no embedded capital gains or losses. Additional details are provided

in Section 3. Figure 1 summarizes the optimal portfolio choice expressed as an equity-to-wealth ratio and capital gain taxes paid through the binomial tree under the LUL- and FUL-based capital gain tax systems and when no capital gain taxes are paid (NCGT).

Figure 1: **Motivating Example**

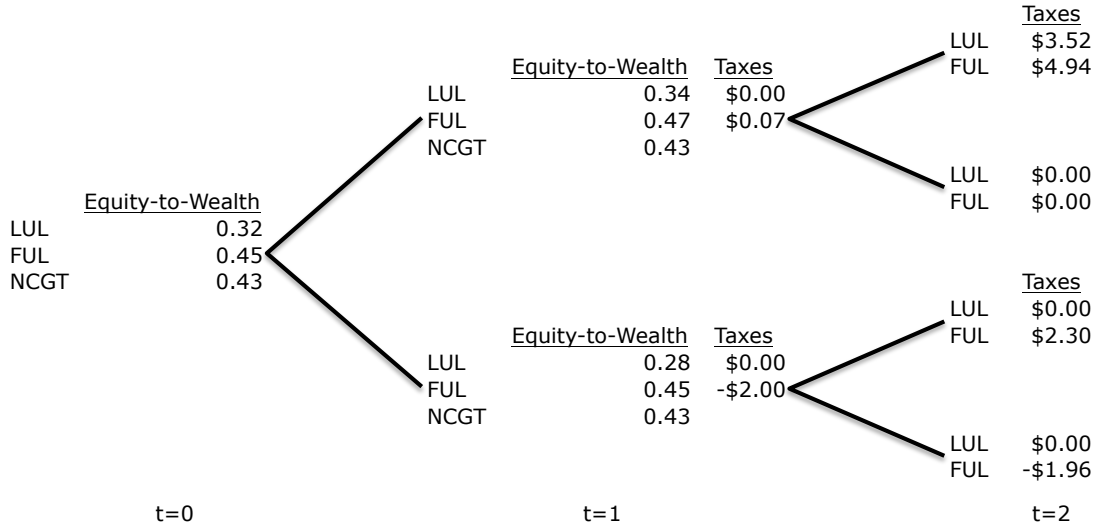


Figure 1 shows that at $t = 0$ the LUL investor chooses an equity-to-wealth ratio of 0.32, which is significantly below the constant equity-to-wealth ratio of 0.43 for the NCGT investor. The figure also shows that the FUL investor's initial equity-to-wealth ratio, at 0.45, is higher than that of the NCGT investor. We note that from an after-tax risk-return tradeoff perspective, an allocation above the NCGT allocation is possible. If the tax reduces the volatility of after-tax returns more than the after-tax risk premium, the after-tax Sharpe ratio is pushed higher implying a higher demand for equity than the NCGT investor. However, from inspecting the FUL case we see that this intuition is misleading. The FUL investor's increased equity demand is driven by the prospect of artificially cushioning the impact of a stock price drop through a tax rebate. Specifically, if the stock price drops at $t = 1$, a tax rebate of \$2 is collected which immediately increases the FUL investor's wealth. If the stock price increases at $t = 1$, both the LUL and the FUL investors are overexposed to equity with embedded capital gain. Given the LUL investor started with a smaller investment, the equity-to-wealth ratio of 0.34 is again smaller than for the NCGT investor. The FUL investor still holds the most equity with an equity-to-wealth ratio of 0.47. This simple short-horizon example shows that it can be quite important how capital losses are treated in optimal portfolio choice problems with capital gain taxation.

Obviously, the above example does not allow for gauging how the use of capital losses drive portfolio choice in long-horizon examples. In this regard, we use the test region iterative contraction method, described in Yang and Tompaidis (2013), to solve investors' lifetime dynamic portfolio choice problem with capital gain taxation since it allows us to handle several endogenous state variables and singularities that are due to capital gain taxation and the limited use of losses.¹ Therefore, our analysis differs from the literature² in the way we model the use of capital losses and in that we seek to understand how LUL strategies evolve over time in realistic settings through a series of backtests using the S&P 500 Index and popular investment strategies such as Small and Large or Value and Growth stock portfolios.

What is wrong with the intuition that after a few years the equity holdings of the FUL and LUL investors are locked-in and the treatment of the use of capital losses becomes irrelevant? Absolutely, nothing. We do find several examples in our backtests, where after only six years both the LUL and FUL are locked-in, the LUL investor does not have any carry-over loss, and expected tax rebates do not play a significant role in the portfolio of the FUL investor. In these cases, we see that the equity-to-wealth ratios equate after six years and from there on move in lockstep until the end of the backtest - only a Japan style stock market meltdown can unlock these equity-to-wealth ratios. However, the data tell a more balanced story than the above intuition suggests as we also see examples where it takes significantly longer before the FUL and LUL investors' equity-to-wealth ratios equate. In the backtests, it takes on average 30 years before the equity-to-wealth ratios converge. Furthermore, the wedge between the FUL and LUL investors' equity-to-wealth ratios is not only long-lived but appears to be also large. We present examples where the LUL equity-to-wealth ratio remains below the FUL equity-to-wealth ratio by at least 10 percent for at least 5 years and by at least 5 percent for at least 20 years. We note that our conclusions are conservative as the current U.S. capital gain tax rate is low relative to historical U.S capital gain tax rates and relative to capital gain tax rates in many other countries, risk aversion can be lower than what we use, capital gain taxes are not forgiven at death in Canada or Europe, and we use a short horizon for the bequest motive. Otherwise, the wedge between LUL and FUL equity-to-wealth ratios would be even larger and longer lived.

¹While alternative numerical solution approaches, such as the one in Brandt et al. (2005), Gallmeyer, Kaniel and Tompaidis (2006), and Garlappi and Skoulakis (2008), exist, we were unable to implement them for the two-stock case with the limited use of losses.

²See the single stock model with capital gain taxes of Dammon, Spatt and Zhang (2001b) and the multiple stock model with capital gain taxes of Gallmeyer, Kaniel and Tompaidis (2006), both of which permit the full or unrestricted use of capital losses.

There is a general pattern in the paths of equity-to-wealth ratios across different backtest periods. Upon entering the stock market, the LUL investor holds less equity than the NCGT investor. On the contrary, upon entering the stock market the FUL investor holds more equity than the NCGT investor as tax rebates from capital losses reduce downside risk. On average, the stock market appreciates over time and the FUL and LUL investors' equity-to-wealth ratios end up higher than their own optimal equity holdings that they would choose without embedded gains or losses. To avoid over-exposure to equity, capital gain taxed investors have an incentive to sell equity but will do so only if the incentive of selling exceeds the tax-induced trading cost. What differentiates the FUL investor from the LUL investor is that the FUL investor has higher incentive to sell than the LUL investor. While embedded capital gains reduce both the probability of collecting a carry-over loss and tax rebates, tax rebates are worth more as they can be utilized immediately. In fact, a carry-over loss in the extreme might remain unused.

Inspecting the optimal no-trade regions sheds further light on the differences and similarities between the FUL and LUL investors. The LUL no-trade region lies below the FUL no-trade region except when the basis-to-price ratio is low, indicating that an FUL investor enters the market at a higher equity-to-wealth ratio, and maintains a significantly larger equity position most of the time. Only when the basis-to-price ratio is sufficiently low, which can happen once the investors become locked-in in their positions due to significant capital gains, do the two no-trade regions overlap.

We also perform a thought experiment to measure the cost for a hypothetical investor, denoted FUL(LUL), who is up to date with the financial literature and follows the FUL investment strategy, but faces the LUL treatment of the U.S. government. We find that the FUL(LUL) investor sells more equity and thus pays more capital gain taxes and that his portfolio contains larger and more persistent carry-over losses than the LUL investor. All of this is driven by the elevated equity-to-wealth ratio of the FUL investor, which the FUL(LUL) investor replicates from the literature. During stock market booms the FUL(LUL) investor's consumption and wealth increases in lockstep with the ones of the FUL investor. During periods when the stock market declines the consumption and wealth of the FUL(LUL) investor can fall more than those of the LUL investors. Hence, the FUL(LUL) investor's consumption and wealth can show higher volatility than those of the FUL, LUL, and NCGT investors. It is not surprising that over time the expected cumulative utility of the FUL(LUL) investor declines gradually and significantly below that of the LUL investor. When young, the FUL(LUL) investor

enjoys a roughly 4 percent higher wealth equivalent than the NCGT investor, i.e., he prefers to be taxed just as the FUL investor. When old, the wealth equivalent of the FUL(LUL) investor is roughly 8 percent below that of the LUL investor and 10 percent below the NCGT investor.

In the two-stock backtests with Small and Large and Value and Growth stocks we see large deviations in the portfolio of the LUL investor relative to the unconstrained optima. For example, we see simultaneously that one equity-to-wealth ratio increases by as much as 10 percent above and the other weight decreases by roughly 7 percent below the unconstrained equity-to-wealth ratios. Such large and persistent deviations from the unconstrained equity-to-wealth ratios that have no embedded capital gains and losses are driven by the fact that both stocks are locked-in but one has more embedded gain than the other. Thus, it is not tax optimal to sell the more locked-in stock and the purchase of the other stock is prevented through an already elevated total equity-to-wealth ratio.

Tax optimal Small and Large and Value and Growth portfolios show that the optimal equity holdings with the limited use of losses remain over extended periods of time significantly lower than the optimal equity holdings with the full use of losses. What is wrong with the intuition that simultaneous capital gains and losses allow to frequently rebalance sufficiently close to the optimal portfolio without any embedded capital gains or losses, thereby substantially reducing the role for carry-over losses, and even capital gain taxes altogether, in optimal portfolios? Everything. First, in the backtests we see that there is a conflict, which often binds, between under-diversification and an over-exposure to equity at the portfolio level. This evidence suggests that simultaneous capital gains and losses do not occur in the data we use. Hence, the role of carry-over losses and, more generally, of capital gain taxation for portfolio choice is not significantly reduced through diversification. Second and more importantly, the intuition is misguided: The more there are simultaneous capital gains and losses in the portfolio the closer the optimal equity allocation can be to the allocation without embedded gains and losses. If so, this would tend to amplify but not reduce the effect of the limited use of losses.

While almost all our focus is intentionally on the backtests, as they are not only easier to interpret than hypothetical simulations but also provide valuable information about the performance of tax managed portfolios in the time series, we do want to emphasize that the optimization problems studied are at the computational forefront of portfolio choice theory. For each stock, we need to keep track of two state variables: the stock holding and the weighted average purchase price. In addition, tracking the unused carry-over loss under the LUL assumption requires one more state variable. Thus, the

one-stock FUL (LUL) case has two (three) state variables and the two-stock FUL (LUL) case has four (five) state variables excluding time. All state variables of the model are endogenous. As a runtime benchmark based on our computing resources, the two-stock LUL portfolio choice problem takes approximately 90 hours to solve using 100 CPUs (2.66GHz) running in parallel on supercomputers.³ Further, we stress that the method used to solve the portfolio choice problem can be applied to other large asset allocation problems.

While our modeling of capital gain taxation through limited use of losses is more realistic compared to the full use of losses, modeling individuals' investment behavior remains a challenge. According to the Department of the Treasury (Office of Tax Analysis), realized capital gains between 1954 and 2009, roughly amount to 1.76% to 7.35% of GDP, are roughly 3.5 times larger than corporate capital gains (1955-1999), and roughly yield 3 times more tax revenue than dividends (2000-2005). Matching these empirical facts requires modeling additional incentives to sell equity, such as inflows through saved income or saved returns on capital and inheritance, outflows such as purchase of real estate, college tuition for offspring and dis-saving over retirement, anticipating changes in the capital gain tax rate over time, and perhaps behavioral biases.⁴ Each additional assumption complicates the formulation of the problem, and makes it difficult to solve with current optimization tools.

The paper is organized as follows. Section 2 describes the portfolio problem. Section 3 provides an example that highlights the intuition behind the role of the limited use of capital losses. Section 4 reports lifetime properties of backtested optimal portfolios and analyzes the economic costs of the use of losses in capital gain taxation portfolio problems. Section 5 concludes. Appendix A gives a thorough description of the problem studied. Appendix B discusses the numerical procedure used.

2 The Consumption-Portfolio Problem

The investor chooses an optimal consumption and investment policy at trading dates $t = 0, \dots, T$. Our assumptions concerning the exogenous price system, taxation, and the investor's portfolio problem are outlined below. A full description of our partial equilibrium setting is given in Appendix A.

³Modeling a third stock requires two more state variables and amplifies the computational cost by at least 10 times using a quasi-random grid, which is parsimonious as compared to the regular grid.

⁴For example, individual investors tend to hold losers while selling winners, i.e., they exhibit loss aversion, which appears inconsistent with a portfolio strategy that minimizes capital gain taxes.

2.1 Security Market

The set of financial assets available to the investor consists of a riskless money market and multiple stocks. The money market pays a continuously-compounded pre-tax rate of return while stocks pay dividends. To characterize the evolutions of prices and payouts, we sample from empirical distributions.

2.2 Taxation

Dividends and interest income are taxed as ordinary income on the date they are paid at rates τ_D and τ_I , respectively. Realized capital gains and losses are subject to a constant capital gain tax rate τ_C .

The tax basis used for computing realized capital gains or losses is calculated as a weighted-average purchase price.⁵ When an investor dies, capital gain taxes are forgiven and tax bases of stocks owned reset to current market prices. This is consistent with the reset provision in the U.S. tax code. Dividend and interest taxes are still paid at the time of death. We also consider the case when capital gain taxes are not forgiven which is consistent with the Canadian and many European tax codes. While we allow investors to “wash sell” and immediately rebalance after they realize capital losses, they are precluded from shorting the stock which eliminates a “shorting against the box” transaction to avoid paying capital gain taxes.⁶

A common assumption regarding the capital gain tax code in the portfolio choice literature is that there are no restrictions on the use of capital losses. It has the computational advantage that capital losses are never carried over and hence the investor does not need to keep track of an additional state variable. Tax codes, however, restrict the use of capital losses. We define the two cases as follows.

Definition 1. *Under the full use of capital losses (FUL) case, an investor faces no restrictions on the use of realized capital losses. When realized capital losses are larger than realized capital gains in*

⁵The U.S. tax code allows for a choice between weighted-average price and exact identification of the shares to be sold. The Canadian and European tax codes use the weighted-average price rule. 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 strategy becomes numerically intractable for a large number of trading periods given the dimension of the state variable increases with time (Dybvig and Koo, 1996; Hur, 2001; DeMiguel and Uppal, 2005). However, for parameterizations similar to those in this paper, DeMiguel and Uppal (2005) numerically show that the certainty-equivalent wealth loss using the weighted-average price basis rule as compared to the exact identification rule is small.

⁶A wash sale is a sale of a financial security with an embedded capital loss and a proximate repurchase (within 30 days before or after the sale) of the same or substantially similar security. We permit wash sales as highly correlated substitute securities, that are not considered substantially similar, typically exist in most stock markets allowing an investor to re-establish a position with a similar risk-return profile after a capital loss. For an analysis of possible portfolio effects of wash sales when adequate substitute securities do not exist, see Jensen and Marekwica (2011). A shorting against the box transaction involves short selling securities that the investor owns to defer tax on capital gains. The Taxpayer Relief Act of 1997 no longer allows delaying taxation through shorting.

a period, the remaining capital losses generate a tax rebate that can be immediately invested.⁷

Definition 2. Under the limited use of capital losses (LUL) case, an investor can only use realized capital losses to offset current realized capital gains. Unused capital losses can be carried forward indefinitely to future trading dates.

We assume that FUL and LUL investors immediately realize all capital losses each period even if they are not used.⁸ Our definition of the limited use of capital losses does not include the ability to use capital losses to offset current taxable income.⁹ Additionally, our analysis does not distinguish between differential taxation of long and short-term capital gains since our investors trade annually.¹⁰

2.3 Investor’s Portfolio Choice Optimization Problem

To finance consumption, the investor trades in risky stocks and the money market. The setting we have in mind is one where a taxable investor trades individual stocks or exchange traded funds (ETFs)¹¹ or some other form of investment strategy that is taxable at the investor level. Given an initial equity endowment, a consumption and security trading policy is *admissible* if it is self-financing, involves no short selling of stocks, and leads to nonnegative wealth over the investor’s lifetime. The investor lives at most T periods and faces a positive probability of death each period.

The investor’s objective is to maximize expected utility of real lifetime consumption and a time of death bequest motive by choosing an admissible consumption-trading strategy given an initial endowment. The utility function for consumption and terminal wealth is of the constant relative risk aversion form with a relative risk aversion coefficient γ .

Using the principle of dynamic programming, the Bellman equation for the investor’s optimization

⁷Definition 1 is used in several papers that study portfolio choice with capital gain taxes (Constantinides (1983); Dammon, Spatt and Zhang (2001a,b, 2004); Garlappi, Naik and Slive (2001); Hur (2001); DeMiguel and Uppal (2005); Gallmeyer, Kaniel and Tompaidis (2006)).

⁸The no-arbitrage analysis in Gallmeyer and Srivastava (2011) shows that, under the LUL case, an investor is indifferent between realizing an unused capital loss or carrying it forward.

⁹In the U.S. tax code, individual investors can only offset up to \$3,000 of taxable income per year with realized capital losses. Allowing for this tax provision requires keeping track of wealth as an extra state variable. Marekwica (2012) shows that asymmetries in the tax code such as the \$3,000 dollar rule introduce the incentive to periodically realize capital gains to allow for using realized losses in the future for tax rebates on income. This feature of the U.S. tax code favors poor LUL investors but likely has only a small impact on most investors. Further, the relevance of the \$3,000 dollar rule has decreased considerably over time as the capital loss limit has not increased since 1978.

¹⁰For such an analysis, see Dammon and Spatt (1996).

¹¹To isolate the effect of the LUL assumption, we abstract away from investing in mutual funds where unrealized capital gain concerns can also be important. Like mutual funds, ETFs must pass unrealized capital gains onto investors generated by portfolio rebalancing. However, many ETFs substantially reduce and in some cases eliminate unrealized capital gains. This is achieved through a “redemption in kind” process described in Poterba and Shoven (2002).

problem, derived in Appendix A, can be solved numerically by backward induction starting at time T . The numerical algorithm is described in Appendix B.

3 A Two Date Example

In this section we return to the two trading date example described in the introduction to highlight the role the limited use of capital losses plays in determining an investor's optimal trading strategy.

Consider that the investor lives with probability one until $T = 2$ and maximizes expected utility of final period wealth over CRRA preferences with a coefficient of relative risk aversion equal to 5. The investor trades in one non-dividend paying stock and a riskless money market. Over time, he pays taxes on the money market's interest payment as well as capital gain taxes on the stock. At T , the portfolio is liquidated and all after-tax wealth is consumed. In this example, no capital gain tax liabilities are forgiven at time T . Endowment consists of one share of stock with a pre-existing tax basis-to-price ratio, $b(0)$, that is varied to capture different tax trading costs. When the $t = 0$ tax basis-to-price ratio is set lower (higher) than one, the investor has a capital gain (loss) in his position.

Using the same notation as Appendix A, the price system parameters are $S_0(0) = S_1(0) = 1$, $r = 0.05$, $\mu = 0.08$, and $\sigma = 0.16$, where S_0 and S_1 denote the money market and stock price, respectively. The rate of appreciation (depreciation) of the stock over one time period is set at $e^\sigma = e^{0.16} = 1.174$ ($e^{-\sigma} = e^{-0.16} = 0.852$). The continuously-compounded expected stock return $\mu = 0.08$ determines the probabilities in the binomial tree. The range for the basis-to-price ratio $b(0)$, $[0.73, 1.38]$, covers the range of the stock price at T . Tax rates are $\tau_I = 0.35$ and $\tau_C = 0.3$.

Figure 2 summarizes the evolution of the optimal portfolio choice expressed as an equity-to-wealth ratio $\bar{\pi}$ (top three plots in the left panel) and the capital gain taxes paid Φ_{CG} (top three plots of the middle panel and all plots in the right panel) conditional on the initial basis-to-price ratio $b(0)$. From Figure 2, we see that a NCGT investor always maintains an equity-to-wealth ratio of approximately 0.43. At $t = 0$, the investor reduces his position from 1 share to 0.43 shares given the stock price is initially one; the proceeds of selling 0.57 shares are invested in the money market. At $t = 1$, when the stock price increases, the investor's fraction of wealth in equity rises above its optimum. The investor then reduces his equity-to-wealth ratio back to 0.43 by selling shares of stock and investing the proceeds in the money market. When the stock price decreases at $t = 1$, the investor is underexposed to equity and buys shares by selling part of the money market investment to again reach an equity-to-wealth

ratio of 0.43.

For a large enough basis-to-price ratio ($b(0) \geq 1.15$), we see from the top left plot of Figure 2 that capital gains tax effects are irrelevant for the LUL investor. In this region, realized capital losses at time $t = 0$ are large enough to cover any possible future capital gain taxes as shown in the Figure 2 tax plots. When the basis-to-price ratio $b(0)$ is between 1.07 and 1.15, the LUL investor still never pays any capital gain taxes over his lifetime, but only by reducing his equity-to-wealth ratio at time $t = 0$ relative to the NCGT case. When $b(0) = 1.07$, the LUL investor's optimal equity-to-wealth ratio reaches a minimum at 0.27. As the basis-to-price ratio falls toward 1.0, the LUL investor optimally holds slightly more equity at $t = 0$.¹² Tax trading costs at $t = 0$ matter for the LUL investor when the basis-to-price ratio falls below 1.0 as the lock-in effect now becomes relevant. Specifically, as the basis-to-price ratio falls, the tax cost of trading at time $t = 0$ begins to dominate the benefit of holding less stock.

For the FUL investor, the ability to collect tax rebates through tax loss selling skews his portfolio choice as his optimal equity-to-wealth ratio always remains above the NCGT case. Additionally, the tax rebate artificially inflates his $t = 0$ wealth $W(0)$ for a basis-to-price ratio above 1, as seen in the bottom left plot of Figure 2. Given the FUL investor's equity-to-wealth ratio is above the NCGT case and his wealth is elevated, his dollar investment in equity at $t = 0$ is also significantly higher than the NCGT case. Further, for a basis-to-price ratio above 1, we see from the bottom middle plot of Figure 2 that the FUL investor's expected utility at $t = 0$ exceeds the expected utility of the NCGT investor, i.e., the FUL investor prefers to be taxed. For a basis-to-price ratio below 1 when there are no tax rebates available at $t = 0$, the probability of collecting tax rebates in the future still skews the FUL investor's equity allocation since he continues to hold more than the NCGT benchmark and the LUL investor. At the lowest initial basis-to-price ratio $b(0) = 0.73$, the FUL investor can never collect a tax rebate in the future. At this point, tax rebates no longer skew the FUL investor's trading strategy implying that LUL and FUL strategies equate.

This simple example shows that the LUL investor's optimal trading strategy at $t = 0$ is sensitive to tax trading costs as captured by the basis-to-price ratio. If current capital losses are large enough to offset all future capital gain taxes, the LUL investor trades as if he is the NCGT investor. For

¹²In long horizon portfolio problems the equity-to-wealth ratio reaches a minimum at $b(0) = 1$, not slightly above it. In the two-period example it is possible to eliminate capital gain taxes in all states by reducing the allocation to equity, as there is a small loss in the portfolio. This is, in general, not feasible in long horizon problems.

no or small capital gains or losses embedded in the current portfolio, future taxes cannot be offset leading to a lower demand for equity than the NCGT investor. If capital gains are large enough and the current allocation is above the NCGT equity-to-wealth ratio, the LUL investor is reluctant to sell as tax trading costs are too high relative to the gains from investing in the money market.

4 Dynamic Tax Trading Strategies

To understand quantitatively the significance of capital losses on investor’s lifetime consumption-portfolio problem, we consider long-dated dynamic consumption-portfolio problems with risky assets whose dynamics are based on the empirical distribution of historical returns. Specifically, we focus on the evolution of backtested optimal portfolios using the time-series of popular investment strategies. In addition, we provide two measures for the economic costs of the limited use of losses.

4.1 Parameterizations

The investor begins trading at age 20 and can live to a maximum of 100 years.¹³ He has a time discount parameter $\beta = 0.96$ and his relative risk aversion coefficient is set at $\gamma = 5$. The bequest motive is set such that the investor plans to provide an equal amount of payment each year in real terms for 20 periods to his heirs.

The tax rates used are set to roughly match those faced by a wealthy investor under the U.S. tax code. We assume that interest is taxed at the investor’s marginal income tax rate $\tau_I = 35\%$. Dividends are taxed at $\tau_D = 15\%$. The capital gain tax rate is set to the long-term rate $\tau_C = 20\%$.¹⁴ To be consistent with the U.S. tax code, capital gain taxes are forgiven at the investor’s death.

To avoid potential biases caused by the Normal return assumption and be more consistent with backtests on historical paths, we adopt the empirical distribution assumption to find the optimal portfolios. Specifically, we compute one stock optimal portfolios using the empirical distribution of

¹³The 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.

¹⁴The U.S. Tax Relief and Reform Act of 2003 changed several features of the tax code with respect to investments. Specifically, the long-term capital gain tax rate dropped from $\tau_C = 20\%$ to $\tau_C = 15\%$ for most individuals. Dividend taxation switched from being linked to investor’s marginal income tax rate to a flat rate of $\tau_D = 15\%$. The 2006 Tax Reconciliation Act extended these rates to be effective until 2010. This was further extended through 2012 in legislation passed by Congress. From 2013, rates reverted to $\tau_C = 20\%$. Beginning in 2013 there is a contribution to Medicare of 3.8% for the lesser of the net investment income or the excess of the adjusted gross income above \$200,000 (\$250,000 for a joint return or half thereof if married taxpayers file separate returns). For a comprehensive summary of U.S. capital gain tax rates through time, see Figure 1 in Sialm (2009).

the S&P 500 Index and two stock optimal portfolios using the empirical distributions of Small and Large and Value and Growth stock portfolios, based on the sorts from Kenneth French’s website.¹⁵ As compared with the Normal distribution, the empirical distribution captures the negative skewness in returns, which implies a higher probability of large negative returns and lower probability of large positive returns. We acknowledge that the constant opportunity set assumption does not allow for any autocorrelation structure over time. Yet, all the historical annual return series used show very low autocorrelations and are statistically insignificant at the 95% confidence level.

4.2 Benchmarks

To understand the role of the LUL assumption on portfolio choice, we compare it to three benchmark portfolio choice problems. The first benchmark is the case when the investor faces no capital gain taxation, abbreviated as NCGT. In this benchmark, the investor still pays dividend and interest taxes. Second, we use the FUL case as a benchmark when the investor faces no restrictions on the use of capital losses. Third, we subject the FUL investor to the limited use of capital losses tax treatment, abbreviated as FUL(LUL). We do this by assuming that the FUL investor keeps investing and consuming at the same rate as a FUL investor even though he never receives a tax rebate but can only use capital losses to offset current or future capital gains as the LUL investor. The FUL(LUL) benchmark helps measuring the economic cost of the FUL rule, a simplified taxation rule, that an investor might follow to reduce the complexity of the portfolio optimization problem.

4.3 Backtesting with the S&P 500 Index: Equity-to-Wealth Ratios

We consider a sequence of investors reaching age 20 in each of the years between 1927 and 1999. When investors turn 20 they start to invest into equity and the riskless money market. Each investor is endowed with 100 dollars. Using the optimal portfolios computed from the empirical distribution of the S&P 500 Index,¹⁶ we let each investor experience the realized path of the S&P 500 Index, contemporaneously with realized paths of dividends, inflation, and interest rates.

Figure 3 plots realized paths of equity-to-wealth ratios of selected backtest windows, with the only difference being the year when a 20-year-old investor starts to invest. From the top right plot of Figure

¹⁵A description of how we sample from historical time series is given in Appendix A.

¹⁶The Internet Appendix presents selected tabulated examples of optimal equity-to-wealth ratios ($\bar{\pi}(t)$) conditional on the beginning period equity-to-wealth and basis-to-price ratios ($\bar{\pi}(t)$ and $b(t)$), at ages 20 and 80.

3, we see that a LUL investor, who enters the market in 1950, optimally chooses to invest 38 percent of his wealth in S&P 500 Index, while the FUL and NCGT investors start with equity-to-wealth ratios of 53 percent and 42 percent, respectively. As time passes, investors dynamically adjust their equity holdings, where the NCGT investor's equity holding stays constant, the FUL investor's equity holding fluctuates between 50 and 56 percent, and the LUL investor's equity holding steadily increases and converges towards the FUL investor's equity holding. We see that it only takes six years for the FUL and LUL equity-to-wealth ratios to converge. From this example, one might conclude that the treatment of capital losses is not crucial for how real life equity-to-wealth ratios evolve.

Yet, the top left (1929 to 1959) and the middle left (1965 to 1995) plots of Figure 3 show examples where the LUL equity-to-wealth ratio remains below the FUL equity-to-wealth ratio for periods of twenty years or more. In these examples, the LUL equity-to-wealth ratio remains 10 percent below the FUL equity-to-wealth ratio for at least five years, middle plots, and remains on average five percent below the FUL equity-to-wealth ratio for at least twenty years, top left and middle left plots. From these examples, we find it more difficult to conclude that the treatment of capital losses is inconsequential for how real life equity-to-wealth ratios evolve over time.

To mitigate cherry-picking concerns, the bottom left plot displays the average path of equity-to-wealth ratios by investor age. The average is computed from 58 investors who enter the stock market at age 20 in 1927, 1928, ..., and 1984, respectively. The 58'th investor enters the equity market in 1984 and reaches age 50 in 2014. We highlight with this exercise that there exists a general pattern in the paths of equity-to-wealth ratios across different backtest periods. At the beginning of each backtest period, the LUL investor holds less equity than the NCGT investor. The reduced equity-to-wealth ratio results from the trade off between harvesting equity risk premium and minimizing capital gains taxation. On the contrary, the FUL investor holds more equity than the NCGT investor because the tax rebates on capital losses reduce the downside risk of equity. As time passes, the NCGT investor holds a constant fraction of his wealth in equity, the FUL investor's equity-to-wealth ratio fluctuates roughly around a constant level, and the LUL investor's equity-to-wealth ratio appears to eventually converge to the level of the FUL investor, although the speed of convergence varies. The periods of 1950 to 1965 and 1980 to 1995 are examples of fast convergence between LUL and FUL. The period of 1965 to 1995 is an example roughly on par with the average convergence in the bottom left plot, which is 30 years. Examples of slow convergence are the first backtest window, 1929 to 1959, and the

last backtest window, 1999 to 2014.

We close the discussion of how the use of capital losses drives equity-to-wealth ratios with the last backtest window of Figure 3. It shows the most recent episode with the first rebalancing taking place in January 1999 and the last one in January 2014. The wide initial gap between the FUL and LUL equity-to-wealth ratios demonstrates a modest sign of convergence of roughly 5 percent over this 15-year period. If similar market conditions repeat in the near future, then only after 45 years or at age 65 will the equity-to-wealth ratios of LUL and FUL equate.

4.4 Backtesting with the S&P 500 Index: Trading and Taxes

To better understand tax induced trading in the backtests, in particular why the FUL and LUL equity-to-wealth ratios over some periods converge faster than over others, we depict investor's optimal trading strategy by showing the optimal no-trade regions and how the equity-to-wealth ratios evolve around the no-trade regions in Figure 4. The figure shows two representative backtest windows from Section 4.3, namely, one example of fast convergence, period 1950 to 1965, and the other example of slow convergence, period 1999 to 2014.¹⁷ For the same backtest windows, we present the evolutions of basis-to-price ratio, carry-over loss to wealth ratio, and the cumulative capital gain taxes in Figure 5 to demonstrate the tax consequences of trading over the realized return path.

The no-trade regions in Figure 4 are the regions in between the upper boundary and lower boundary. The left plots of Figure 4 show the upper and lower boundaries of the FUL (green dashed line) and LUL (blue solid line) investors with zero carry-over loss at age 20 (dark color) and age 80 (light color). In the right plots of Figure 4 we show how a 1% (red dashed line) and a 10% (beige dotted line) carry-over loss to wealth ratio shrink the no-trade region of the LUL investor at age 20. Roughly, we see that a 1% (10%) carry-over loss ratio shrinks the no-trade by one (two) third relative to the case when there is no loss in the portfolio (blue solid line).

The no-trade region drives the investor's optimal trading in the following way. Whenever an equity-to-wealth ratio is above (below) the no-trade region it is optimal to trade back to the upper (lower) boundary of the no-trade region.¹⁸ This implies selling when above the no-trade region, buying when

¹⁷Figure 4 only shows the first 10 years of period 1950-1965 and the first 8 years of period 1999-2007.

¹⁸Portfolio choice with capital gain taxes can be compared to portfolio choice with transaction costs; see for example Dumas and Luciano (1991), Liu and Loewenstein (2002), the multiple risky asset analysis in Liu (2004), and the literature therein. Specifically, from this literature we know that there exists a no-trade region, in which investors optimally refrain from trading. Once there is trade, investors will trade back to the boundary if there is only a proportional transaction cost or trade back inside the no-trade region if there are both fixed and proportional costs.

below the no-trade region, and no-trade when in the no-trade region.

From Figure 4, we see that the LUL no-trade region lies below the FUL no-trade region for a wide range of basis-to-price ratios. The difference is the largest when the basis-to-price ratio is around one and diminishes as the basis-to-price ratio decreases. As the basis-to-price ratio decreases, investors face more embedded gains in their stock holdings and are less likely to incur losses in the future, thus the differences in treating capital losses become less important. Further, the shape of the FUL and LUL no-trade regions are different especially for the upper boundaries which are more relevant for investor's trading decisions in a stock market that is on average appreciating. As the basis-to-price ratio decreases, both the FUL and LUL upper boundaries increase. This is due to the fact that as both investors become more locked-in, they are more reluctant to sell their elevated equity positions due to higher capital gain taxes per unit of stock sold. In addition, the FUL upper boundary has a much smaller negative slope than the LUL upper boundary since when the basis-to-price ratio is close to one the FUL upper boundary is above the LUL one and eventually these two boundaries converge as the basis-to-price ratio decreases. The difference in shape implies that on average the FUL investor's equity-to-wealth ratio is above the no-trade region right away after he enters the stock market and remains above it most of the time. The LUL investor's equity-to-wealth ratio instead on average moves into the no-trade region and stays there over several trading periods after he enters the stock market. Over time, the average basis-to-price ratio declines for both investors. Eventually, the LUL investor's equity-to-wealth ratio also increases above the upper boundary of the no-trade region and from there on we see that the LUL investor sells equity shares just as the FUL investor. Once the upper boundary of the LUL and FUL no-trade regions coincide, the equity-to-wealth ratios equate and move in lockstep.

Inspecting the equity-to-wealth ratios in the left plots of Figure 4 of the FUL dynamic tax trading strategies over the backtest period from 1950 to 1960, we observe that all entering equity-to-wealth ratios are above the no-trade region, except in 1958. Hence, in all but one year the exiting equity-to-wealth ratio is lower than the entering one implying that the FUL investor is selling equity, realizing capital gains, and pays capital gain taxes as shown in the bottom left plot of Figure 5. The exiting equity-to-wealth ratios gradually increases over time as the upper boundary of the no-trade region slightly increases for lower basis-to-price ratios. The FUL investor is locked-in from the start and gradually becomes even more locked-in. In contrast to the FUL investor, the LUL investor enters the

no-trade region just after entering the stock market and remains there until 1955, which is the first year when the LUL investor sells equity. In 1956, just after 6 years, the equity-to-wealth ratios of both investors are almost equal. From 1957 on, even in the no-trade region, the equity-to-wealth ratios move in lock-step. This is entirely driven by the overlap between the upper boundary of the LUL and FUL no-trade regions, which occurs slightly below a basis-to-price ratio of 0.4. The last observation we make is that over the entire 10 years neither of the investors ever purchases equity as entering equity-to-wealth ratios are never below the no-trade regions.

Turning to the 1999 to 2007 backtest, what differentiates this period from the period of 1950 to 1960 is the market crash in 2000. Over three successive years from 2000 to 2002 the stock market depreciates and the basis-to-price ratios are pushed up from below 1 with embedded gains to above 1 with embedded losses. As shown in Figure 5, both the FUL and LUL investors optimally realize their capital losses. While the FUL investor collects tax rebates evidenced by the decrease in cumulative taxes, the LUL investor accumulates carry-over losses roughly equal to 10 percent of wealth. Starting from 2003, the stock market recovers and keeps appreciating until 2008. Over this period, armed with an unused carry-over loss the LUL investor's upper boundary of the no-trade region is pushed down significantly as shown in Figure 4. As a result, the LUL investor starts selling equity in 2004 at a rate higher than what he would sell if there was no carry-over loss. Specifically, the investor keeps on selling in 2005, 2006, and 2007, while he would choose not to trade if there was no carry-over loss. Thus, the non-zero carry-over loss implies a significant and persistent gap between the upper boundaries of the FUL and LUL no-trade regions. This results in a significant and persistent gap in equity-to-wealth ratios as both the FUL and LUL investors keep on selling equity back to their own upper boundaries after 2004. In other words, a carry-over loss slows down the convergence between the FUL and LUL equity-to-wealth ratios.

Briefly, to complete the view of the dynamic tax trading strategies, in Figure 5, we plot investors' tax basis, the carry-over loss to wealth ratio, and the cumulative capital gain taxes. In both examples, all taxed investors show almost identical tax bases. This is due to the fact that the tax basis is computed as a weighted average purchase price. Only the purchase of equity can alter the tax basis, but not the selling of equity. Over the period of 1950 to 1965, large amount of stock purchases only occur at the beginning of the backtest period. Afterwards, taxed investors keep selling stock most of the time and purchase stock only a few times at very small amounts. Over the period of 1999 to 2014,

the FUL and LUL investors buy and sell stock synchronously for almost the same amount.

4.5 Backtesting with the S&P 500 Index: Wealth and Consumption

The consequences of the limited use of capital losses are not limited to investors' equity holdings and capital gain taxes, but are also reflected in the investors' wealth and consumption. We plot realized paths of investors' wealth and consumption in dollars, wealth and consumption scaled by the wealth and consumption of the NCGT benchmark, respectively, and GARCH(1,1) volatilities of wealth and consumption for the period 1999 to 2014 in Figure 6. Figures showing wealth and consumption over other backtested periods are presented in the accompanying Internet Appendix.

From the top and middle left plots of Figure 6, we see that the wealth paths implied by different tax trading strategies all closely follow the market cycles, where the relative differences between the investors are highlighted through scaling by the wealth of the NCGT benchmark. The wealth implied by the FUL tax trading strategy is higher than all other strategies at the end of the backtest period. It is also higher than the other strategies at market peaks and remains on top most of the time over the backtest period. This wealth difference is driven by the FUL investor's higher equity-to-wealth ratios relative to the LUL and NCGT investors, which matters when returns are positive on average, and by tax rebates when returns are negative, which are available only to the FUL investor. The FUL(LUL) investor, who always maintains an equity-to-wealth ratio identical to that implied by the FUL strategy, has the second largest wealth at the end of the backtest period. We note, however, that the wealth of the FUL(LUL) investor shows the highest volatility and that over periods when the stock market declines it is lower than for all other strategies. This is because the FUL(LUL) investor holds as much equity as the FUL investor but is not cushioned by the tax rebates during market crashes.

The market cycles in the wealth paths carry over to the consumption paths as shown in the top and middle right plots of Figure 6. Although investors smooth consumption in a countercyclical fashion relative to the wealth dynamics, considerable variations in the consumption paths remain, as indicated by the GARCH(1,1) consumption volatility in the bottom-right plot. It is remarkable that the FUL(LUL) investor experiences the highest consumption volatility while the LUL investor shows the lowest consumption volatility over the entire backtest period. While we do frequently observe in the other backtests that the GARCH(1,1) consumption volatility of the NCGT strategy is lower than the LUL strategy, it can be said that in almost all backtests we observe that the FUL(LUL)

strategy shows the largest consumption volatility. Importantly, the difference between the GARCH consumption volatility of the FUL(LUL) and the LUL based investment provides a first, albeit indirect, glance into the real life cost of following the FUL strategy when the use of losses is restricted. The consumption volatility patterns are also reflected in the wealth volatility shown in the bottom left plot.

4.6 The Economic Costs of the LUL and the FUL Cases

While the dynamics of investors' wealth and consumption implied by optimal tax trading are interesting in their own right, it is important to develop a better understanding of the economic costs of the use of capital losses. To quantify the economic cost, we plot the expected cumulative utility and the wealth equivalent scaled by the wealth equivalent of the NCGT investor by age for an FUL, an FUL(LUL), an LUL, and an NCGT investor in Figure 7. The expected cumulative utilities are calculated from 50,000 simulation paths constructed by randomly sampling the historical annual returns of the S&P 500 Index. Each sampled path has 80 periods to reflect the investment horizon between age 20 and 100.

We see from the left plot in Figure 7 that the expected cumulative utility of the LUL investor closely follows the one of the NCGT investor, but always lies below it due to the tax friction. Even when the LUL investor ages there is only a slight increase in the difference between his cumulative utility and that of the NCGT investor. Due to the tax rebates, the FUL investor's expected cumulative utility is larger than the one of the NCGT investor and the difference increases over time. The tax rebates increase consumption and wealth artificially, which increases expected cumulative utility. In contrast, the expected cumulative utility of the FUL(LUL) strategy slowly declines by age relative to the FUL strategy due to the lack of tax rebates, unused carry-over losses, and higher volatilities of wealth and consumption. The cost of ignoring the limitation on carry-over loss accumulates over time. Eventually, the FUL(LUL) benchmark shows the lowest expected utility, which is more than 35 percent below that of the LUL investor at age 100.

For easier interpretation of economic significance, we calculate the wealth equivalent from the expected cumulative utility and rescale it by the wealth equivalent of the NCGT investor, shown in the right plot of Figure 7. The FUL investor is roughly and constantly over his life 4 percent better off than the NCGT investor at the same age, i.e., prefers to be taxed and would be willing to pay for it.

The LUL investor is roughly and constantly 2 percent worse off than the NCGT investor. Economically, the difference of 6 percent between FUL and LUL is quite large. The FUL(LUL) investor starts out just as the FUL investor but by age 40 his wealth equivalent coincides with the one of the NCGT investor and by age 50 coincides with that of the LUL investor. By age 100, the wealth equivalent of the FUL(LUL) investor is 8 percent below the LUL investor, 10 percent below the NCGT investor, and 14 percent below the FUL investor. In other words, ignoring the limited use of capital losses built into the tax codes all over the world will cost an individual investor 8 percent of his entire lifetime wealth.

4.7 Robustness

Optimal portfolios and through that our backtests depend on the capital gain tax rate. To understand the impact of the capital gain tax rate, we also study the Capital Gain Tax 30% Case. A 30% rate roughly equals the 28% rate imposed after the U.S. 1986 Tax Reform Act and is consistent with the long-term capital gain tax rate paid in many European countries. For example, the capital gain tax rates in Denmark, Finland, Norway and Sweden are currently 27%, 30%, 28%, and 30%, respectively. In 2009, Germany's individual capital gain tax rate rose to approximately 28% from 0%.¹⁹ To illustrate a case where stock holdings decrease for the NCGT investor and hence the dollar value of tax-loss selling decreases for the FUL investor, the Higher Risk Aversion Case assumes that γ increases to 10. The No Tax Forgiveness at Death Case assumes capital gain taxes are assessed when the investor dies, a feature consistent with Canadian and European tax codes. Finally, the Bequest Motive 100 Years Case shows the sensitivity of our main analysis to the horizon of the bequest motive. The results of these robustness checks to the one stock backtest are given in the Internet Appendix.

Briefly, we find that a higher capital gains tax rate, a lower risk aversion, no tax forgiveness at death, and a longer horizon for the bequest motive all increase the initial difference between LUL and FUL portfolio strategies and slow down the convergence of the FUL and LUL strategies over time. Hence, the robustness checks show that the main parametrization employed in Subsection 4.1 is conservative. For example, after increasing the capital gain tax rate to 30%, the wealth equivalent of the FUL(LUL) investor by age 100 is 20 percent below the LUL investor, 22 percent below the NCGT investor, and 29 percent below the FUL investor. Therefore, the analyses in Subsections 4.3, 4.6, and

¹⁹The German capital gain tax rate is 25% plus a church tax and tax to finance the five eastern states of Germany.

4.8 provide lower bounds on the differences between LUL and FUL as well as LUL and FUL(LUL).

4.8 Backtesting with Small and Large and Value and Growth Strategies

It is natural to ask how stock portfolios behave when capital gains and losses occur simultaneously across different stocks in the portfolio. A common intuition, which we show to be wrong, is that simultaneous capital gains and losses substantially reduce the role for the use of losses in portfolio choice problems with capital gain taxation, by facilitating frequent rebalancing sufficiently close to the optimal portfolio without embedded capital gains or losses, or even to the equity allocations of the NCGT investor.

Due to the challenges of dimensionality and singularity, it is very difficult to numerically solve a long-horizon portfolio-choice problem with the limited use of capital losses for more than two stocks. Even for the two-stock case, it is still challenging to apply numerical solution approaches such as the one in Brandt et al. (2005), Gallmeyer, Kaniel and Tompaidis (2006), and Garlappi and Skoulakis (2008). Instead, we use the test region iterative contraction method, developed in Yang and Tompaidis (2013), to solve the optimization problem. The Internet Appendix shows that the test region iterative contraction method and the method in Gallmeyer, Kaniel and Tompaidis (2006) produce equity-to-wealth ratios that are indistinguishable for the one stock case with the limited use of losses.²⁰

As for the one-stock backtests with the S&P 500 Index, we consider a series of investors who start trading once they reach age 20. For Small and Large we use data from 1945 to 2013 and for Value and Growth from 1985 to 2013. We pick different data samples to obtain positive weights in the strategies over the entire periods.²¹ Using the optimal portfolios computed from the empirical distribution of i) Small and Large and ii) Value and Growth, we present the evolution of equity-to-wealth ratios and other quantities from 1965 to 1995 and 1999 to 2014 in Figures 8 and 9 for Small and Large and from 1999 to 2014 for Value and Growth in Figure 10.

Figures 8-10 confirm that there is a large and persistent wedge between the total equity-to-wealth

²⁰We note that Gallmeyer, Kaniel and Tompaidis (2006) is based on a random search algorithm, which is reliable for the one stock LUL case and the two stock FUL case but requires relatively large computational costs. Our method solves the first order conditions and more efficiently achieves the same level of accuracy.

²¹At least one of the two “stocks” in the Small and Large portfolio or the Value and Growth portfolio exhibits a significantly lower before tax expected return than the other stock or the S&P 500 Index. The after tax expected returns, which matter for the equity-to-wealth ratios, are even lower. Since we preclude shorting, it implies that equity-to-wealth ratios can be zero. Further, it is not possible to judge whether an equity-to-wealth ratio jumping back and forth between 0 and 0.02 is real or a numerical error pattern. We, therefore, search for a period that ensures that the weights on both stocks in both backtests stay away from the zero bound at all times over the entire backtest period.

ratios, shown in the top right plots, of the LUL and FUL investors. While it is not possible to directly compare the total equity-to-wealth ratios from any two-stock example with an one-stock example, we understand that convergence between the FUL total equity-to-wealth ratio and the LUL total equity-to-wealth ratio is largely driven by realized returns. From the two-asset parameters in the Internet Appendix we see that Small, Large, Value, and Growth have higher realized returns than the return of the S&P 500 Index. Therefore, the basis-to-price ratios of Small, Large, Value, and Growth decline faster than for the S&P 500 Index, which speeds up convergence.

Strikingly, while our two-stock examples are restrictive for our understanding of how diversification effects are traded off with optimal tax trading it turns out that they are not restrictive when it comes to simultaneous capital gains and losses. If we can ascertain that capital gains and losses actually cancel out in a portfolio with a large number of individual stocks, then the optimal equity-to-wealth ratios of the LUL investor will be close to his unconstrained optimum, which is lower than the NCGT benchmark. The longer one can maintain a portfolio in which capital gains and losses cancel out or, at least, cancel out to a large extent, the larger is the difference between FUL and LUL total equity-to-wealth ratios and the more persistent is this wedge. This insight also implies that we are again very conservative in our analysis. It is hard to find investment strategies that are even more successful than Small, Large, Value, and Growth. It might be possible to find examples, unlike the ones in Figures 8 to 10, in which gains and losses do indeed occur in a more synchronous fashion. Therefore, portfolios that underperform the S&P 500 Index, that do generate enough losses to offset the gains, or both, will have a larger and more persistent wedge between FUL and LUL total equity-to-wealth ratios.

Next we further inspect the time series properties of the respective weights in Small and Large and Value and Growth shown in the top left and top middle plots of Figures 8, 9, and 10. First, in the time series we see large deviations of the LUL investor's equity-to-wealth ratios from the initial LUL equity-to-wealth ratios at age 20 and from the equity-to-wealth ratios used by the untaxed investor. Such deviations can be as large as 10 percent of wealth. Frequently, the deviations go in opposite directions for the two "stocks." For example, from Figure 8 we see that between 1980 and 1990, the LUL investor's equity-to-wealth ratio in Small "stocks" is more than 10 percent elevated relative to the equity-to-wealth ratio chosen by the NCGT investor and the equity-to-wealth ratio in Large "stocks" is roughly 7 percent below the equity-to-wealth ratio chosen by the NCGT investor. Second, we see that the FUL and LUL equity-to-wealth ratios in each stock can have any ordering, i.e., LUL

above FUL or FUL above LUL, even though at the portfolio level we always have that the total LUL equity-to-wealth ratio is below or at the total FUL equity-to-wealth ratio as in the one-stock case. Third, while the equity-to-wealth ratios of FUL and LUL investor converge at the portfolio level, they do not necessarily converge at the stock level, at least not over the same period of time.

The above observations relate to the fact that besides the tax impact there is a trade off between deviating from the optimal total equity-to-wealth ratio and the diversification in the portfolio. The desire to control total equity risk through the total equity-to-wealth ratio is at least as significant as the desire to optimize diversification. Specifically, from the basis-to-price ratios, middle left and middle plots, and stock shares, bottom left and bottom middle plot in Figure 8, it is evident that the LUL investor sells Small and Large shares every year starting in 1975. From 1980 on, there is no carry-over loss in the portfolio and thus each sale involves paying capital gain taxes. After 20 years, the cumulative capital gain tax reaches 18 dollars. To understand these trades it is useful to inspect basis-to-price ratios, middle left and middle middle plots, and stock shares, bottom left and bottom middle plots. We see that right after 1975 the Small stock shows higher tax trading costs as it has a lower basis-to-price ratio because the realized returns of Small stocks are on average higher than those of Large stocks. For the same reason the equity-to-wealth ratio in the Small stock deviates more and more from the optimal equity-to-wealth ratio without capital gains and losses. To prevent further deterioration of the diversification in the portfolio, we see more selling of the Small stock than the Large stock, which is evident from the steeper drop in stock shares for the Small stock in the bottom left plot.²² Although the Large stock shows a significantly lower equity-to-wealth ratio over this period than the optimal equity-to-wealth ratio without capital gains and losses, we still see that shares are being sold, to not let the total equity-to-wealth allocation increase by too much.

The FUL investor's initial equity-to-wealth ratio in the stocks deviates considerably from those of the NCGT or LUL investor. This is visible in all three backtests. For example, from the top middle plot of Figure 8 we see that the equity-to-wealth ratio for Large stocks is 6 to 7 percent above the equity-to-wealth ratio of the LUL and NCGT investors. Interestingly, the bottom middle plot shows that after 1977 the FUL investor's stock shares in Large stocks is basically flat. Our intuition for this result is that since the Large stock has a larger basis-to-price ratio its equity-to-wealth ratio is kept high to maximize the dollar amount from a tax rebate in case of a significant decline in the price of

²²The slope of the stock shares is informative here as for both stock shares the range from low to high equals 0.3.

Large stocks. Only after 1990, we see that shares in Large stocks decline to reduce the overallocation to Large stocks. This is consistent with our intuition as by that time expected tax rebates for Large stocks are unlikely, due to their very low basis-to-price ratio.

As for the one-stock case using the S&P 500 Index, we also consider the backtest period from 1999 to 2014. In fact, Large stocks, likely by definition but certainly over the years from 1999 to 2014, show return dynamics that are substantially similar to the S&P 500 Index. This is confirmed by the basis-to-price ratio of Large stocks that follow a pattern that is comparable to the evolution for the S&P 500 Index over the same backtest period. As a result, the Large stock also preserves a similar trading pattern as the S&P 500, shown in the bottom middle plot of Figure 9. What is different from the backtest with the S&P 500 Index is that the average allocation to Large stock is smaller than the average allocation to S&P 500; hence, after going through the same market crash a smaller amount of tax rebate for the FUL investor and a smaller amount of carry-over loss for the LUL investor are realized. The consequence from this for the LUL investor is that he buys Large stock less aggressively over market downturns and sells Large stock less aggressively over major market booms than the FUL and NCGT investors. As the bottom left plot of Figure 9 shows, the LUL, FUL, FUL(LUL), and NCGT sell Small stocks most of the time between 1999 and 2014, with the exception of the period after the stock market crash in 2008 where total equity-to-wealth ratios of the LUL and FUL investor are very close to the initial total equity-to-wealth ratios without embedded capital gains or losses.

Overall, the evolution of stock specific and total equity-to-wealth ratios, and for the other quantities discussed, of the Value and Growth two-stock portfolio from 1999 to 2014 is similar to the Small and Large backtest over the same period. What we can add is that significant increases in the basis-to-price ratio of one or both stocks can lead to significant rebalancing in the portfolio. By 2007 both the LUL and the FUL investor have a large tilt to Value in their portfolio. Selling of Value is prevented by the lock-in through the rather low basis-to-price ratios. Buying of Growth is prevented by the high total equity-to-wealth ratio. From the bottom left and bottom middle plots of Figure 10, which show the stock share in Value and Growth, we see that the LUL and FUL investors somewhat increase their share in Value and significantly buy into Growth stocks after the 2008 stock market decline. Without the stock market decline, which drives down the total equity-to-wealth ratio, such a large rebalancing would involve substantial tax trading costs in Value or an overexposure to equity.

5 Conclusion

We integrate the limited use of losses assumption into multiple stock portfolio problems with capital gain taxation and show that requiring that capital losses can only be used to offset current or future realized gains significantly changes the after-tax risk-return tradeoff of holding equity. In contrast, a full use of losses investor's trades are artificially impacted by tax rebates. These tax rebates act as an income process that pays off in down markets leading to a misleading higher demand for equity relative to an untaxed investor when capital gains are not too large in the existing portfolio. Tax rebates skew optimal wealth, collected taxes, and total dollar investment in equity over an investor's life. The motives for capturing tax rebates are strong enough to generate a counterfactual welfare result — the full use of losses investor prefers to being subjected to capital gain taxes rather than being untaxed.

Through backtesting, using the time series and empirical distribution of the S&P 500 Index and Value and Growth and Small and Large portfolios we document average optimal equity holdings with limited use of losses. Specifically, carry-over losses drive a wedge between the upper boundaries of the limited use of losses and full use of losses no-trade regions. Therefore, on average we see that optimal equity holdings for the the limited use of losses are over an extended period of time significantly lower compared to various benchmark cases including the one where the use of capital losses is unrestricted. In the case of multiple stocks, with Value and Growth and Small and Large portfolios, we see in the backtests that allocations with the limited use of losses remain persistently under-diversified.

Our portfolio choice model is useful in capturing differences between investors subject to capital gain taxation with limited use and full use of losses. To match empirical facts, such as the amount of capital gains realized by individual investors, additional assumptions are required. These assumptions include allowing the investor to earn income, face liquidation shocks (for example to purchase a house, help a child go to college, or due to changes in health), or face a changing capital gain tax rate. We leave these extensions for future research.

A Investor Consumption-Portfolio Choice Problem Description

The mathematical description of the consumption-portfolio choice problem outlined in Section 2 is now presented. Our portfolio choice problem with multiple risky stocks is based on the setting of Gallmeyer, Kaniel and Tompaidis (2006). The major difference is that our work incorporates the limited use of capital losses with no short selling.

A.1 Security Markets and Empirical Distributions

The economy is discrete-time with trading dates $t = 0, \dots, T$ and a time step of 1 year. The investor trades a riskless money market and multiple stocks at each trading date. The riskless money market pays a continuously compounded pre-tax time-varying interest rate $r(t)$. Stock market investment opportunities are represented by dividend-paying stocks with time- t ex-dividend prices $S_n(t)$ for $n = 1, \dots, N$. Each stock pays a pre-tax dividend of $\delta_n(t) S_n(t)$ at time t where $\delta_n(t)$ is stock n 's dividend yield. Stock prices, dividends, and interest rate are all measured in nominal terms. The effect of inflation $i(t)$ is adjusted in the investor's objective.

For the evolutions of stock prices, dividend yields, interest rate, and inflation rate we employ empirical distributions by sampling from historical time series with equal probability assigned to each period. To preserve the correlation structure, all quantities are sampled simultaneously from the same period.

In the one stock analysis we employ the following time series (from Robert Shiller's website) for the period of 1927 to 2013: the S&P 500 index and its dividend yields, the U.S. annual interest rate, and the U.S. annual inflation rate calculated from the Consumer Price Index. For the analysis with two risky assets we use historical returns on stock portfolios such as Small (stocks) and Large (stocks) or Value and Growth (from Kenneth French's website) for the period of 1945 to 2013 and 1985 to 2013, respectively.²³ Note that we back out the dividends of these investment strategies from return series with and without dividends.

A.2 Investor's Optimization Problem

An investor endowed with initial wealth chooses an optimal lifetime consumption and investment policy in the presence of realized capital gain taxation. We assume that the investor makes annual decisions starting at age 20 corresponding to $t = 0$ and exits the economy with certainty at age 100 implying $T = 80$. At time $t < T$, the investor faces a positive probability of death each period, measured by the conditional death probability of $1 - e^{-\lambda t}$ or survival probability of $e^{-\lambda t}$. The single-period hazard rate $\lambda_t > 0$, $t = 0, \dots, T - 1$, are calibrated to the 1990 U.S. Life Table, compiled by the National Center for Health Statistics. At time T , $\lambda_T = \infty$ since the investor lives for at most T periods.

A.2.1 Interest and Dividend Taxation

The investor faces three forms of taxation in our analysis: interest taxation, dividend taxation, and capital gain taxation. Interest income is taxed as ordinary income at the constant rate τ_I , while dividend income is taxed at the constant rate τ_D . Thus the time- t tax-adjusted risk-free rate is

$$R_f(t) = (1 - \tau_I) \exp(r(t)) + \tau_I$$

²³To simplify, we do not allow for short positions in the optimization problem. Therefore, we cannot use for Small and Large and Value and Growth the period of 1927 to 2013 as the period involves short positions in unconstrained optimization, which then imply one stock investments over extended periods of time under short selling restriction.

and the time- t dividend-tax-adjusted excess return of stock n is

$$R_n(t) = \frac{S_n(t+1)}{S_n(t)} [1 + \delta_n(t+1)(1 - \tau_D)] - R_f(t).$$

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

A.2.2 Trading Strategies

The investor enters time t with wealth $W(t)$, which is defined as time- t wealth before any capital gain taxes are paid, but after dividend and interest taxes are paid. As our model is homogeneous in wealth, we define all time- t dollar variables such as consumption, stock holdings, realized capital gains/losses, capital gain taxation, as a fraction of time- t wealth $W(t)$.

Entering time t , the investor holds an entering stock portfolio with weights $\{\underline{\pi}_n(t)\}_{n=1}^N$ and the remaining wealth is invested in the money market account. Exiting time t , the investor chooses his consumption $c(t)$ and an exiting stock portfolio, $\{\bar{\pi}_n(t)\}_{n=1}^N$, which will be held until time $t+1$. Rebalancing the stock portfolio from $\underline{\pi}(t)$ to $\bar{\pi}(t)$ is subject to capital gain taxation $\phi(t)$. We write the wealth growth rate between time t and time $t+1$ as

$$\frac{W(t+1)}{W(t)} = \sum_{n=1}^N \bar{\pi}_n(t) R_n(t) + [1 - c(t) - \phi(t)] R_f(t), \quad (\text{A.1})$$

where the investor's wealth after consumption and capital gain taxes $W(t)[1 - c(t) - \phi(t)]$ grows at the tax-adjusted risk-free rate $R_f(t)$ for one period and the stock positions contribute to growth through the dividend-tax-adjusted excess stock returns $\{R_n(t)\}_{n=1}^N$.

An admissible time- t trading strategy $\bar{\pi}(t)$ satisfies no short selling constraints

$$\bar{\pi}_n(t) \geq 0, n = 1, \dots, N \quad (\text{A.2})$$

and a margin constraint

$$(1 - m_+) \sum_{n=1}^N \bar{\pi}_n(t) \leq 1 - c(t) - \phi(t), \quad (\text{A.3})$$

where $1 - m_+$ denotes the fraction of equity that is marginable. Throughout, we use $m_+ = 0.5$ which is consistent with the Federal Reserve Regulation T for initial margins.

After choosing an admissible exiting portfolio $\bar{\pi}(t)$ at time t , the investor holds the portfolio for one period, which leads to the entering portfolio next period

$$\underline{\pi}_n(t+1) = \bar{\pi}_n(t) \frac{S_n(t+1)}{S_n(t)} \left(\frac{W(t+1)}{W(t)} \right)^{-1}, n = 1, \dots, N, \quad (\text{A.4})$$

where portfolio weights change from time t to $t+1$ due to stock returns over that period and changes in investor's wealth since portfolio weights are scaled by wealth.

A.2.3 Capital Gain Taxation and Limited Use of Capital Losses

Computing the capital gain taxes requires keeping track of the past purchase price of each stock. We define the nominal tax basis, $B_n(t)$, as the weighted-average purchase price of stock n after rebalancing at time t , and define the relative tax basis or basis-price ratio, $b_n(t)$, as $b_n(t) = B_n(t-1)/S_n(t)$, $n = 1, \dots, N$.

When stock n has an embedded capital gain ($b_n(t) < 1$), the investor can either increase or decrease his position, where reducing his position ($0 \leq \bar{\pi}_n(t) \leq \underline{\pi}_n(t)$) results in a realization of the embedded capital gain. When stock n has an embedded capital loss, ($b_n(t) > 1$), based on Gallmeyer and Srivastava (2011) the investor is always weakly better off realizing all capital losses today even if he cannot use them immediately. Thus, the total realized capital gains of all stocks net of immediately realized capital losses $g(t)$ is calculated as

$$g(t) = \sum_{n=1}^N (1 - b_n(t))^+ (\underline{\pi}_n(t) - \bar{\pi}_n(t))^+ - \sum_{n=1}^N (b_n(t) - 1)^+ \underline{\pi}_n(t), \quad (\text{A.5})$$

where $x^+ \triangleq \max(x, 0)$. The first summation in (A.5) aggregates all realized gains and the second summation aggregates all realized losses. The net realized capital gain g_t is the difference between these two summations.

In the FUL case, the investor either pays capital gain taxes if there is net capital gains ($g(t) > 0$) or collects tax rebates (negative taxes) if there is net capital losses ($g(t) < 0$). The amount of capital gain taxes/rebates at time t for the FUL case is calculated as

$$\phi^{FUL}(t) = \tau_C g(t), \quad (\text{A.6})$$

where τ_C is the capital gain tax rate.

In the LUL case, realized capital gains are subject to capital gain taxes in the same way as the FUL case. However, realized capital losses are treated differently. Instead of generating tax rebates, realized capital losses can only be used to offset current realized capital gains or be carried forward to offset future realized capital gains. As a result, an additional state variable is required: the carry-over loss $l(t) \geq 0$ which accumulates unused past-realized capital losses up to time t . Thus, the incurred capital gain taxes at time t for the LUL case is calculated as

$$\phi^{LUL}(t) = \tau_C (g(t) - l(t))^+. \quad (\text{A.7})$$

Any unused capital losses at time t are accumulated and carried forward to the next period in the carry-over loss account

$$l(t+1) = \left(\frac{W(t+1)}{W(t)} \right)^{-1} (l(t) - g(t))^+, \quad (\text{A.8})$$

where the wealth growth rate appears due to the fact that $l(t)$ and $l(t+1)$ are scaled by $W(t)$ and $W(t+1)$ respectively. When an investor dies, capital gain taxes are forgiven for both the FUL investor and the LUL investor and the tax basis resets to the current market price. This is consistent with the reset provision in the U.S. tax code.

The evolution of the basis-price ratio depends on the evolution of stock prices and investor's trading behaviour. If there is an embedded capital loss in stock n at time t , the investor immediately liquidates stock n . After liquidation the time- t basis-price ratio resets to $b_n(t) = 1$ because any new purchase of stock n is at current market price. The basis-price ratio next period simply reflects the return of stock n from time t to $t+1$; i.e., $b_n(t+1) = S_n(t)/S_n(t+1)$. If there is an embedded capital gain in stock n at time t and the investor reduces his position in stock n , the tax basis is unchanged and the return of stock n drives the basis-price ratio next period; i.e., $b_n(t+1) = [S_n(t)/S_n(t+1)] b_n(t)$. If the investor decides to purchase additional shares of stock n at time t while facing an embedded capital gain in stock n , the resulting position in stock n is a mixture of past-purchased shares and new-purchased shares. The past-purchased shares have weight $\underline{\pi}_n(t)$ and basis-price ratio $b_n(t)$ and the new-purchased shares have weight $\bar{\pi}_n(t) - \underline{\pi}_n(t)$ and basis-price ratio of 1. The combined position of stock n has a basis-price ratio that is the weighted-average of $b_n(t)$ and 1 adjusting for the stock

return $S_n(t)/S_n(t+1)$

$$b_n(t+1) = \frac{S_n(t)}{S_n(t+1)} \left[\frac{\underline{\pi}_n(t) b_n(t) + (\bar{\pi}_n(t) - \underline{\pi}_n(t))}{\bar{\pi}_n(t)} \right].$$

To summarize, the basis-price ratio evolves as

$$b_n(t+1) = \begin{cases} \frac{S_n(t)}{S_n(t+1)} & \text{if } \underline{\pi}_n(t) = 0 \text{ or } b_n(t) > 1, \\ \frac{S_n(t)}{S_n(t+1)} \left[\frac{\underline{\pi}_n(t)b_n(t) + (\bar{\pi}_n(t) - \underline{\pi}_n(t))^+}{\underline{\pi}_n(t) + (\bar{\pi}_n(t) - \underline{\pi}_n(t))^+} \right] & \text{otherwise,} \end{cases}, n = 1, \dots, N. \quad (\text{A.9})$$

A.2.4 Bellman Equation

The investor's objective is to maximize his expected lifetime utility from real consumption and terminal wealth at the time of death by choosing an admissible trading strategy given an initial endowment. We assume that the investor and his heirs have identical preferences²⁴ of the constant relative risk aversion (CRRA) form with a coefficient of relative risk aversion γ and a common time preference parameter β . Using the principle of dynamic programming, we can describe the investor's lifetime consumption and portfolio choice problem in a recursive form as the Bellman equation.

The Bellman equation for the FUL investor is

$$V(t, \underline{\pi}(t), b(t)) = \max_{c(t), \bar{\pi}(t)} e^{-\lambda t} \frac{c(t)^{1-\gamma}}{1-\gamma} + (1 - e^{-\lambda t}) \frac{\alpha}{1-\gamma} + e^{-\lambda t} \beta E_t \left[\left(\frac{e^{-i(t)} W(t+1)}{W(t)} \right)^{1-\gamma} V(t+1, \underline{\pi}(t+1), b(t+1)) \right] \quad (\text{A.10})$$

for $t = 0, 1, \dots, T-1$ subject to the wealth evolution equation (A.1), the stock proportion dynamics (A.4), the basis-price ratio evolution equation (A.9), the no short selling constraints (A.2), the margin constraint (A.3), the realized capital gain equation (A.5), and the FUL capital gain taxation equation (A.6). Note that $\underline{\pi}(t)$, $\bar{\pi}(t)$, and $b(t)$ are vectors of length N that capture the time- t entering position, the exiting position, and the basis-price ratio for each stock. Thus, there are $2 \times N + 1$ state variables in the FUL case including the time dimension.

For the LUL investor, an additional state variable is needed, $l(t)$, the carry-over loss up to time t . The Bellman equation for the LUL investor is

$$V(t, \underline{\pi}(t), b(t), l(t)) = \max_{c(t), \bar{\pi}(t)} e^{-\lambda t} \frac{c(t)^{1-\gamma}}{1-\gamma} + (1 - e^{-\lambda t}) \frac{\alpha}{1-\gamma} + e^{-\lambda t} \beta E_t \left[\left(\frac{e^{-i(t)} W(t+1)}{W(t)} \right)^{1-\gamma} V(t+1, \underline{\pi}(t+1), b(t+1), l(t+1)) \right] \quad (\text{A.11})$$

for $t = 0, 1, \dots, T-1$ subject to the wealth evolution equation (A.1), the stock proportion dynamics (A.4), the basis-price ratio evolution equation (A.9), the carry-over loss evolution equation (A.8), the no short selling constraints (A.2), the margin constraint (A.3), the realized capital gain equation (A.5), and the LUL capital gain taxation equation (A.7). Note that the value function $V(t)$ defined in equation (A.10) and (A.11) is scaled by $W(t)^{1-\gamma}$. This consistent with the fact that $c(t)$ is scaled by $W(t)$ and the investor has a CRRA preference with a relative risk aversion γ .

If the investor survives at time t , with conditional probability $e^{-\lambda t}$, he chooses the consumption and investment plan to maximize the sum of the utility from current consumption $c(t)$ and the expected future utility measured by the value function $V(t+1)$ adjusting for inflation $i(t)$, time discount β ,

²⁴ Assuming different weights on investor's own consumption and his heirs' consumption is equivalent to changing the number of bequest years.

and the wealth growth rate $W(t+1)/W(t)$.

If death occurs at time t , with conditional probability $1 - e^{-\lambda t}$, the investor's assets totaling $W(t)$ are liquidated and distributed to the investor's heirs as an equal amount of payment, p , each year in real terms for η periods starting from time t , where p is calculated as

$$W(t) = p + e^{\bar{i}-\bar{r}}p + e^{2(\bar{i}-\bar{r})}p + \dots + e^{(\eta-1)(\bar{i}-\bar{r})}p \Rightarrow p = \frac{1 - e^{\bar{i}-\bar{r}}}{1 - e^{(\bar{i}-\bar{r})\eta}} W(t).$$

Thus, the total bequest utility of those η payments is

$$u(p) + \beta u(p) + \beta^2 u(p) + \dots + \beta^{\eta-1} u(p) = \alpha \cdot u(W(t))$$

and the bequest factor α is

$$\alpha = \frac{1 - \beta^\eta}{1 - \beta} \left(\frac{1 - e^{\bar{i}-\bar{r}}}{1 - e^{(\bar{i}-\bar{r})\eta}} \right)^{1-\gamma},$$

where \bar{i} and \bar{r} are inflation rate and after-tax interest rate that apply to the η periods of bequest. Throughout, we use $\eta = 20$ years and \bar{i} and \bar{r} equal to their historical averages from 1927 to 2013.

B Numerical Methodology

To numerically solve the Bellman equations (A.10) and (A.11), we extend the methodology of Brandt et al. (2005) and Garlappi and Skoulakis (2008) to incorporate endogenous state variables and constraints on portfolio weights. In addition, since the state variable evolution is given by functions with non-differentiable kinks, the Bellman equation corresponds to a singular stochastic control problem that we solve employing a domain decomposition of the state space. We first briefly sketch the algorithm before providing additional details. A full description can be found in Yang and Tompaidis (2013).

B.1 Sketch of Algorithm

Step 1 - Domain Decomposition

- a. The state space is decomposed into degenerate and non-degenerate regions. The degenerate region corresponds to when a stock's basis-price ratio is above 1. The solution at a point in the degenerate region is mapped to a solution at a point in the non-degenerate region.
- b. For a point in the non-degenerate region, the choice space is decomposed by a partition so that, in each subset of the partition, the evolution rules of all state variables are differentiable.

Step 2 - Dynamic Programming

- a. For each time step, starting at the terminal time and working backward, a quasi-random grid is constructed in the non-degenerate region of the state space. For each point on the grid, the value function, the optimal consumption, and the optimal portfolio decisions are computed.
- b. The value function is approximated using a set of basis functions, consisting of radial basis functions and low order polynomials. This approximation is used in earlier time steps to compute conditional expectations of the value function.

Step 3 - Karush-Kuhn-Tucker (KKT) Conditions

To solve the Bellman equation for each point on the quasi-random grid in the non-degenerate region and for each partition in the choice space, the following steps are performed.

- a. A Lagrangian function is constructed for the value function using the portfolio position constraints, the corresponding Lagrange multipliers, and the state variable evolution.
- b. The system of first order conditions (KKT conditions) are constructed from the Lagrangian function.
- c. The solution to the KKT conditions is found using a double iterative process:
 - i. An approximate optimal portfolio is chosen and the corresponding approximate optimal consumption is computed.
 - ii. Given the approximate optimal consumption, the corresponding approximate optimal portfolio is updated by solving the system of KKT conditions, where the conditional expectations in the KKT conditions are approximated using a cross-test-solution regression:
 1. A quasi-random set of feasible allocations and consumptions is chosen.
 2. For each feasible choice, the required conditional expectations are computed using the approximate value function from the next time step that was already computed.
 3. For each feasible choice, the computed conditional expectations are projected on a set of basis functions of the choice variables.
 4. The resulting system of approximated KKT conditions is then solved.
 - iii. The consumption choice is then updated to the choice corresponding to the new approximate optimal portfolio.
 - iv. Step (ii) and (iii) are repeated until the consumption and portfolio choices converge. Each repetition employs a smaller region in which feasible portfolio choices are drawn. The region is chosen based on the location of the previously computed approximate optimal portfolio. This is the *test region contraction* step.

We now provide a more detailed description of each step of the algorithm using the case of limited use of losses as an example. The case of full use of losses is similar but less complicated.

B.2 Algorithm Step 1 - Domain Decomposition

The first step in the domain decomposition is to decompose the state space into a degenerate and a non-degenerate region. The solution at any point in the degenerate region can be mapped to the solution at a point in the non-degenerate region, and the problem is only solved over the non-degenerate region. The degeneracy arises when the basis-price ratio of a stock is above 1, in which case it is optimal to immediately liquidate the position and add the realized capital loss to the carry-over loss account. Given any point in the state space $(\hat{\pi}(t) \in \mathbb{R}_+^N, \hat{b}(t) \in \mathbb{R}_+^N, \hat{l}(t) \in \mathbb{R}_+)$, we can define the following sets

$$\begin{aligned}
 \text{Index set of all risky assets:} & \quad \Omega = \{1, \dots, N\} \\
 \text{Index set of all degenerate assets:} & \quad \Omega_t^D = \left\{ n \in \Omega : \hat{b}_n(t) > 1 \right\} \\
 \text{Index set of all non-degenerate assets:} & \quad \Omega_t^{\bar{D}} = \left\{ n \in \Omega : \hat{b}_n(t) \leq 1 \right\}
 \end{aligned}$$

and identify an equivalent point $(\underline{\pi}(t), b(t), l(t))$ in the non-degenerate region of the state space, such that

$$\begin{aligned} V(t, \underline{\pi}(t), b(t), l(t)) &= V\left(t, \dot{\underline{\pi}}(t), \dot{b}(t), \dot{l}(t)\right) \\ \bar{\pi}^*(t, \underline{\pi}(t), b(t), l(t)) &= \bar{\pi}^*\left(t, \dot{\underline{\pi}}(t), \dot{b}(t), \dot{l}(t)\right) \\ c^*(t, \underline{\pi}(t), b(t), l(t)) &= c^*\left(t, \dot{\underline{\pi}}(t), \dot{b}(t), \dot{l}(t)\right) \end{aligned}$$

where

$$\underline{\pi}_n(t) = \begin{cases} 0 & \text{if } n \in \Omega_t^D \\ \dot{\underline{\pi}}_n(t) & \text{if } n \in \Omega_t^{\bar{D}} \end{cases}, \quad b_n(t) = \begin{cases} 1 & \text{if } n \in \Omega_t^D \\ \dot{b}_n(t) & \text{if } n \in \Omega_t^{\bar{D}} \end{cases}, \quad l(t) = \dot{l}(t) + \sum_{n \in \Omega_t^D} (\dot{b}_n(t) - 1) \dot{\underline{\pi}}_n(t)$$

The second step employed in the domain decomposition is to partition the choice space at each point in the non-degenerate region such that, in each subset of the partition, the evolution rules of all state variables are differentiable. This is achieved by choosing the partition $(\Omega_t^{RP}, \Omega_t^{IP})$, where

$$\begin{aligned} \text{Index set of all non-degenerate reduce-position assets: } \quad \Omega_t^{RP} &= \left\{ n \in \Omega_t^{\bar{D}} : \bar{\pi}_n(t) \leq \underline{\pi}_n(t) \right\} \\ \text{Index set of all non-degenerate increase-position assets: } \quad \Omega_t^{IP} &= \left\{ n \in \Omega_t^{\bar{D}} : \bar{\pi}_n(t) > \underline{\pi}_n(t) \right\} \end{aligned}$$

To find the optimal solution for each point in the non-degenerate region of the state space, we solve for each possible partition of the choice space and choose the solution with the higher value of objective.

B.3 Algorithm Step 2 - Dynamic Programming

The solution methodology is based on dynamic programming that solves the Bellman equation (A.11) backward in time. At each time, the non-degenerate region of the state space is discretized by a grid. For the LUL case, there are five state variables at each time. The problem of discretizing a high-dimensional space using a regular grid is that the number of grid points increases exponentially with the number of dimensions. To overcome this problem we employ the quasi-random grid method for discretization. Quasi-random grids are often used in the literature for sampling high-dimensional spaces, have well-understood properties, and several algorithms for their generation exist (see Press et al. (2007)).

Once the optimal strategy and the value function levels are computed for all points in the quasi-random grid at a particular time, the value function for any point in the state space is approximated by projecting the values on a set of basis functions. Some form of approximation is necessary, since we need to estimate the value function at arbitrary points in the state space in order to compute the conditional expectations that arise naturally when the optimization problem is solved at grid points in the previous time slice. In the literature different approximations have been used, including a linear rule (see Gallmeyer, Kaniel and Tompaidis (2006)) and projection on polynomials of the state variables (see Brandt et al. (2005)). We choose an approximation scheme that proceeds in two steps. First, we project the value function on a set of low order polynomials of the state variables. Second, we approximate the residuals with a set of radial basis functions. Each radial basis function is defined by its weight, center, and width. We adjust the number of centers, the location of each center, the corresponding widths, and the corresponding weights to achieve a good approximation of the value function. Additional details of the radial basis function approximation are in Fasshauer (2007).

B.4 Algorithm Step 3 - Karush-Kuhn-Tucker (KKT) Conditions

To derive the KKT conditions at each grid point, we construct the Lagrangian function that combines the objective function and the constraints on the choice variables. Since the LUL capital gain taxation equation (A.7) and the carry-over loss evolution equation (A.8) are non-differentiable when $l(t) = g(t)$, it is necessary to write two versions of the Lagrangian and solve them separately depending on whether $g(t) \geq l(t)$ or $g(t) \leq l(t)$. Assuming $g(t) \geq l(t)$ at $(\underline{\pi}(t), b(t), l(t))$, the Lagrangian is

$$\begin{aligned} \mathcal{L}(c_t, \bar{\pi}(t), \lambda_t^g, \lambda_t^+, \lambda_t^m, \lambda_t^{RP}, \lambda_t^{IP}) &= e^{-\lambda_t} \frac{c(t)^{1-\gamma}}{1-\gamma} \\ &+ e^{-\lambda_t} \beta E_t \left[\left(\frac{e^{-i(t)} W(t+1)}{W(t)} \right)^{1-\gamma} V(t+1, \underline{\pi}(t+1), b(t+1), l(t+1)) \right] \\ &+ \lambda_t^g [g(t) - l(t)] + \sum_{n \in \Omega} \lambda_t^{+,n} \bar{\pi}_n(t) + \lambda_t^m \left[1 - c(t) - \phi^{LUL}(t) - (1 - m_+) \sum_{n \in \Omega} \bar{\pi}_n(t) \right] \\ &+ \sum_{n \in \Omega_t^{RP}} \lambda_t^{RP,n} [\underline{\pi}_n(t) - \bar{\pi}_n(t)] + \sum_{n \in \Omega_t^{IP}} \lambda_t^{IP,n} [\bar{\pi}_n(t) - \underline{\pi}_n(t)], \end{aligned} \quad (\text{B.1})$$

where the Lagrangian is a function of choice variables and Lagrange multipliers; λ_t^g is the Lagrange multiplier corresponding to the constraint $g(t) \geq l(t)$; λ_t^+ are the Lagrange multipliers corresponding to the no short selling constraints; λ_t^m is the Lagrange multiplier corresponding to the margin constraint; and $\lambda_t^{RP}, \lambda_t^{IP}$ are the Lagrange multipliers corresponding to the partitioning of the choice variable space. Assuming $g(t) \leq l(t)$ at $(\underline{\pi}(t), b(t), l(t))$, the Lagrangian is the same as (B.1) except that $g(t) - l(t)$ is replaced by $l(t) - g(t)$.

The KKT conditions are derived by differentiating the Lagrangian (B.1) with respect to the choice variables and Lagrange multipliers, which requires the partial derivatives $\partial/\partial \bar{\pi}_n(t), n = 1, \dots, N$ and $\partial/\partial c(t)$ of the conditional expectation

$$E_t [A | \underline{\pi}(t), b(t), l(t), c_t, \bar{\pi}(t)], \quad (\text{B.2})$$

where

$$A = \left(\frac{e^{-i(t)} W(t+1)}{W(t)} \right)^{1-\gamma} V(t+1, \underline{\pi}(t+1), b(t+1), l(t+1)).$$

To estimate (B.2) at $(\underline{\pi}(t), b(t), l(t))$, we generate a set of test solutions for the choice variables, $(c^{(j)}(t), \bar{\pi}^{(j)}(t)), j = 1, \dots, J$, where J is the number of test solutions, and calculate the conditional expectation for each test solution

$$E_t \left[A | \underline{\pi}(t), b(t), l(t), c^{(j)}(t), \bar{\pi}^{(j)}(t) \right].$$

The test solutions need to be generated consistently with the partition of the choice space in which the problem is solved. The J values of the conditional expectation are projected onto a set of basis functions $f_k(\cdot, \cdot), k = 1, \dots, K$, where K is the number of basis functions, using OLS regression:

$$E_t \left[A | \underline{\pi}(t), b(t), l(t), c^{(j)}(t), \bar{\pi}^{(j)}(t) \right] = \sum_{k=1}^K \omega_k f_k \left(c^{(j)}(t), \bar{\pi}^{(j)}(t) \right) + \varepsilon_j, j = 1, \dots, J$$

The estimated coefficients $\{\omega_k\}_{k=1}^K$ are used to approximate the conditional expectation (B.2) and its

partial derivatives for an arbitrary choice of $c(t)$ and $\bar{\pi}(t)$ at $(\underline{\pi}(t), b(t), l(t))$

$$\begin{aligned}
E_t [A | \underline{\pi}(t), b(t), l(t), c_t, \bar{\pi}(t)] &\approx \sum_{k=1}^K \omega_k f_k(c_t, \bar{\pi}(t)), \\
\frac{\partial}{\partial \bar{\pi}_n(t)} E_t [A | \underline{\pi}(t), b(t), l(t), c_t, \bar{\pi}(t)] &\approx \sum_{k=1}^K \omega_k \frac{\partial}{\partial \bar{\pi}_n(t)} f_k(c_t, \bar{\pi}(t)), n = 1, \dots, N, \\
\frac{\partial}{\partial c(t)} E_t [A | \underline{\pi}(t), b(t), l(t), c_t, \bar{\pi}(t)] &\approx \sum_{k=1}^K \omega_k \frac{\partial}{\partial c(t)} f_k(c_t, \bar{\pi}(t)).
\end{aligned}$$

The choice of basis functions $\{f_k(\cdot, \cdot)\}_{k=1}^K$ is not unique. We use polynomials up to the order of two; i.e., quadratic functions. As a result, the partial derivatives of the conditional expectation in the KKT conditions become linear functions of the choice variables. This setup eliminates the possibility of having multiple solutions to the KKT conditions, which is often the case when high-order polynomials are used as basis functions. To account for the inaccuracy in approximating conditional expectations with quadratic functions, we use an iterative scheme, where we successively reduce the size of the region from which the test solutions are drawn.

References

- Brandt, M.W., Goyal, A., Santa-Clara, P., Stroud, J.R., 2005. A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Review of Financial Studies* 18, 831–873.
- Constantinides, G., 1983. Capital market equilibrium with personal tax. *Econometrica* 51, 611–636.
- Dammon, R., Spatt, C., 1996. The optimal trading and pricing of securities with asymmetric capital gains taxes and transaction costs. *Review of Financial Studies* 9, 921–952.
- Dammon, R., Spatt, C., Zhang, H., 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, 583–616.
- Dammon, R., Spatt, C., Zhang, H., 2004. Optimal asset location and allocation with taxable and tax-deferred investing. *Journal of Finance* 59, 999–1037.
- DeMiguel, V., Uppal, R., 2005. Portfolio investment with the exact tax basis via nonlinear programming. *Management Science* 55, 277–290.
- Dybvig, P., Koo, H.K., 1996. Investment with taxes. Washington University.
- Fasshauer, G.E., 2007. Meshfree Approximation Methods with MATLAB. Singapore: World Scientific Publishing Company.
- Gallmeyer, M., Kaniel, R., Tompaidis, S., 2006. Tax management strategies with multiple risky assets. *Journal of Financial Economics* 80, 243–291.
- Gallmeyer, M., Srivastava, S., 2011. Arbitrage and the tax code. *Mathematics and Financial Economics* 4, 183–221.
- Garlappi, L., Naik, L.V., Slive, J., 2001. Portfolio selection with multiple assets and capital gains taxes. University of British Columbia.
- Garlappi, L., Skoulakis, G., 2008. A state-variable decomposition approach for solving portfolio choice problems. University of British Columbia.
- Hur, S.K., 2001. Optimal portfolio selection with personal tax. University of Chicago.
- Jensen, B.A., Marekwica, M., 2011. Optimal portfolio choice with wash sale constraints. *Journal of Economic Dynamics and Control* 35, 1916–1937.
- Marekwica, M., 2012. Optimal tax-timing and asset allocation when tax rebates on capital losses are limited. *Journal of Banking & Finance* 36, 2048–2063.

- Poterba, J.M., Shoven, J.B., 2002. Exchange-traded funds: A new investment option for taxable investors. *American Economic Review* 92, 422–427.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). Cambridge University Press.
- Sialm, C., 2009. Tax changes and asset pricing. *American Economic Review* 99, 1356–1383.
- Yang, C., Tompaidis, S., 2013. An iterative simulation approach for solving stochastic control problems in finance. BI Norwegian Business School and University of Texas at Austin.

Figure 2: **Example.** The figure reports the optimal trading strategy for an LUL, an FUL, and an NCGT investor as a function of the investor's basis-to-price ratio, $b(0)$, assuming the investor owns one share of stock at time $t = 0$. The left panel summarizes after-tax optimal portfolio choices as a fraction of wealth $\bar{\pi}$ at $t = 0$ and $t = 1$ and wealth $W(0)$. The middle panel shows capital gain taxes paid Φ_{CG} at $t = 0$ and $t = 1$ as well as the investor's expected utility at $t = 0$. The right panel summarizes capital gain taxes paid at $t = 2$ when the investor consumes. 'Up' and 'Dn' denote up and down moves through the binomial tree. Parameters are given in Section 3.

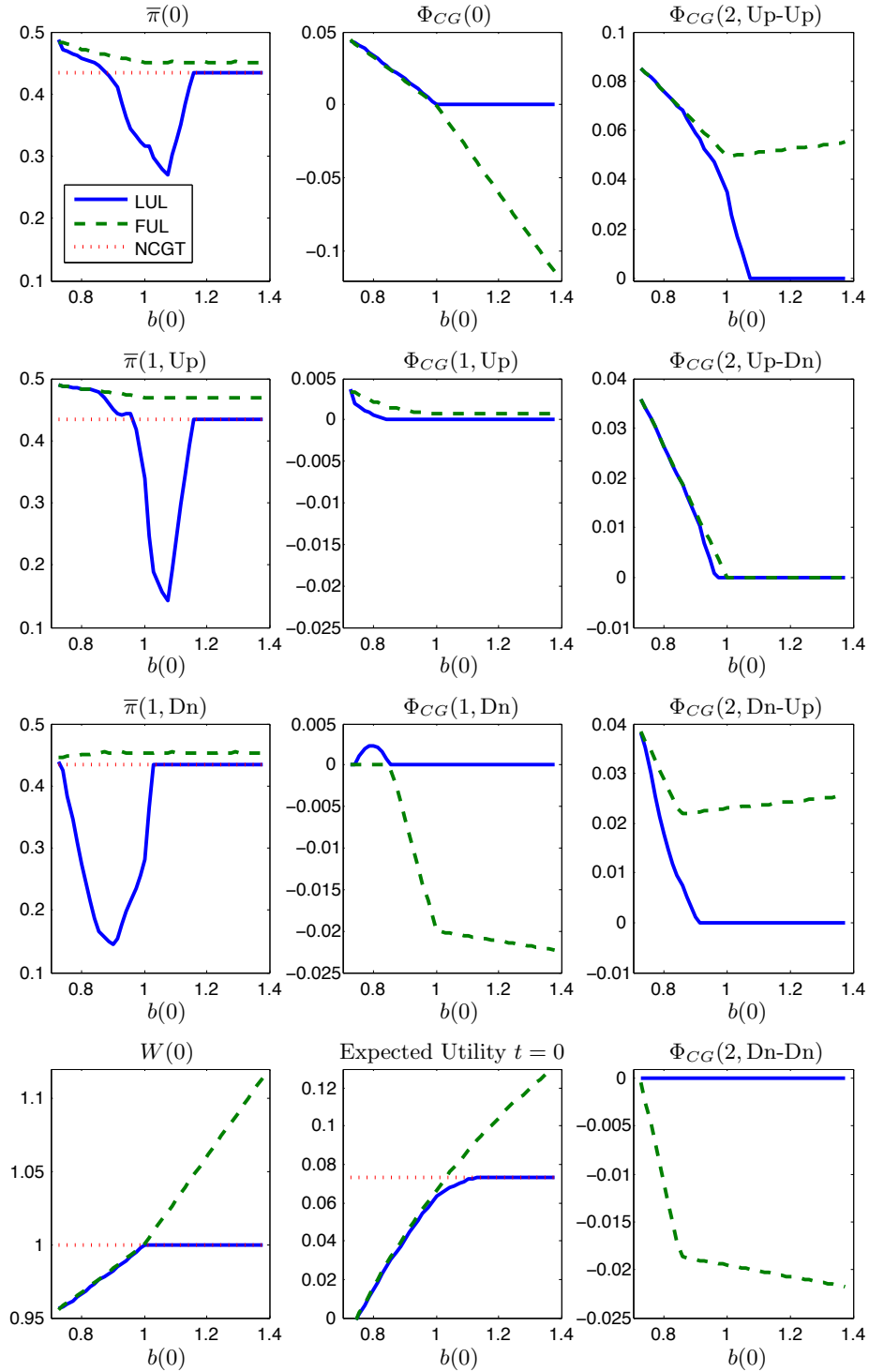


Figure 3: **Backtesting Tax Trading Strategies using the S&P 500 Index.** The figure reports backtests of the optimal trading strategy, expressed as an equity-to-wealth ratio, for an FUL, an LUL, and an NCGT investor. Equity-to-wealth ratios are computed by dividing the value of the equity position by the value of the investor’s wealth after subtracting dividend and interest taxes, but before consumption and capital gains taxes. The investor’s tax trading strategy along the historical path of the S&P 500 Index assumes that the backtests start at the beginning of the first year of each backtest window when the investor is 20 years old with initial stock position of 0, initial tax basis of 1, and initial carry-over loss of 0. The top left and middle left plots show allocations over 30 years with 31 optimal trading strategies starting in 1929 and 1965, respectively. The top right, middle right, and bottom right plots show allocations over 15 years with 16 optimal trading strategies starting in 1950, 1980, and 1999, respectively. The bottom left plot shows the equity-to-wealth ratio averaged over 58 investors who start investing in 1927, 1928, ..., and 1984, respectively, when they are 20 years old. The 58’th investor enters the equity market in 1984 at age 20 and reaches age 50 in 2014. Tax rates are set to match the U.S. tax code (see Subsection 4.1). Optimal portfolios conditional on the beginning period equity-to-wealth and basis-to-price ratios and on age implied by S&P 500 Index returns over the period of 1927 to 2013 are summarized in the Internet Appendix.

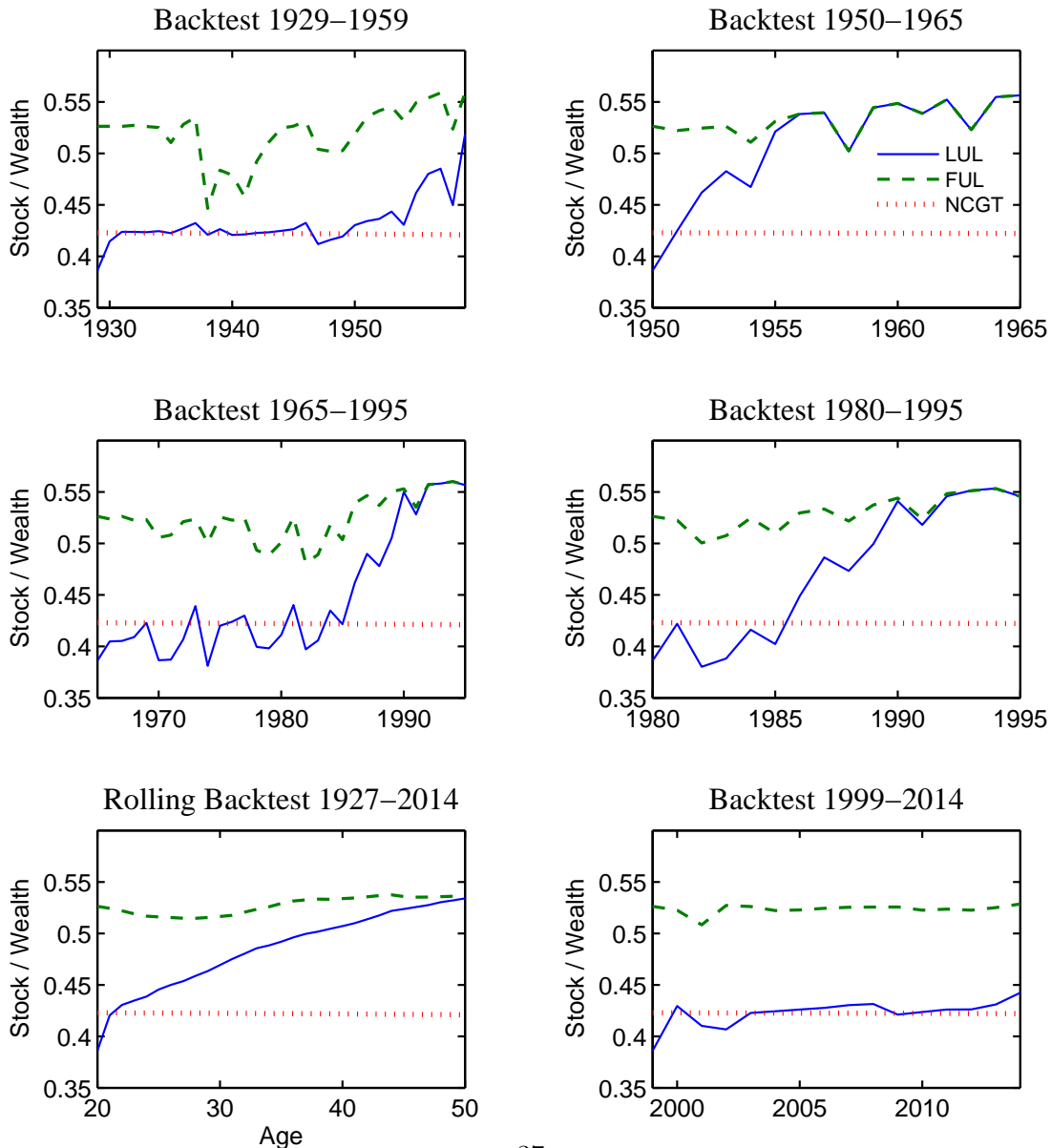


Figure 4: **Notrade Regions using the S&P 500 Index.** The figure reports entering and exiting equity-to-wealth ratios for an FUL and an LUL investor for the backtested periods starting in 1950 and 1999 (see Figure (3)) conditional on the basis-to-price ratio. The area inside the dark (light) solid blue line shows the notrade region for the LUL investor at age 20 (80). The area inside the dark (light) green dashed line shows the notrade region for the FUL investor at age 20 (80). The optimal strategy implies that above (below) the notrade region investors sell (buy) equity shares to trade onto the boundary of the notrade region. The right plots showing the backtest starting in 1999, present the notrade region for the LUL investor at age 20 conditional on a carry-over-loss to wealth ratio of 0 (blue solid line), 0.01 (red dashed line), and 0.1 (beige dotted line), respectively.

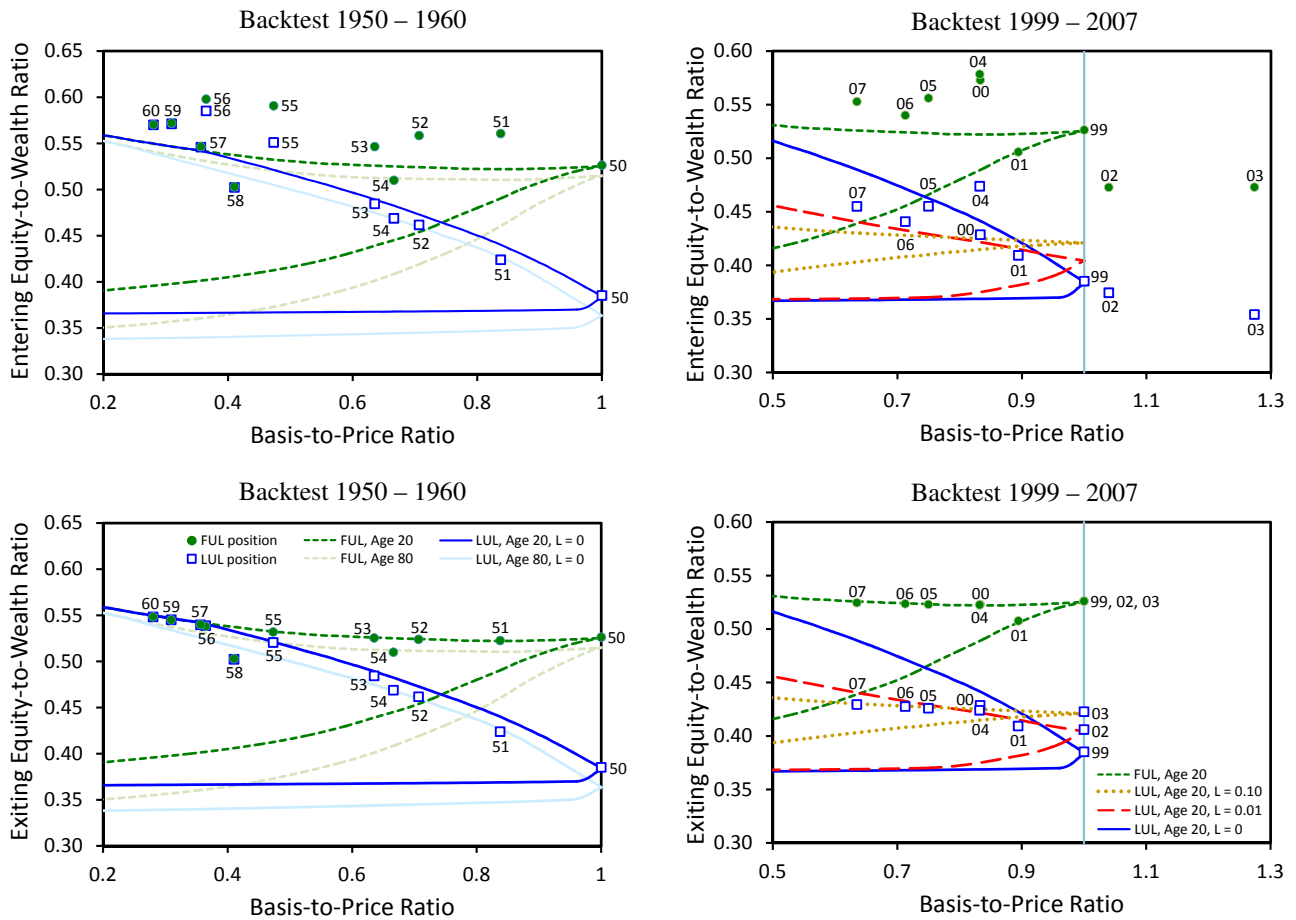


Figure 5: **Taxes and Trading using the S&P 500 Index.** The figure reports backtests of the tax basis, carry-over-loss-to-wealth ratio, stock shares, and cumulative capital gain taxes for an FUL, an FUL(LUL) (not shown when identical to FUL), an LUL, and an NCGT investor implied by the tax trading strategy for the periods starting in 1950 and 1999 (see Figure (3)). Figures corresponding to the other backtested periods shown in Figure (3) are delegated to the Internet Appendix.

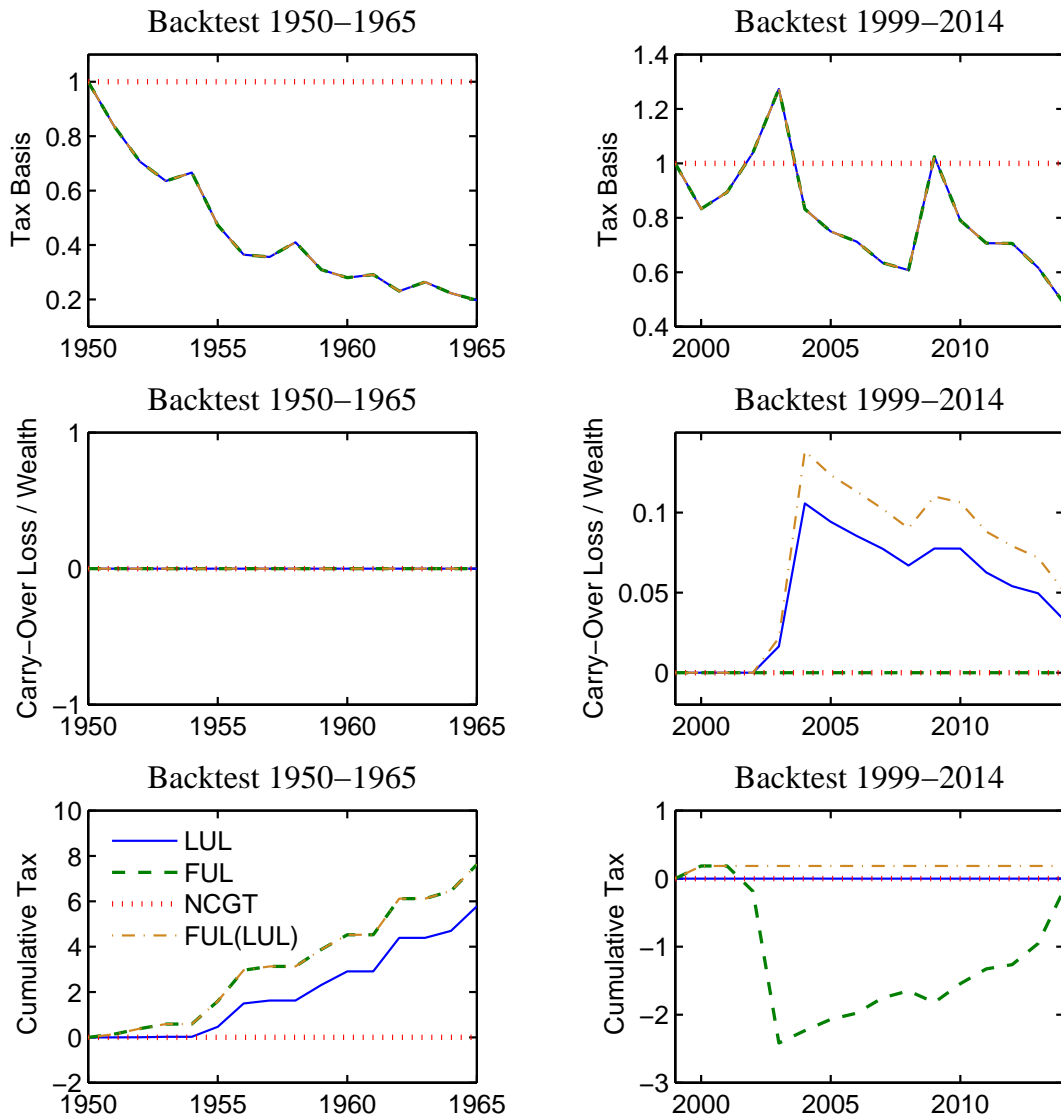


Figure 6: **Consumption and Wealth implied by Tax Trading Strategies using the S&P 500 Index.** The figure reports backtests of wealth, consumption, wealth scaled by NCGT wealth and consumption scaled by NCGT consumption, and GARCH(1,1) wealth and consumption volatility of an FUL, an FUL(LUL) (not shown when identical to FUL), an LUL, and an NCGT investor implied by the tax trading strategy for the period starting in 1999 (see Figure (3)). Figures corresponding to the other backtested periods shown in Figure (3) are delegated to the Internet Appendix.

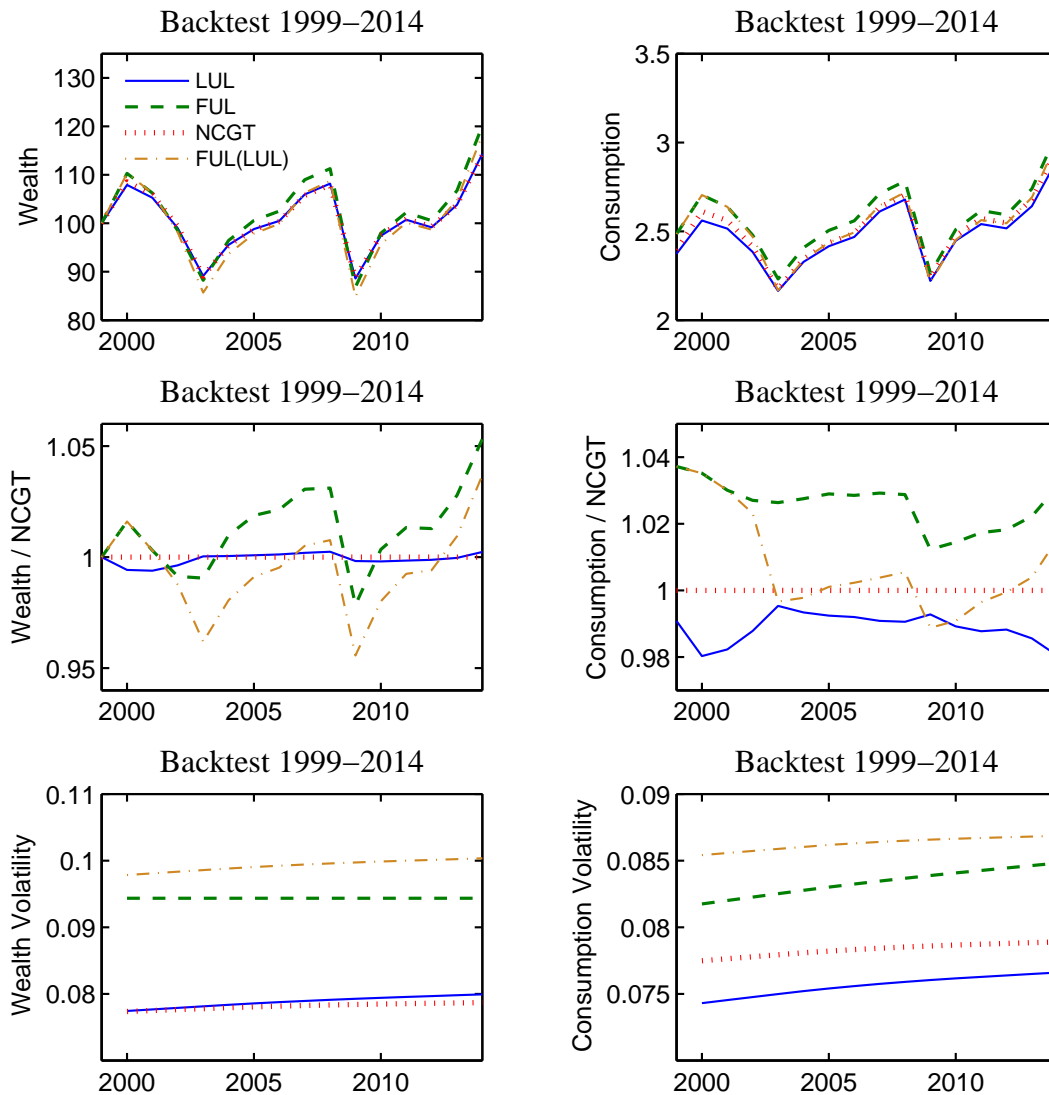


Figure 7: **The Costs of the Tax Treatment of Capital Losses.** The figure reports the expected cumulative utility and the wealth equivalent scaled by the wealth equivalent of the NCGT investor for an FUL, an FUL(LUL), an LUL, and an NCGT investor. The expected cumulative utility and the wealth equivalent are averaged over 50,000 simulation paths sampled from the S&P 500 Index. Each path has 80 periods (age 20 to 100). Tax rates are set to match the U.S. tax code (see Subsection 4.1).

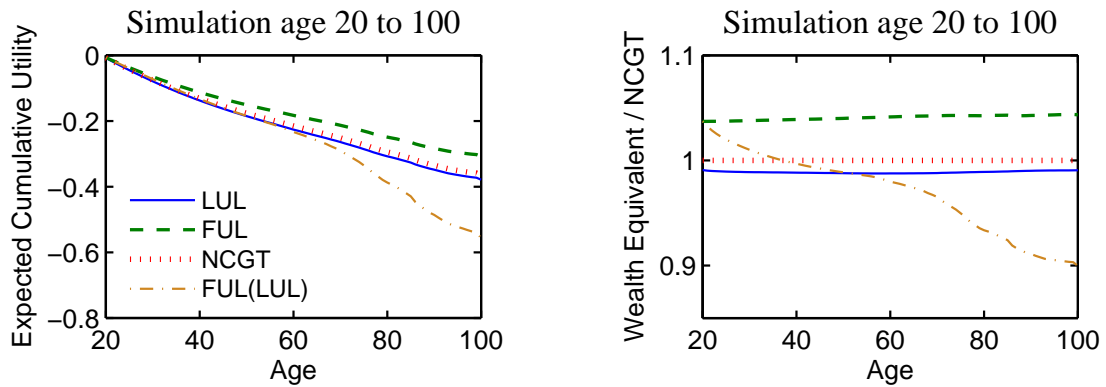


Figure 8: **Backtesting Tax Trading Strategies using Small and Large I.** The figure reports backtests of the optimal trading strategies (Small-to-wealth and Large-to-wealth plus total equity-to-wealth ratios), optimal tax bases, carry-over-loss-to-wealth ratio, stock shares, and cumulative capital gain taxes for an FUL, an FUL(LUL) (not shown when identical to FUL), an LUL, and an NCGT investor. Equity-to-wealth ratios are computed by dividing the value of the equity position by the value of the investor's wealth after subtracting dividend and interest taxes, but before consumption and capital gains taxes. The investor's tax trading strategy along the historical path of Small stocks and Large stocks assumes that the backtests start at the beginning of 1965 when the investor is 20 years old with initial stock position of 0, initial tax basis of 1, and initial carry-over loss of 0. Additional backtested optimal Small and Large strategies are delegated to the Internet Appendix. Tax rates are set to match the U.S. tax code (see Subsection 4.1).

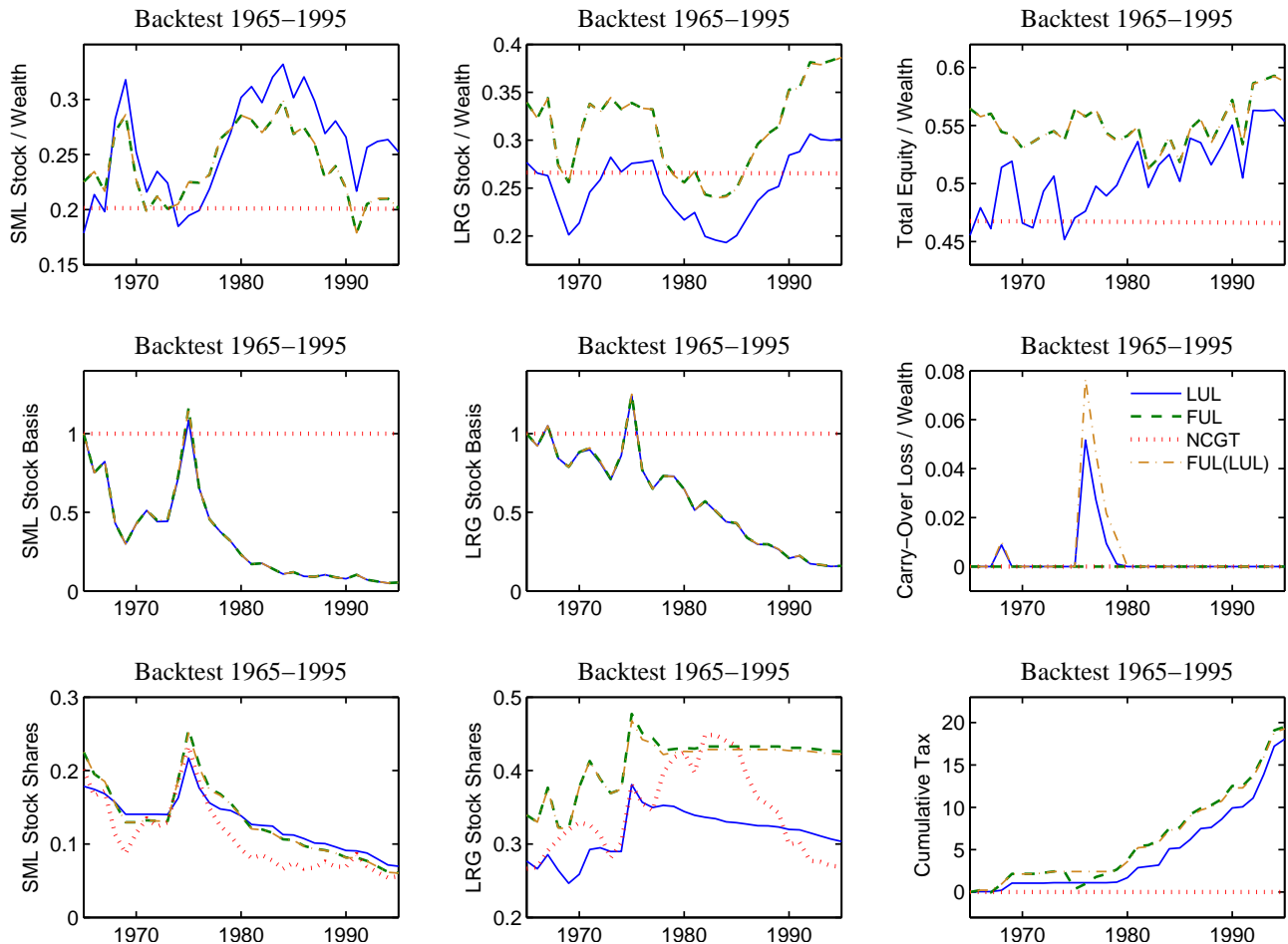


Figure 9: **Backtesting Tax Trading Strategies using Small and Large II.** The figure reports backtests of the optimal trading strategies (Small-to-wealth and Large-to-wealth plus total equity-to-wealth ratios), optimal tax bases, carry-over-loss-to-wealth ratio, stock shares, and cumulative capital gain taxes for an FUL, an FUL(LUL) (not shown when identical to FUL), an LUL, and an NCGT investor. Equity-to-wealth ratios are computed by dividing the value of the equity position by the value of the investor’s wealth after subtracting dividend and interest taxes, but before consumption and capital gains taxes. The investor’s tax trading strategy along the historical path of Small stocks and Large stocks assumes that the backtests start at the beginning of 1999 when the investor is 20 years old with initial stock position of 0, initial tax basis of 1, and initial carry-over loss of 0. Additional backtested optimal Small and Large strategies are delegated to the Internet Appendix. Tax rates are set to match the U.S. tax code (see Subsection 4.1).

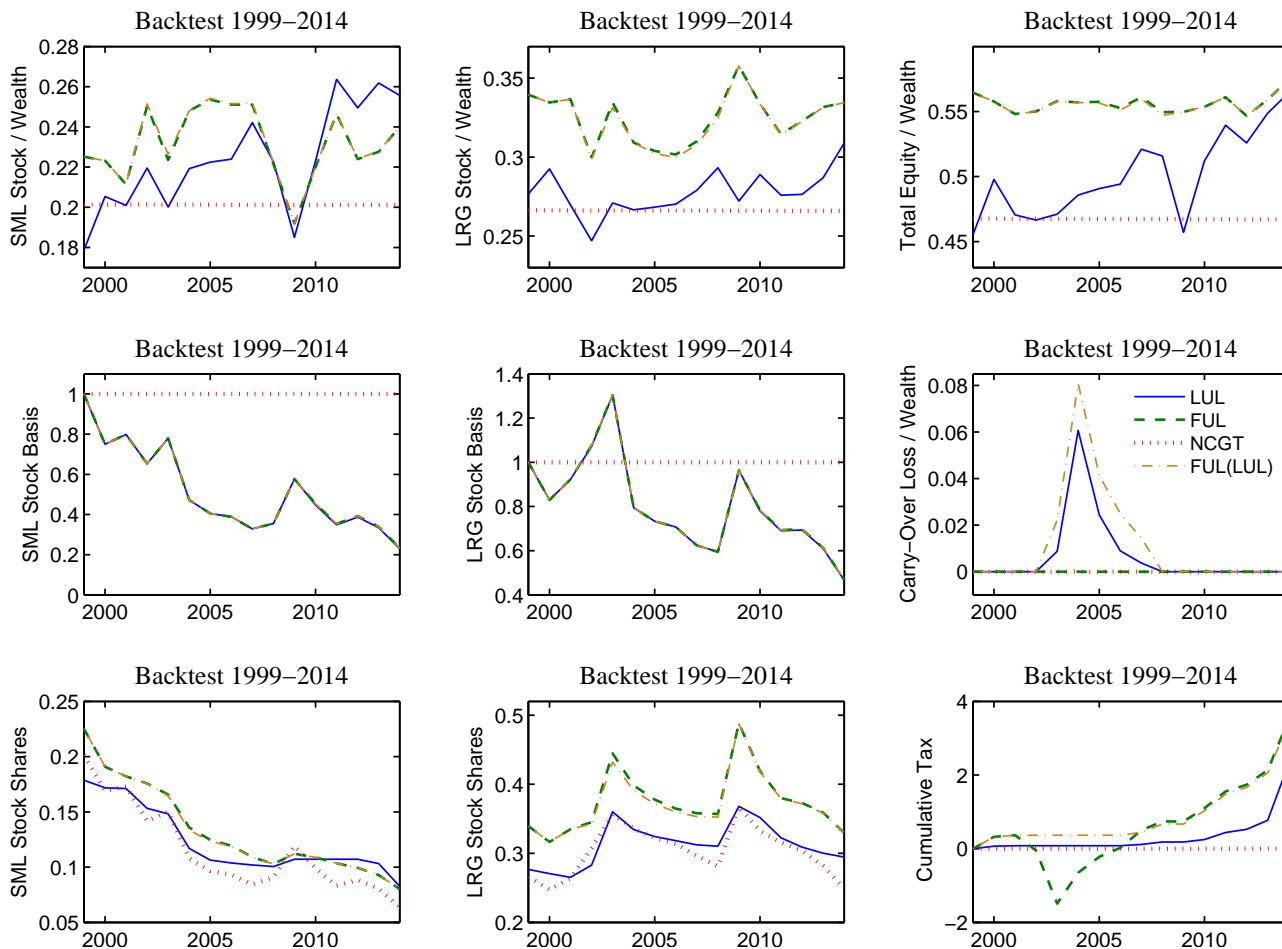


Figure 10: **Backtesting Tax Trading Strategies using Value and Growth.** The figure reports backtests of the optimal trading strategies (Value-to-wealth and Growth-to-wealth plus total equity-to-wealth ratios), optimal tax bases, carry-over-loss-to-wealth ratio, stock shares, and cumulative capital gain taxes for an FUL, an FUL(LUL) (not shown when identical to FUL), an LUL, and an NCGT investor. Equity-to-wealth ratios are computed by dividing the value of the equity position by the value of the investor’s wealth after subtracting dividend and interest taxes, but before consumption and capital gains taxes. The investor’s tax trading strategy along the historical path of Value and Growth assumes that the backtests start at the beginning of 1999 when the investor is 20 years old with initial stock position of 0, initial tax basis of 1, and initial carry-over loss of 0. Tax rates are set to match the U.S. tax code (see Subsection 4.1).

