

Tax Collection from Realized Capital Gains on Equity*

Paul Ehling[†]

BI Norwegian Business School

Stathis Tompaidis[‡]

McCombs School of Business

University of Texas at Austin

Chunyu Yang[§]

BI Norwegian Business School

Current Draft: July 2019

*We thank Harjoat Bhamra (discussant), Richard Priestley, and seminar participants at a Research Day and a lunch seminar of the Finance Department at BI Norwegian Business School, Deakin Business School, Monash Business School, 1st Asset Pricing Conference at the Collegio Carlo Alberto, and Université Paris-Dauphine for helpful comments. Ehling and Yang acknowledge funding support from the Centre for Asset Pricing Research (CAPR) at BI Norwegian Business School. We thank the Texas Advanced Computing Center (TACC) at the University of Texas at Austin (www.tacc.utexas.edu) and UNINETT, Norway's National Research and Education Network (www.uninett.no/en), for providing computing resources.

[†]paul.ehling@bi.no

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

[§]chunyu.yang@bi.no

Abstract

The tax rate on capital gains of equity securities has varied substantially over time and correlates negatively with realized capital gains and collected taxes. Our model shows that investors who anticipate the dynamics of the capital gains tax rate in their bond-equity mix, realize more gains the higher the realized return on equity, the lower the capital gain tax rate, and the higher the capital loss carried over. Simulating the behavior of a calibrated population of investors at an annual frequency, yields a correlation between the model-based aggregate capital gains taxes paid and the data of at least forty-one percent.

Keywords: time-varying capital gains taxation, tax-optimal rebalancing, tax collection from realized capital gains

JEL Classification: G11, H20

1 Introduction

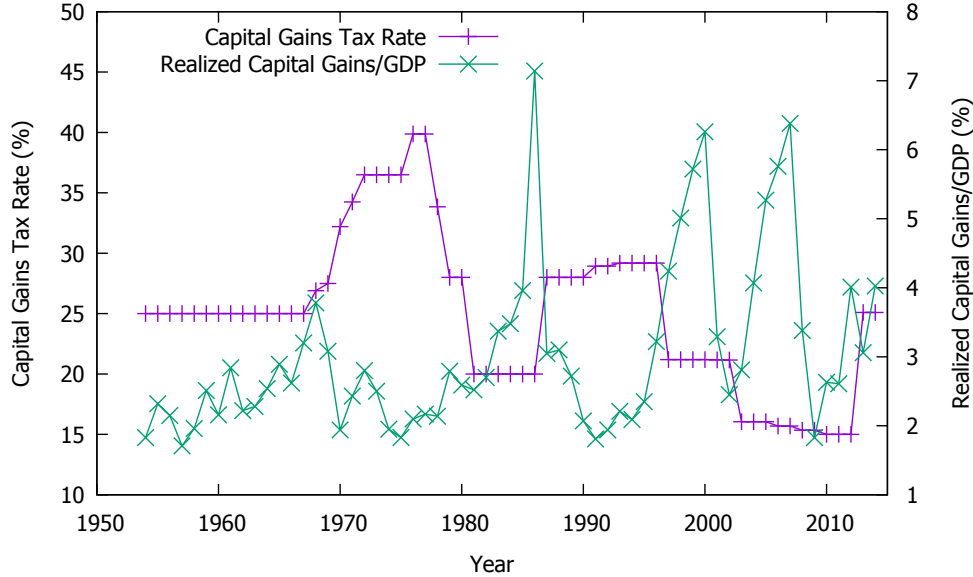
Between 1954 and 2014, the capital gains tax rate in the United States varied between 15 and almost 40 percent.¹ During the same period, realized capital gains of equity securities as a percentage of gross domestic product (GDP) varied between 1.71 and 7.35.² Figure 1 plots the two time-series and indicates that a decrease in the tax rate leads to intensified liquidations of positions with embedded gains, while an increase in the tax rate leads to a decrease in the realization of gains. Specifically, from Figure 1, we see that realized gains spike up during three periods of historically low capital gains tax rates and decline sharply during periods of elevated tax rates such as the 1970s or after 1987 when the maximum tax rate on capital gains increased from 20 to 28 percent. The contemporaneous correlation between the maximum tax rate on capital gains and the amount of realized gains as a percentage of GDP is -50 percent.

How responsive realizations of capital gains and capital gains tax collection are to the level of the capital gains tax rate has been and still is the subject of continuing policy debate. Two questions have been explored in the literature. Do reduced marginal tax rates on realized gains lead to increased revenue? And, is the responsiveness of capital gains realizations to a changing capital gains tax rate transitory or permanent? The literature has tried to answer these questions empirically. While the results in Feldstein, Slemrod, and Yitzhaki (1980) suggest that if marginal tax rates on realized gains were reduced, the revenue collected from realized gains would increase, overall, the empirical literature has documented mixed findings using different data sources. For example, the review articles by Auerbach (1988) and Poterba (2002) point out that much of the debate regarding the responsiveness of capital gains realizations to the level of the capital gains tax rate is fueled by conflicting results obtained with cross-sectional micro data, panel micro data, and aggregate

¹See Sialm (2009) and <https://taxfoundation.org/federal-capital-gains-tax-rates-1988-2013>.

²See <https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/Taxes-Paid-on-Capital-Gains-for>Returns-with-Positive-Net-Capital-Gains.pdf>.

Figure 1: **Time-variation in the Capital Gains Tax Rate and Realized Gains.** The figure plots the U.S. maximum capital gains tax rate and realized capital gains of equity securities as a percentage of gross domestic product (GDP), over the period of 1954 to 2014.



data.³ How to reconcile the findings at the micro level and at the aggregate level remains a challenge. On the issue of transitory vs. permanent effects, studies such as Auten and Clotfelter (1982) suggest that the permanent effects can be smaller than the transitory effects, while Burman and Randolph (1994) do not find statistically significant permanent effects and, thus, question whether reducing capital gains tax rates can permanently raise revenue. More recently, Dowd, McClelland, and Muthitacharoen (2015) separate the data based on the presence of carryover losses to improve on the identification of the tax elasticity. Even then, it appears difficult to settle these issues empirically without a structural model based on theoretical results.

Our paper offers a micro-founded approach that sheds light on both questions. We solve an individual investor's lifetime consumption-portfolio choice problem, where we allow the capital gains tax rate to vary over time.⁴ The investor takes into account the dynamics

³See Bogart and Gentry (1995) for evidence using interstate comparisons.

⁴There exists an extensive portfolio choice literature focusing on various intricacies of the capital gains tax code. See, for example, Dammon, Spatt, and Zhang (2001), Gallmeyer, Kaniel, and Tompaidis (2006), and Ehling, Gallmeyer, Srivastava, Tompaidis, and Yang (2018), among many others. Our paper differs from this literature since we consider a time-varying capital gains tax rate, we provide a simulation of the behavior

of the capital gains tax rate while rebalancing his portfolio. We learn that realized gains covary positively with realized returns on equity, negatively with the capital gains tax rate, and positively with capital losses carried over from previous periods. Further, the investor anticipates declines in the capital gains tax rate when it is relatively high and, hence, is willing to implement conditional equity-to-wealth ratios that are significantly larger than what comparative static cases with a constant tax rate suggest.

To answer the question to what extent our individual-based model can explain the aggregate economic data in Figure 1, we take the optimal trading strategy of individual investors as given, and simulate a population of investors who enter the market over the period of 1954 to 2014 and experience the historically realized returns and capital gains tax rates. Then, we aggregate investors' quantities such as realized gains, carryover losses, and paid taxes into population quantities using either a model-based population wealth distribution or an empirical population wealth distribution. The main result from this calibrated simulation is that the correlation between our model-generated aggregate capital gains taxes scaled by wealth and the annual U.S. capital gains tax collection scaled by GDP is at least forty-one percent.

From our portfolio choice model, we see that the interplay between stock returns, capital gains tax rates, and carryover losses mainly drive the dynamics of aggregate capital gains taxes. We interpret these dynamics through the decisions of individuals whether to hold on to their current level of equity, purchase or sell stocks, and whether their trading leads to realized gains or losses. Holding on to the level of equity from the previous period after a change in the stock price can be optimal when capital gains are taxed, or, more generally, when investors face trading frictions.⁵ When rebalancing is not optimal, we say that the portfolio is in the no-trade region. With capital gains taxation, the no-trade region depends on the tax basis, the carryover loss, the evolution of equity returns and capital gains tax

of cohorts of investors, and we compare our results to data on the tax collection from realized capital gains on equity.

⁵It is well known that with a friction such as capital gains taxes it can be optimal to not rebalance a portfolio after a shock. See Dumas and Luciano (1991) for an example with transaction costs.

rates, and the investor's age.

It is possible that, even without variation in the capital gains tax rate, investors optimally sell part of their equity position with embedded gains. When is that so? It is when the investor's entering equity share is above the no-trade region, where it is optimal to reduce the equity holding to the upper boundary of the no-trade region through selling. If the entering equity share is in the no-trade region, a positive return can move the equity share above the no-trade region in the next period. The larger is the per-period return, the more likely it is that an equity share increases from inside the no-trade region to above it. Alternatively, if the entering equity share is on the upper boundary of the no-trade region any positive return will move it above the no-trade region and selling back to the upper boundary is optimal. Given the statistical properties of annual equity returns, investors are more likely to face an entering equity share that is within or above the no-trade region rather than below the no-trade region, where it is optimal to buy equity.

Why is it optimal to sell equity with embedded gains when the capital gains tax rate declines? Solving the problem of the investors shows that the level of the capital gains tax rate can have a large impact on the no-trade region. For a given level of the entering equity share above the no-trade region, implying embedded gains, the lower is the current or newly applicable capital gains tax rate, the lower is the upper boundary of the no-trade region, and, thus, the larger is the amount of shares sold optimally. In addition, even investors who previously were in the no-trade region might sell equity with embedded gains provided their new no-trade region, implied by the now lower tax rate, shrinks below the entering equity share. Therefore, a lower capital gains tax rate implies intensified selling by all investors who would sell under the previous tax rate, and selling by some investors who would not sell under the previous tax rate. It follows that the aggregate capital gains tax collection can move counter to the tax rate, as can be seen in Figure 1 for the years 1997 and 2003.

The tax collection from realized capital gains on equity, however, does not go hand in hand with variations in the capital gains tax rate. In this regard, we point to the period of

1978 to 1981, where the tax rate declined from almost 40 percent to 20 percent within four years, which is the largest capital gains tax rate cut in history. Over this period, one would expect a surge in tax collection from realized capital gains on equity. Yet, we only see a significant increase in the tax collection in 1986. What breaks the link between changes in the tax rate and changes in tax collection are capital losses. Realized capital losses can be used to offset current capital gains or be carried forward indefinitely to offset capital gains in the future. If the investor is armed with a carried over capital loss that is larger than the current period gain, he uses the carryover loss to avoid paying taxes when rebalancing. In general, the larger is the carryover capital loss the smaller is the no-trade region, implying that, at least, a portion of equity can be sold without incurring capital gains taxes. In the simulation, a significant portion of young investors enter the period of 1978 to 1985 with accumulated carryover losses. This explains why over the period 1978 to 1981 the aggregate capital gains tax collection does not rise at an extraordinary rate. In the simulated data there are other episodes with large carryover losses across investor cohorts, and such episodes can persist for more than a decade. Periods that produce large and persistent carryover losses include the aftermath of the dot-com boom period and the Great Recession of 2008.

Summarizing, with respect to the policy debate, we find that a reduction in the tax rate on realized gains usually does lead to an increased tax revenue. Carryover losses, however, can attenuate or even completely offset any increase in the rate of realization of gains leading to no increase in revenue not just for a few investor cohorts but even at the aggregate level. Therefore, the contemporaneous correlation between the tax rate on capital gains and the amount of realized capital gains is negative but larger than -1. Within our partial equilibrium analysis it is challenging to consider the permanent effect of a current decrease in the capital gains tax rate. However, we conjecture that it should be much smaller than the transitory effect. This is because once the transitory effect, which in essence is the decrease in the upper boundary of the no-trade region, took place, the optimal trading evolves around the new and lower upper boundary of the no-trading region which implies a different level for

the equity-to-wealth ratio but very similar dynamics around it.

Our micro-founded approach explains a significant portion of the time-variation of the aggregate US tax collection data over a long period of time based on a set of parsimonious assumptions on dynamics of the capital gains tax rate, the dynamics of stock market returns, and investors' preferences.⁶ First, our model only considers a static opportunity set with one risky asset representing the aggregate market index. This simplified assumption allows us to tease out the first-order effect of equity returns on the capital gains taxes paid, although it may bias against us finding a significant correlation between the model-generated data and the true data because many investors hold individual stocks rather than the market index.

Second, to solve the portfolio choice problem, we have to make assumptions about the dynamics of the capital gains tax rate. In this regard, we show that the transition probabilities of the tax rate evolution can have large effects on the conditional optimal equity holdings, but only minor effects on the correlation between our model-generated aggregate capital gains taxes and the data. Why? This is because while conditional equity holdings differ, the dynamics of equity holdings do not differ much: When the equity holding is within the no-trade region, a large enough return can move it above the no-trade region next period for very different levels of the upper boundary of the no-trade region. Once the equity holding reaches the upper boundary, any positive return will move the equity holding above the no-trade region next period no matter what the level of the upper boundary of the no-trade region is. A decline in the capital gains tax rate will shift the upper boundary of the no-trade region down no matter what the current level of the upper boundary of the no-trade region is. Hence, our simulation results are robust in that the contemporaneous correlation between the total amount of realized capital gains as a percentage of wealth in the simulation and the total amount of realized capital gains as a percentage of GDP is stable across all considered

⁶In our setting, the portfolio choice problem has three endogenous state variables: the stock holding, the weighted average purchase price, and the carryover loss. We also have to track one exogenous state variable, namely the capital gains tax rate, plus time. As a run-time benchmark based on our computing resources, our portfolio choice problem with time-varying capital gains tax rate takes approximately 20 hours to solve using 100 CPUs (2.66GHz) running in parallel. Adding one more state variable would increase the computational cost by at least a factor of 10.

transition regimes.

Third, due to innate heterogeneity, investors may have many different trading motives other than the return- and tax-induced trading motive in our model. All trading motives can contribute to the taxes paid on capital gains in the data. Overall, assuming a homogeneous set of investors, who have only a tax-induced trading motive, biases against finding a significant correlation between the model-generated data and the true data. More specifically, our investors are heterogeneous only in age or, equivalently, heterogeneous in when they enter the market. Further, one example of age heterogeneity that matters for taxes paid on capital gains that we do not model is the exact timing of the tax rate changes. On occasions when the capital gains tax rate changes, such as 1997 and 2003, the actual effective date of the new tax rate can be significantly different across different investors due to complicated phase-in rules that differentiates investors based on their holding period, which depends on age, and the realization date.⁷ Modeling such phase-in rules goes beyond the goals of this paper. Again, we expect that ignoring this specific intricacy of the tax code biases against finding a correlation between the model generated data and the true data.

Although we assume that the distribution of stock returns before-tax is static, time-varying capital gains tax rates can be thought of as inducing after-tax, time-varying and state-dependent, expected returns. The seminal contribution in this literature is Merton (1971), in which the optimal consumption and allocation strategies are available in closed form for special cases. More recently, Campbell and Viceira (1999), Campbell and Viceira (2001), Campbell, Cocco, Gomes, Maenhout, and Viceira (2001), Brandt, Goyal, Santa-Clara, and Stroud (2005), Brandt and Santa-Clara (2006), and Garlappi and Skoulakis (2010), among others, use numerical methods to approximate the solution to specific optimal consumption and asset allocation decisions with time-varying returns. Another related study

⁷The details on the phase-in for the 1997 tax reform — called the Taxpayer Relief Act — can be found here: <https://www.scribd.com/document/24597976/PL-105-34-Taxpayer-Relief-Act-of-1997>. The details on the phase-in for the 2003 tax reform — called the Jobs and Growth Tax Relief Reconciliation Act — can be found here: <https://www.scribd.com/document/24597313/PL-108-27-Jobs-and-Growth-Tax-Relief-Reconciliation-Act-of-2003>.

is Gârleanu and Pedersen (2013), in which a dynamic portfolio policy is derived when trading is costly and time-varying returns are predictable. Our paper differs from this strand of the literature in that the tax-induced trading motive allows us to connect the portfolio choice theory with aggregate data on realizations of capital gains and the resulting tax collection.

2 Investor’s Consumption-Portfolio Choice Problem

In this section, we briefly describe the investor’s consumption-portfolio choice problem with capital gains taxation, following closely the literature.⁸

Consumption and trading take place at discrete times. All investors have access to a riskless money market, which pays a continuously compounded time-varying interest rate, and a risky stock, which pays out a time-varying dividend. An investor’s wealth comprises of holdings in the stock and in the money market account. At each date, the investor rebalances his portfolio, consumes, and pays taxes on interest and dividend income and, if he sells equity with embedded gains, on realized capital gains.

We assume that dividend and interest income are taxed at constant rates, while the capital gains tax rate varies over time. Capital gains are determined by the difference between the price equity is sold for and its absolute tax basis, that is, the weighted-average purchase price of a stock. The relative tax basis or basis-to-price ratio, that is, the ratio of the absolute tax basis divided by the current level of the stock price, measures the tax friction. A stock with an embedded capital gain has a basis-to-price ratio that is smaller than one, while a stock with an embedded capital loss has a basis-to-price ratio larger than one. When an investor with an embedded capital gain reduces his position, he realizes at least part of the embedded capital gain and pays capital gains taxes. When an investor holds a position with an embedded capital loss, he liquidates the position and realizes the entire capital loss even if he cannot use it immediately. After liquidation, the basis-to-price ratio resets to one because any new purchase of the stock is at the current market price and the

⁸The technical description of the consumption-portfolio choice problem is in Appendix A.

unused loss can be carried forward in the form of a carryover loss.

The investor derives utility from real consumption each period and from terminal wealth at the time of his death. The investor's utility over consumption and bequest is given by a function with constant relative risk aversion (CRRA) adjusted for his time discount rate. When the investor dies, all assets are liquidated and distributed to his heirs in annual payments over a period of years, where the payments are held constant in real terms.

Given an initial wealth, the investor's objective is to maximize his expected lifetime utility from real consumption and terminal wealth at the time of death by choosing a budget-feasible trading strategy. Using the principle of dynamic programming, we describe and solve the investor's lifetime consumption and portfolio choice problem in a recursive form as the solution to a Bellman equation.

3 Parametrization

Investor: Starting from age 20 the investor rebalances at an annual frequency. Each period the investor faces a conditional death probability of $1 - e^{-\lambda_t}$ or survival probability of $e^{-\lambda_t}$. The single-period hazard rate $\lambda_t > 0$, $t = 0, \dots, T - 1$, is calibrated to the 1990 U.S. Life Table of the National Center for Health Statistics, where at T , $\lambda_T = \infty$. We assume that the investor dies with certainty at age 100 years; i.e., $T = 80$. His time discount rate is $\beta = 0.96$, and his relative risk aversion coefficient is $\gamma = 5$. The investor implements his bequest motive by providing a constant annual payment in real terms for $\eta = 20$ years to his heirs using his liquidated wealth at death.

Moments: We employ the joint empirical distribution of the S&P 500 Index, the nominal interest rate, the inflation rate, and the dividend yield.⁹ Specifically, the empirical distribution of the S&P 500 Index captures negative skewness in stock market returns, which implies a higher probability of large negative returns and lower probability of large posi-

⁹The joint moments are summarized in Table 1 of the Internet Appendix.

tive returns.¹⁰ Since stock prices, dividends, and interest rates are all in nominal terms the investor's objective is adjusted for inflation.

Tax Code: The tax rates used are set to approximately match those faced by a wealthy investor under the U.S. tax code. Specifically, interest is taxed at the investor's marginal income tax rate $\tau_I = 35\%$;¹¹ dividends are taxed at $\tau_D = 15\%$.

We follow the US tax code in that, at the investor's death, interest and dividend taxes are paid but capital gain taxes are forgiven.

Evolution of the Capital Gains Tax Rate: To gauge the impact of a transition regime for the capital gains tax rate on the optimal portfolio holding, we consider three regimes. These three transition regimes capture a wide spectrum of possible dynamics of the capital gains tax rate. Specifically, we model the evolution of the capital gains tax rate, $\tau_C(t)$, as a Markov chain (i) calibrated on historical transitions (base case), (ii) a no-transition regime, and (iii) a uniform transition regime.

In the base case, a historical transition regime, the levels and the transition probability matrix are matched to the historical tax rates over the period of 1954 to 2014. In particular, we assume that the capital gains tax rate can take the following six values: 15%, 20%, 25%, 28%, 35%, and 40%. The associated transition probability matrix is given by

	15%	20%	25%	28%	35%	40%	
15%	9/10		1/10				
20%	1/12	5/6		1/12			
25%			1/16	15/16			,
28%		1/7		11/14	1/14		
35%				1/7	5/7	1/7	
40%					1/2	1/2	

(1)

¹⁰Given the moments of stock returns, neither a no-short-sale constraint nor a margin constraint with a 50 percent margin requirement bind.

¹¹According to the 2017 tax brackets (for taxes due on April 17, 2018), the highest marginal income rate is 39.6 percent. The 35 percent tax rate that we use applies, for example, to an income of up to 470,700 for married taxpayers filing jointly or a qualifying widow.

where entries with probability zero are not shown.

For the two other cases, we consider 31 possible capital gains tax rates ranging from 10% up to 40% with a step size of 1%. Under the no-transition regime, an investor thinks that the capital gains tax rate will never change again despite the fact that it did change in the past. Thus, the transition probability matrix has ones on the diagonal and zeros elsewhere. Under the uniform transition regime, an investor assumes that all capital gains tax rates are equally likely in the next period. Thus, the transition probability matrix has a constant probability of $1/31$ everywhere. The historical transition regime is between the no-transition and the uniform transition regimes and is more similar to the no-transition regime as the historical transition probability matrix is closer to the identity matrix.

4 Conditional Strategies and No-trade Regions

In our model, an investor's optimal equity holding, expressed as exiting equity-to-wealth ratio, is a function of age, entering equity-to-wealth ratio, basis-to-price ratio, carryover loss-to-wealth ratio, current capital gains tax rate, and the transition regime of the capital gains tax rate.

4.1 Conditional Exiting Equity-to-Wealth Ratios

Figure 2 presents optimal exiting equity-to-wealth ratios as a function of the entering equity-to-wealth ratio and basis-to-price ratio for a 20 year old investor, while fixing the other state variables. The figure shows the exiting equity-to-wealth ratios for the three transition regimes introduced in Section 3: uniform, historical, and no-transition when the current capital gains tax rate is at 20 and 40 percent, respectively, and the carryover loss-to-wealth ratio is zero. We see that the uniform transition regime leads to the highest equity-to-wealth ratios while the no-transition regime produces the lowest equity-to-wealth ratios. The monotonicity in the ordering of the equity-to-wealth ratios across the regimes is consistent

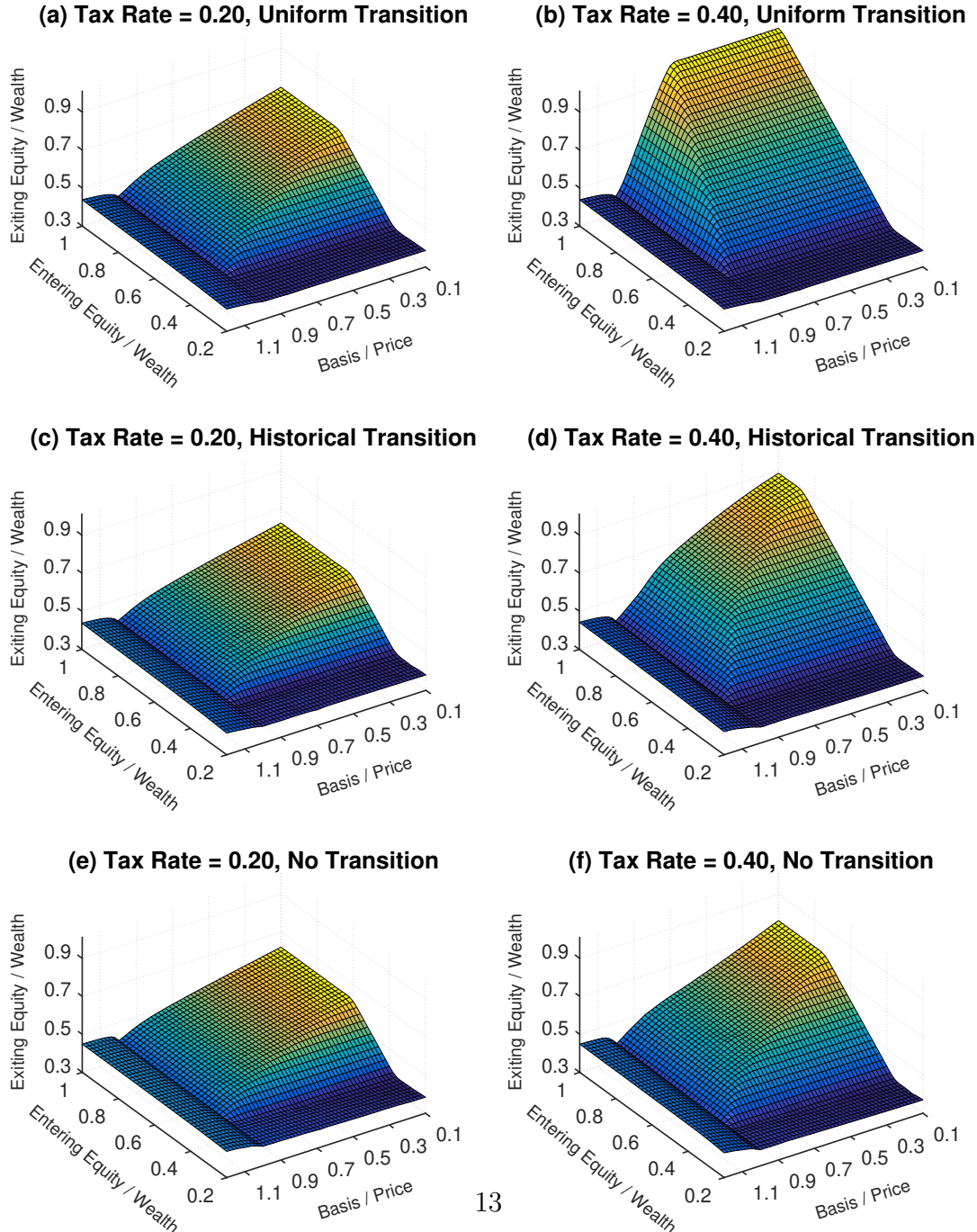
with the interpretation that the historical transition regime is in-between of the no-transition and the uniform transition regimes.

Panels (e) and (f) of Figure 2 correspond to the no-transition regime, which is equivalent to a constant tax rate regime. These figures confirm the standard tradeoff between the marginal tax costs of trading and deviating from the optimal unconstrained equity-to-wealth ratio with a current basis-to-price ratio of one: the lower the basis-to-price ratio, the higher the marginal tax costs of selling.¹² Thus, more equity is held optimally, except for very low entering equity-to-wealth ratios, where investors have an incentive to buy rather than sell. The influence of the basis-to-price ratio on the optimal exiting equity-to-wealth ratios is the stronger the larger is the entering equity-to-wealth ratio, except for very high entering equity-to-wealth ratios. This is because the capital gains tax paid on realized gains increases both with a larger equity position and with a lower basis. Further, the higher the constant tax rate, 40 percent in panel (f) versus 20 percent in panel (e), the higher the tax friction and the more equity is held optimally.

Under the historical transition regime and the uniform transition regime, panels (a) to (d), investors perceive time-varying capital gains tax rates. Given the current capital gains tax rate, investors face a similar tradeoff between the marginal tax costs of trading and deviating from the optimal unconstrained equity-to-wealth ratio as in the no-transition regime. Perceiving the possibility of future tax rate changes induces a timing behavior, where the investor is willing to bear with a temporarily high equity-to-wealth ratio when the current capital gains tax rate is high and is expected to fall in the future. In particular, when an agent takes the time variation in the capital gains tax rate into account, he knows that 40 percent is a high tax rate from a historical perspective, and he also knows that the optimal holdings at a lower tax rate are lower. It can then be optimal to postpone a reduction in the elevated holdings until the tax rate is indeed lower to reduce capital gain taxes paid. Since, in the case of uniform transition, the probability for a lower tax rate next period is higher,

¹²The optimal unconstrained equity-to-wealth ratio in our model differs from the optimal equity-to-wealth ratio in a frictionless world as it takes into account the tax friction in the future.

Figure 2: **Conditional Exiting Equity-to-Wealth Ratios.** The figure plots optimal exiting equity-to-wealth ratios as a function of the entering equity-to-wealth ratio and the basis-to-price ratio for an investor with age 20 assuming a carryover loss-to-wealth ratio of zero. The top plots show optimal exiting equity-to-wealth ratios for an investor with no information about the evolution of tax rates, the middle plots show optimal exiting equity-to-wealth ratios for an investor with perfect information about the transition probabilities of tax rates, and the bottom plots show optimal exiting equity-to-wealth ratios for an investor who is ignorant about the evolution of tax rates. In the plots on the left, the current capital gains tax rate is 20 percent and on the right, the current capital gains tax rate is 40 percent. Additional figures with optimal exiting equity-to-wealth ratios, including the corresponding figures for an investor with age 80, are in the Internet Appendix.

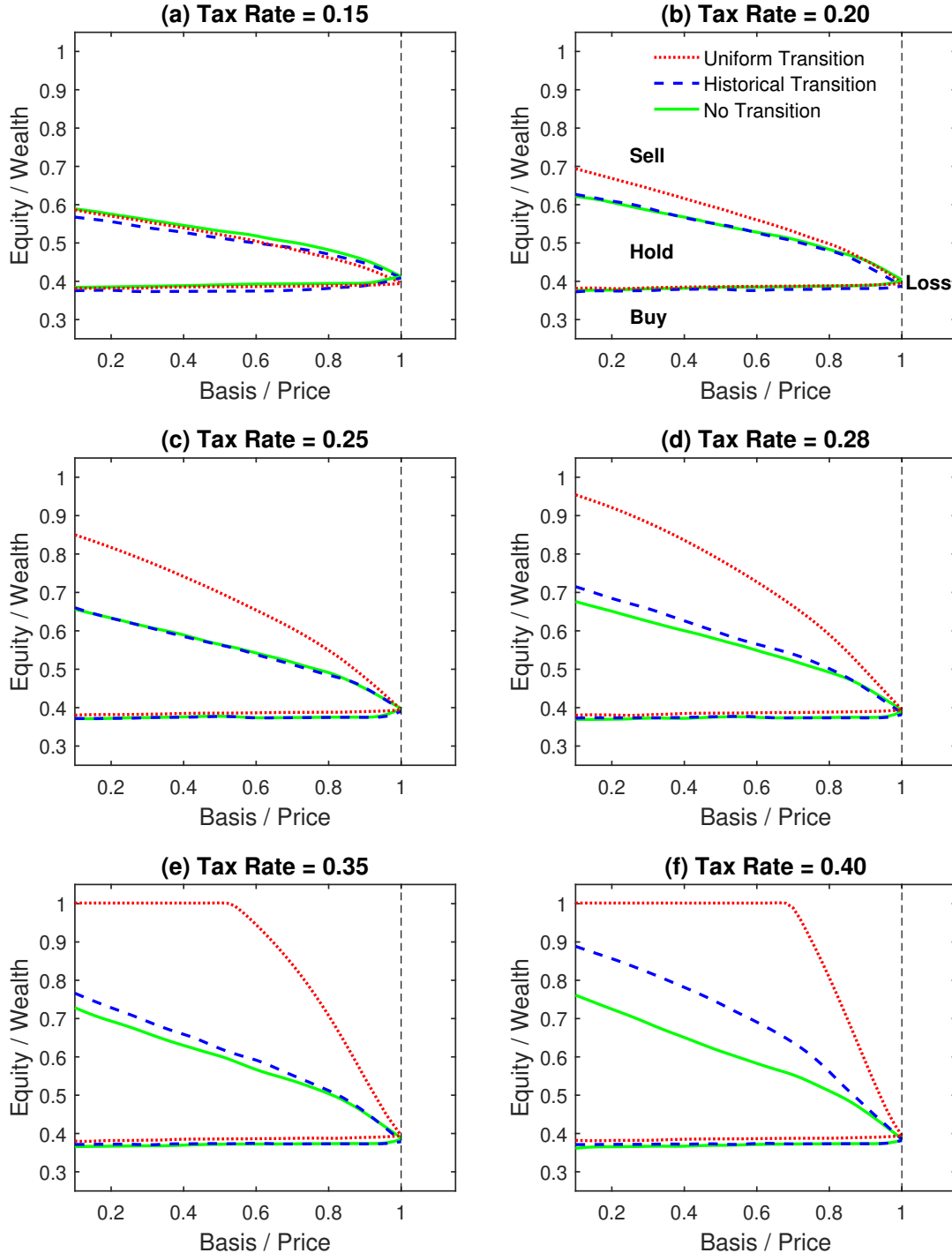


relative to the case of historical transition, the incentive to not reduce an elevated equity holding this period is higher, as can be seen from comparing panels (b) and (d).

4.2 No-trade Regions

An intuitive way to present optimal portfolios with a tax friction is to inspect the no-trade regions. What differentiates our portfolio choice model from a frictionless model is that, due to the tax friction, it can be optimal to not trade, while in the frictionless model after each price change an agent immediately trades back to the optimal equity-to-wealth ratio. Figure 3 plots sell, hold (no-trade), and buy regions for equity-to-wealth ratios as a function of the basis-to-price ratio for an investor with age 20 and a basis-to-price ratio of one. In general, entering a period above the no-trade region implies selling all the way down to the upper boundary of the no-trade region and paying capital gains taxes. Entering a period in the no-trade region implies inaction, and entering below the no-trade region implies buying all the way up to the lower boundary of the no-trade region. Figure 3 shows these regions for the cases when the current capital gains tax rate is 15 (a), 20 (b), 25 (c), 28 (d), 35 (e), and 40 (f) percent. We notice large differences between the no-trade regions, driven by both the current level of the capital gains tax rate and the transition regime. For each of the three transition regimes, increasing the current level of capital gains tax rate leads to larger no-trade regions. While fixing the current level of capital gains tax rate, differences among different transition regimes vary. Specifically, when the current capital gains tax rate is 15 percent there is almost no difference between the no-trade regions of the three transition regimes. Starting from a 20 percent capital gains tax rate, we see that the upper boundary of the no-trade region with uniform transition increases significantly, relative to the two other cases. At a 40 percent capital gains tax rate, all three transition regimes show significantly different upper boundaries of their no-trade regions. The lower boundaries of the no-trade regions are neither very sensitive to the current capital gains tax rate nor to the transition regime.

Figure 3: **No-trade Regions.** The figure plots optimal sell, hold (no-trade), and buy regions for equity-to-wealth ratios as a function of the basis-to-price ratio for an age-20 investor facing a current capital gains tax rate of 15 (a), 20 (b), 25 (c), 28 (d), 35 (e), and 40 (f) percent. Each plot shows optimal sell, hold, and buy regions for an investor with no information about the evolution of tax rates (uniform transition, dotted line), an investor with perfect information about the transition probabilities of tax rates (historical transition, dashed line), and an investor who is ignorant about the evolution of tax rates (no-transition, solid line). Corresponding figures for an investor with age 80, are in the Internet Appendix.



Our focus is on selling of equity with embedded gains; i.e., on realizing capital gains. From Figure 3 we identify two cases in which it is optimal to sell equity with embedded gains: when the investor’s portfolio is close to the upper boundary of the no-trade region and a positive return moves it to the sell region; and when the capital gains tax rate drops. A large enough return can push a portfolio from the no-trade region into the sell region. This leads to optimal selling of equity with embedded gains back to the upper boundary of the no-trade region and, hence, to capital gains taxes. Once the portfolio is on the upper boundary of the no-trade region, every positive return will push the portfolio into the sell region. Given that annual returns are more likely to be positive than negative, selling is more likely than inaction and buying. What is striking about Figure 3 is that independently of the transition regime, the lower the capital gains tax rate the smaller the no-trade region. Therefore, conditional on the basis-to-price ratio, for most investors a declining capital gains tax rate is a major opportunity to optimally reduce their equity-to-wealth ratio closer to the unconstrained optimum even if it involves selling equity with embedded gains and paying a potentially large amount of capital gain taxes. This is the case even without any liquidity needs, which are outside of our portfolio choice model.¹³

5 Simulating a Population of Investors

Given the solutions to the individual investor’s portfolio choice model, we simulate a population of investors from 1954 to 2014, where each investor represents an age group in a calendar year, and compare the aggregate population quantities generated by our simulation, such as the aggregate realized capital gains and the aggregate capital gains taxes paid, versus the US tax collection data. We obtained the, publicly available, US tax collection data from the Statistics of Income site at the Internal Revenue Service.

¹³It is often argued that the rational reason for the sale of equity with embedded capital gains must be liquidation shocks such as college tuition for offspring, dissaving over retirement, divorce, long-term unemployment, purchase of real estate, or severe illness. While liquidation shocks certainly play a role for the sales of equity with embedded capital gains, in this paper we instead ask when it is optimal to rebalance an elevated equity position with embedded capital gains without a liquidation shock.

We split the simulation in three steps. In Step 1, we construct an initial cross-section of investors at the end of 1953. In Step 2, we take the initial cross-section as given and simulate forward a panel of individual investors from 1954 to 2014, which corresponds to the period with available US tax collection data on realized gains and paid capital gains taxes. In Step 3, we aggregate the quantities obtained from simulating the behavior of individual investors at the population level.

5.1 Constructing the Initial Cross-Section

Due to the lack of data on the cross-section of US investors' stock holdings and basis-to-price ratios, we construct the initial cross-section of US investors at the end of 1953 by considering a group of individual investors who enter the market at different times. As shown in Figure 4, we let the first investor enter the market in 1875 at age 20 – in 1953 this specific investor is 98 years old. To construct the initial cross-section, the last investor enters the market in 1953, just before our sample period starts, at age 20. All investors enter the market with 100 dollars of wealth, zero carryover loss, and zero embedded gains or losses implying a basis-to-price ratio of one. Investors hold their portfolios to the end of next year subject to the same market conditions, namely, historical S&P 500 returns, risk-free rates, and inflation rates. They then rebalance their portfolios independently at the end of each year according to a rebalancing strategy described below.

Before 1913, no capital gains taxes were collected in the US. Hence, in the simulation all investors rebalance their portfolios according to an optimal strategy without capital gains taxes before 1913. The collection of capital gains taxes starts in 1913 at a tax rate of 7 percent from 1913 to 1921 and 12 percent from 1922 to 1953. Over this period, we assume that all investors follow the no-transition strategy by assuming that the capital gains tax rate will stay at 7 percent or 12 percent indefinitely in the future. At the end of 1953, which is when the tax collection data start, the population of investors is made up of 79 cohorts with different age, wealth, basis-to-price ratio, equity-to-wealth ratio, and carryover

loss-to-wealth ratio. This cross-section serves as the initial cross-section for our simulation from 1954 to 2014.

5.2 Simulating Individual Investors

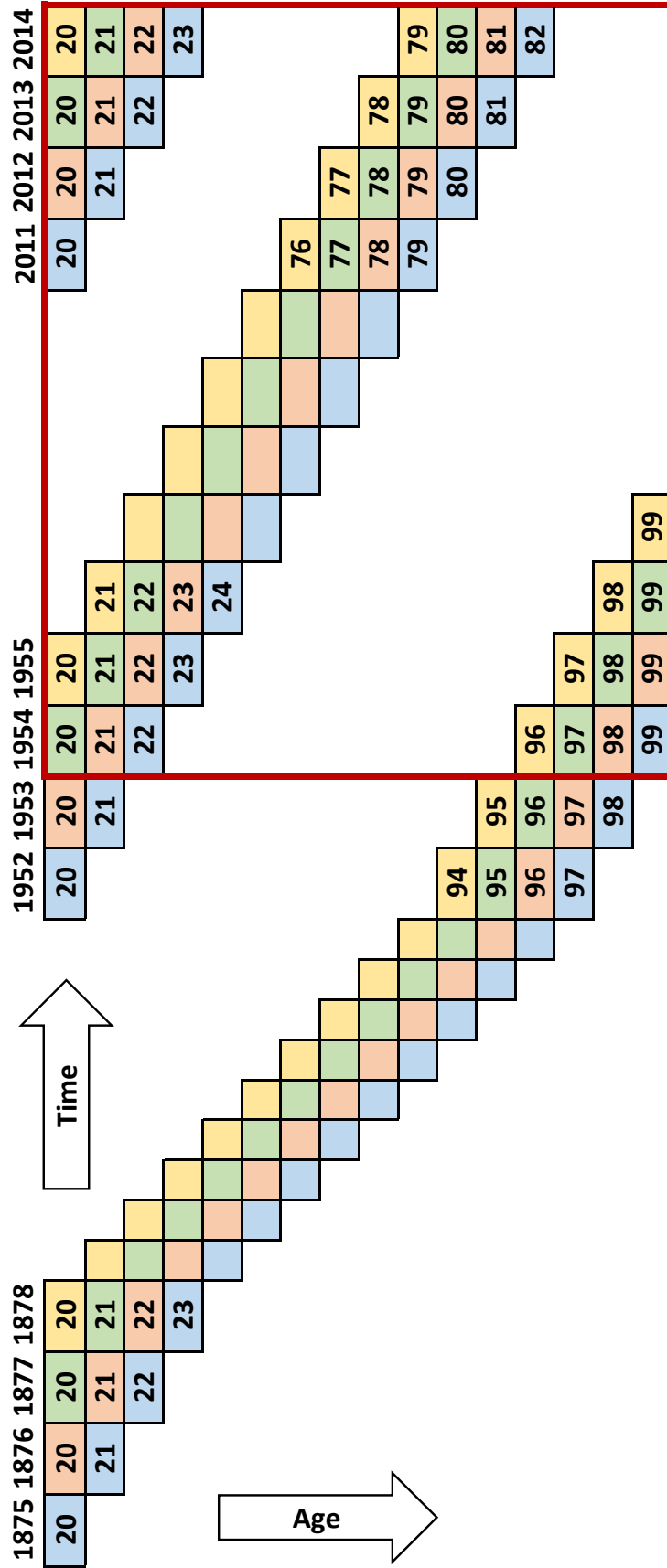
Given the initial cross-section in 1953, we simulate forward the cross-section of investors at different ages, year by year, from 1954 to 2014. As shown in Figure 4, every year there is a new investor, representing a cohort, entering the market at age 20 with 100 dollars in wealth, zero carryover loss, and basis-to-price ratio of one. Every year, the existing investors age by one year and an investor with age 100 exits the market. Starting from 1954, we have a complete cross-section of investors at all ages. The cross-sectional variation within each year is driven by the variation in the initial cross-section for investors who were born before 1954 and the variation in the entering time for investors who were born after 1954.

Within the same year, all investors experience the same S&P 500 return and dividend, risk-free rate, inflation rate, and capital gains tax rate. However, they rebalance their portfolios differently given that they have different entering equity-to-wealth ratios, different basis-to-price ratios, different carryover loss-to-wealth ratios, and different ages. All investors rebalance their portfolios following their individually optimal strategy under one of the following transition regimes: uniform transition, historical transition, or no-transition. For each transition regime, we keep track of the following quantities for each investor: age, wealth, basis-to-price ratio, equity-to-wealth ratio, carryover loss-to-wealth ratio, realized capital gains-to-wealth ratio, and paid capital gains taxes-to-wealth ratio.

5.3 Aggregating Individual Age Cohorts to the Population

So far, our simulations follow each individual investor, corresponding to an age cohort, in the population and the simulation quantities are all scaled by each investor's wealth. To aggregate individual investors to the population level, we need a population wealth distribution across age. We use three different ways to construct the population wealth distribution: the

Figure 4: **Simulating Individual Investors.** The figure shows how we simulate individual investors with the goal to have an entire distribution of the population, that is 80 investor types, by 1954. In the simulation, the first investor enters the market in 1875 at age 20 and reaches 1954 at age 99. The last investor enters the market in 2014 at age 20. This allows producing a complete cross-section of investors with differing age, wealth, basis-to-price ratio, equity-to-wealth ratio, and carryover-loss-to-wealth ratio at the end of 1953.



model-driven wealth distribution; an empirical wealth distribution; and a uniform wealth distribution, where both the empirical wealth distribution and the uniform wealth distribution are independent of our model.

The model-driven wealth distribution is obtained by multiplying two terms at each age, namely the individual investor’s wealth at each age and the population density at each age. We obtain individual investor’s wealth at each age from our simulation of individual investors in Step 1 and Step 2; thus, this is a model-driven quantity. We obtain the population density at each age using the hazard rate in 1990 US life table with a constant 1% population growth rate. The population density is time-independent while the cross-section of investors’ wealth is time-dependent. Hence, the model-driven wealth distribution is time-dependent. Further details of the model-driven wealth distribution can be found in Appendix B.

For the empirical wealth distribution, we employ the data from the Survey of Income and Program Participation (SIPP) in 2013.¹⁴ SIPP reports both the mean household’s financial wealth for different age groups and the total number of households per age group, which allows us to compute the total wealth per age group and the wealth distribution across age groups. To split the weights on each age group into weights on each individual age, we use linear interpolation and extrapolation.

Finally, we use an uninformative wealth distribution, the uniform distribution, to aggregate the cross-section of investors at different ages to the population level and report the population realized capital gains scaled by the population wealth and the population paid capital gains taxes scaled by the population wealth.

6 Inspecting the Cross-Section of Investors

To help understand the main simulation results, we inspect the cross-section of realized capital gains scaled by wealth, paid capital gains taxes scaled by wealth, carryover losses scaled

¹⁴SIPP discloses data every two years over the period of 1993 to 2013. For simplicity, we use the 2013 distribution. Applying a time-dependent empirical distribution from 1993 to 2013 leads to only minor differences in our simulation results.

by wealth, and basis-to-price ratios. Figure 5 summarizes these cross-sectional quantities in four heat maps for the base case with no-transition for the initial cross-section and historical transition regime thereafter.

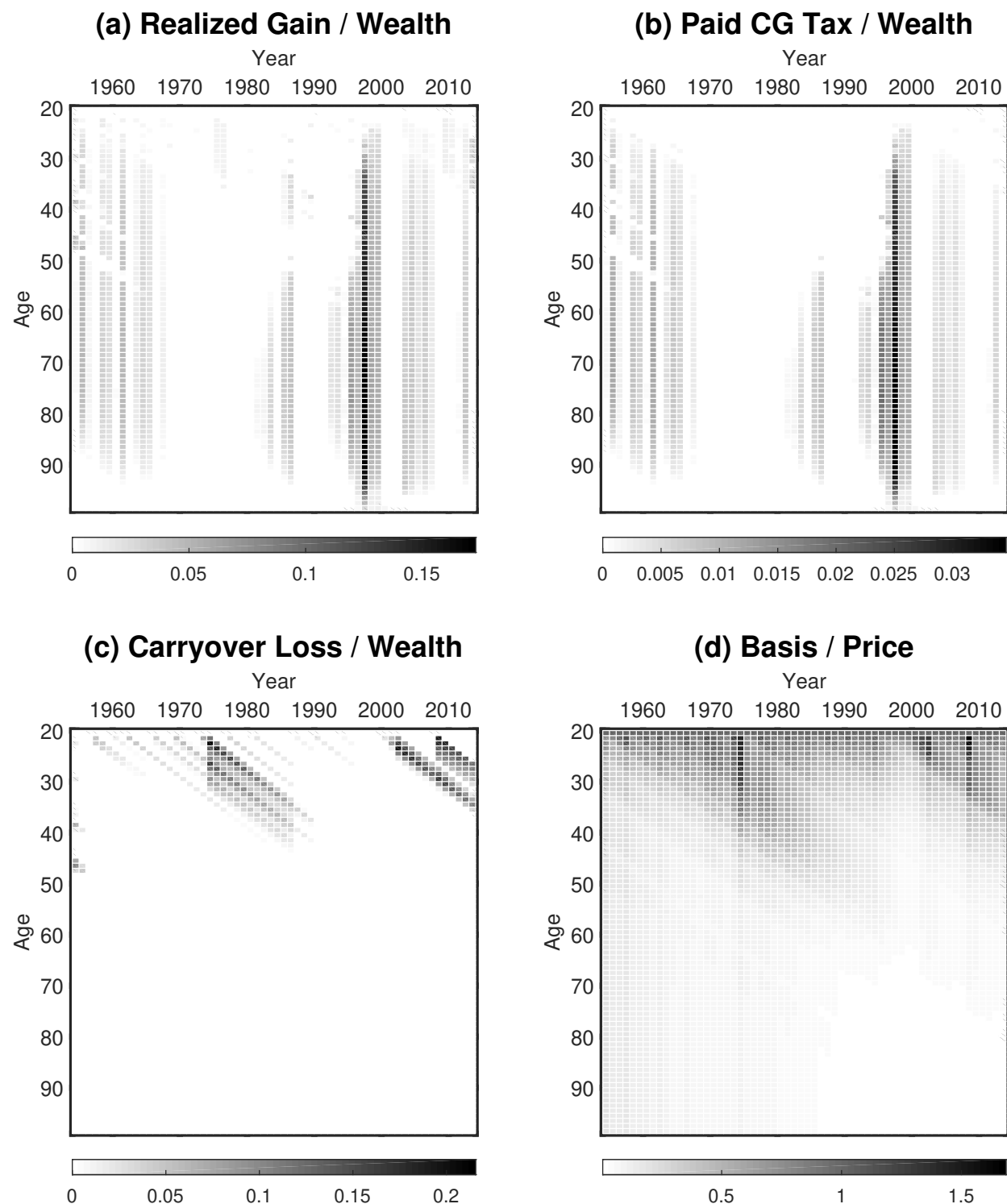
Panel (a) in Figure 5 shows the cross-section of the realized capital gains scaled by wealth for each age group from 1954 to 2014. The scale of the heat map ranges from white, implying no realization of gains, to black, indicating the largest individual realized capital gains scaled by wealth (17.3%) that occurs in the simulation. We see large variations of realized capital gains scaled by wealth across the age groups. Young investors experience little or no realized gains, simply because they just entered the stock market, whereas old investors postpone more and more the realization of gains since capital gains taxes are forgiven at death. A main driver of within age variation of the realized capital gains scaled by wealth is realized returns after an investor turns 20.¹⁵ These experienced returns also drive the variation of the entire age cross-section over time, which is larger than the variation across age. For example, we see that at the end of the dot-com boom period almost all age groups realize large amounts of taxable gains. Further, we see that in the second half of the 70s and first half of the 80s many age groups realize very small amounts of capital gains or even no gains.

The variation of the cross-section of carryover losses over time is shown in panel (c) of Figure 5. Here, the largest carryover loss scaled by wealth that occurs in the simulation is 21.6%. Three episodes produce large carryover losses that are optimally carried over for more than a decade. These occurred in the early 70s, the aftermath of the dot-com boom period, and the Great Recession. During these periods young investors experience negative returns that can overturn any accumulated positive returns to produce substantial carryover losses.¹⁶ According to the simulation, investors that are older than 50 years do not have significant carryover losses.

¹⁵The largest aggregate realized gain occurs in 1997 at 12.7% of wealth. This is due to both a large positive stock market return of 29% and a tax rate cut from 29% to 21%. In 1995 with an even larger stock market return of 35% the aggregate realized gain is only 3% of wealth.

¹⁶We note that some middle aged investors start with carryover losses in 1954, which are driven by returns before the tax collection data starts.

Figure 5: **Heat Map of Population Quantities.** The figure shows heat maps of the cross-section of realized capital gains scaled by wealth, paid capital gains taxes scaled by wealth, carryover losses scaled by wealth, and basis-to-price ratios for each age group from 1954 to 2014. The scale of the heat map ranges from white, corresponding to zero, to black, corresponding to the largest observation.

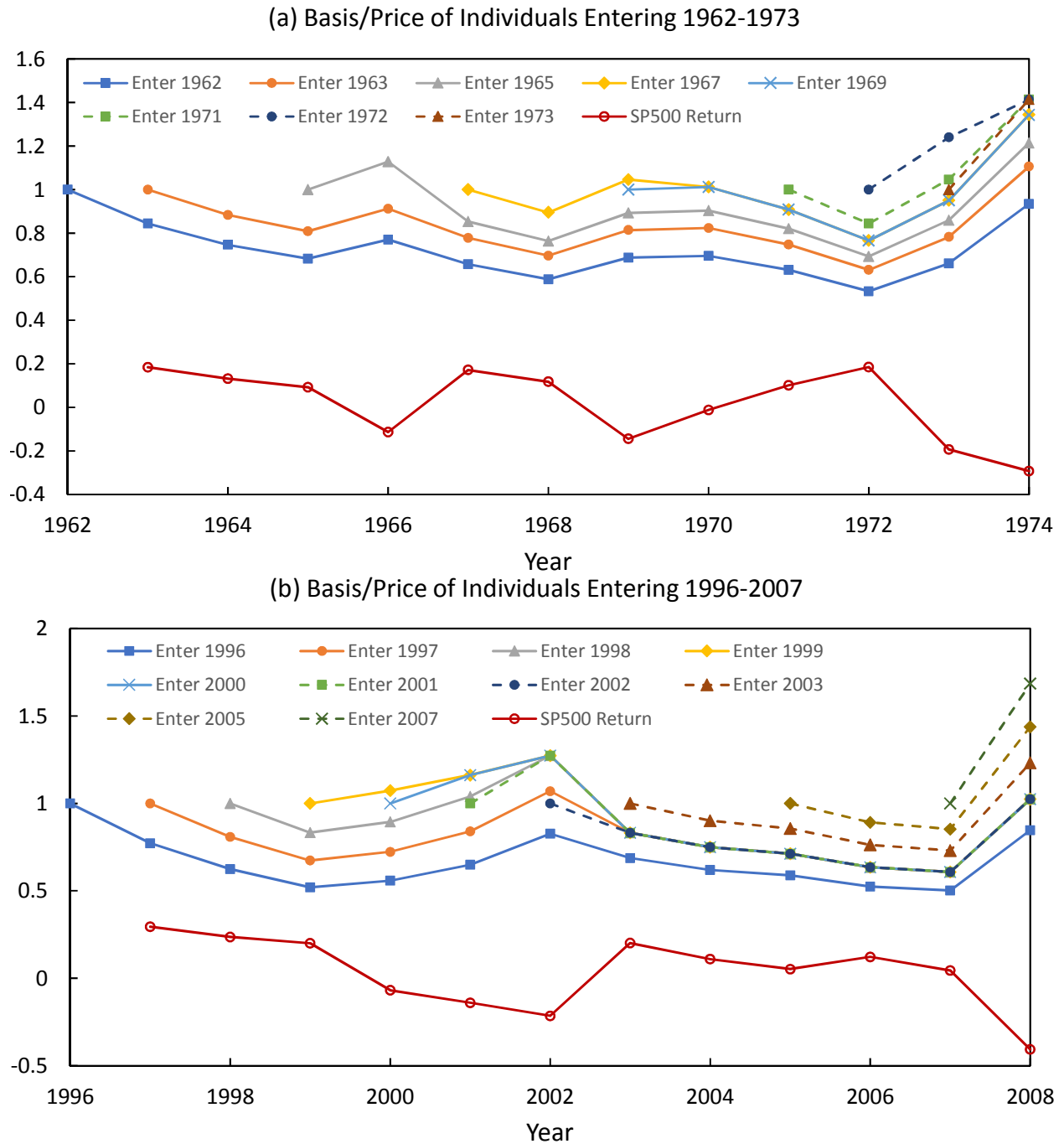


Combining the two left plots containing the realized gains and the available carryover losses, we obtain the cross-section of paid capital gains taxes scaled by wealth, shown in panel (b) of Figure 5. A carryover loss reduces the tax friction by shrinking the no-trade region and it, therefore, facilitates trading towards the unconstrained optimal equity-to-wealth ratio. We see that it can, in some cases, completely offset all realized gains in a period.

Panel (d) shows the cross-section of the basis-to-price ratios from 1954 to 2014. The basis-to-price ratio summarizes the impact of the tax friction, where a basis-to-price ratio of one implies that currently the unconstrained equity-to-wealth ratio is attainable; a basis-to-price ratio above one implies that a loss can be realized and carried forward, implying that the basis-to-price ratio resets to one; and a basis-to-price ratio below one implies that the investor faces embedded gains and a no-trade region. In the data, positive returns are realized more often than negative returns. Hence, over time, as investors' age the basis-to-price ratio declines substantially below one.

Two episodes with accumulation of large carryover losses stand out in the heat map in panel (c) of Figure 5. One in the early 1970s and the other one in the early 2000s. In Figure 6, we plot selected individual cohorts' basis-to-price ratio paths over these two episodes. Between 1962 and 1972, there is a series of moderate positive and negative returns which keep the basis-to-price ratios of the young cohorts around one. This sequence is then followed by two successive years with large losses, namely in 1973 and 1974. As can be seen from panel (a) in Figure 6, this leads to significant capital losses in the cross-section. Specifically, in 1974 only the cohort that entered in 1962 does not have a carryover loss. In the early 2000s, the losses in 2000, 2001, and 2002 fuel in a similar fashion the second episode of accumulation of carryover losses across many cohorts. The evolution of carryover losses across the young cohorts suggests that the path dependence of losses can be crucial for understanding data on capital gains tax collection. Consequently, from Figure 6 one can conjecture that to match the data over the period of 1954-2014, our model must be able to

Figure 6: **Evolution of Basis-to-Price Ratios.** The figure shows the evolution of basis-to-price ratios across cohorts accompanied by the return on the S&P 500 Index between 1962 and 1974 (panel (a)) and 1996 and 2008 (panel (b)) for a select group of cohorts.



capture this long-horizon path-dependency across cohorts.

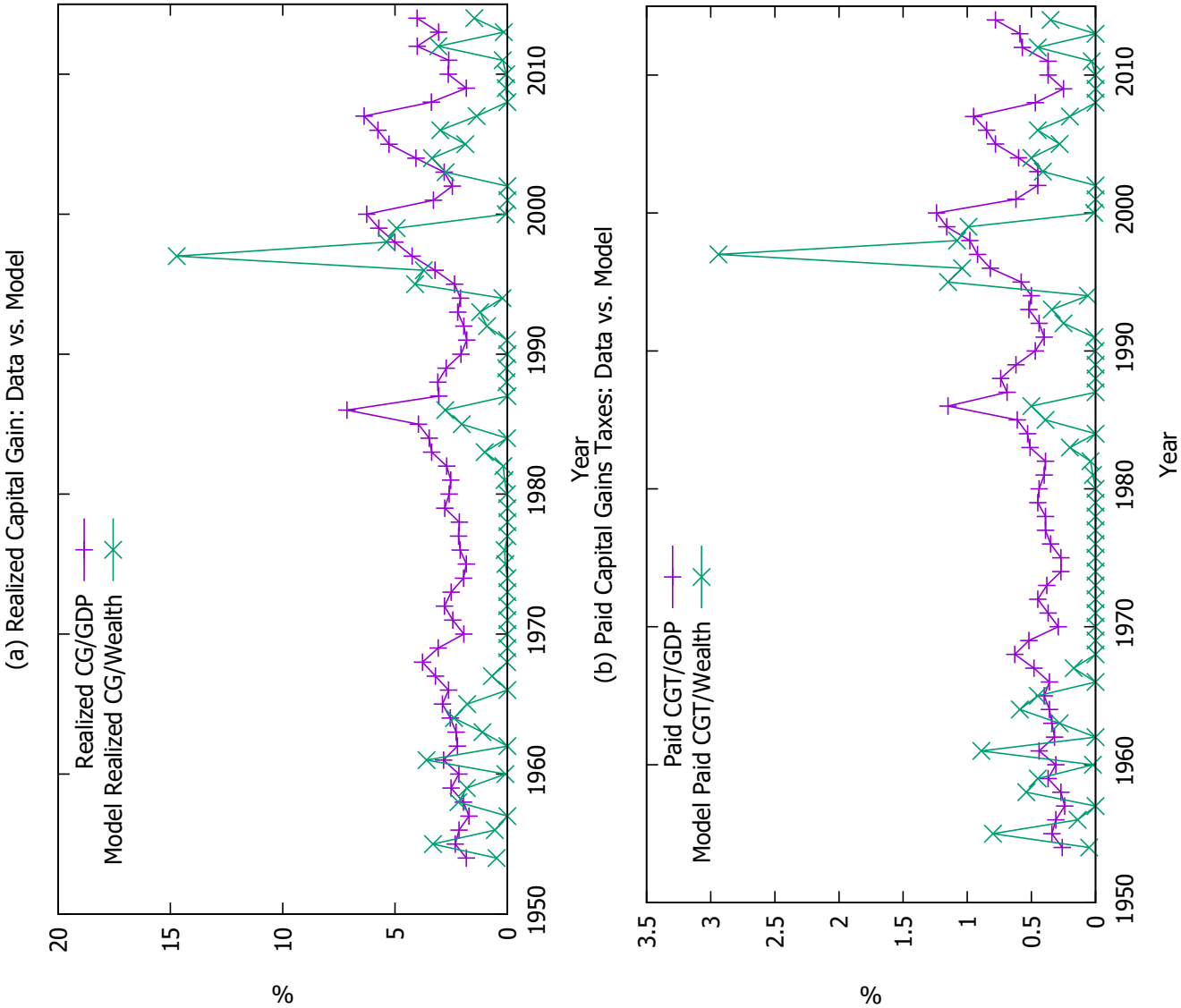
7 Aggregate Dynamic Tax Trading Strategies

We now compare the evolution of the aggregate realization of capital gains and total taxes collected from the realization of capital gains on equity securities between our simulation and the tax data. Since the data are scaled by GDP, the aggregate realized capital gains from the model are scaled by wealth to facilitate comparison.

Panel (a) in Figure 7 shows the evolution of realized capital gains in the data versus the simulated realized capital gains using the model-driven wealth distribution. We see that the model reproduces the data quite well. The contemporaneous correlation is 38.1% percent. Specifically, the model replicates smaller ups and downs in the actual capital gains, for example, in the 50's and 60's, which are mainly driven by returns as the capital gains tax rate does not change and capital losses matter for a significant part of the population only at the start of our simulation. The simulation also reproduces the larger increases and declines of capital gains in the 90's and after 2000. Further, there is a level effect in the data in the sense that there always are realizations of capital gains while in our model without liquidity shocks realizations can literally be zero or very close to zero.

Panel (b) in Figure 7 shows the scaled evolution of aggregate capital gains taxes paid versus the scaled simulated taxes paid. Comparing the top with the bottom plot, we see that realized capital gains and taxes collected from realized capital gains correlate highly in the data and the model. This is an expected result, but we caution the reader that since a decrease in the capital gains tax rate leads to an immediate increase in the realization of capital gains for all investors without capital losses the relation between realization and taxes could be negative. Capital losses can also reduce the correlation or even eliminate it. Overall, though, realized capital gains and capital gains taxes paid show positive correlations. Importantly, the simulated capital gains taxes to wealth produce a large contemporaneous

Figure 7: Realized Capital Gains and Tax Collection. The figure shows the evolution of realized capital gains to GDP versus the simulated capital gains to wealth in the top plot and the evolution of capital gains taxes to GDP versus the simulated capital gains taxes to wealth in the bottom plot over the period 1954-2014.



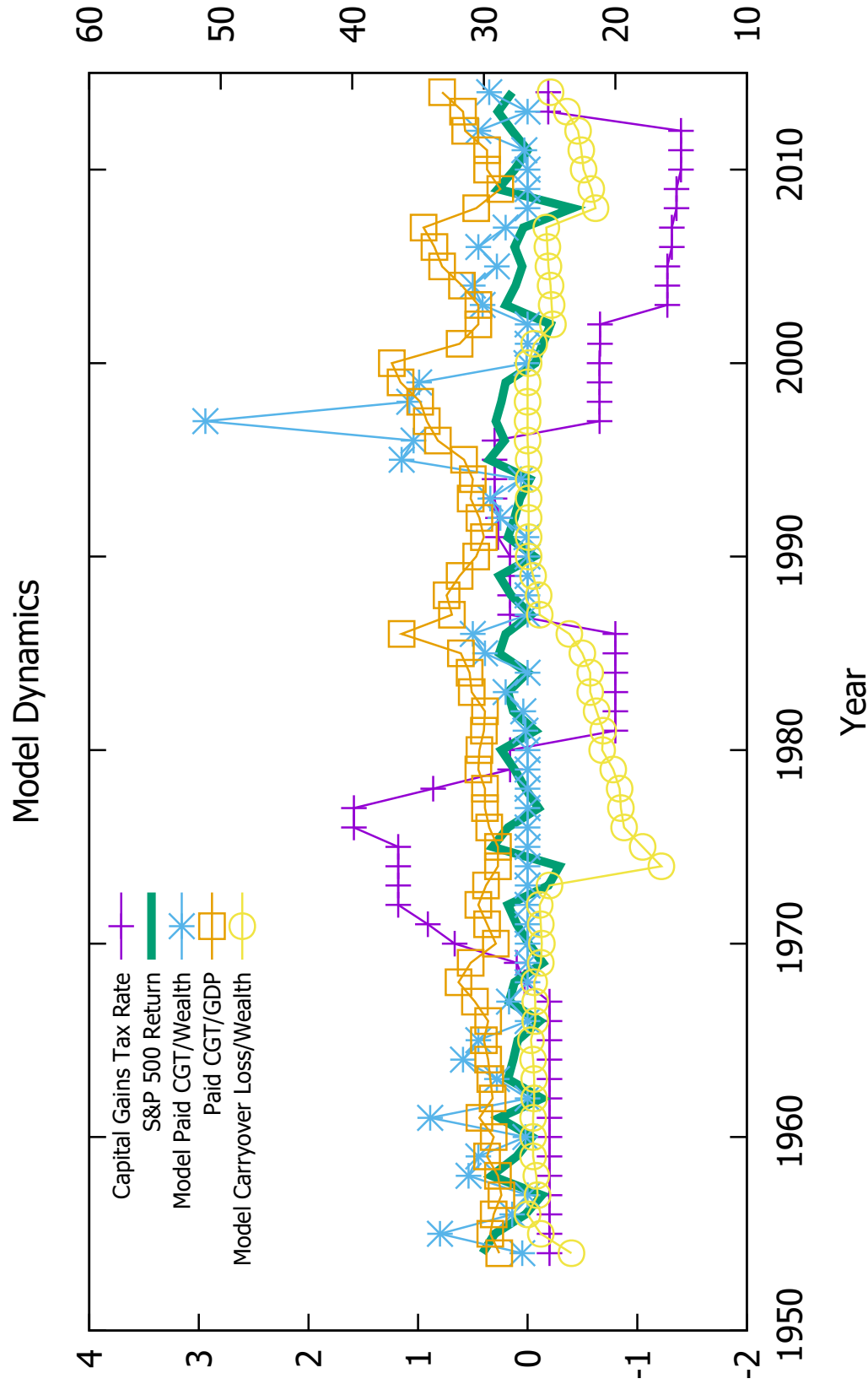
correlation of 41.6% percent with the capital gains taxes to GDP in the data.

Figure 8 sheds light on the results in Figure 7. The figure presents the time series of the aggregate capital gains tax to wealth ratio and carryover loss to wealth ratio over the period 1954 to 2014, simulated by the calibrated model. These time series are juxtaposed against the actual capital gains tax rate, the realized S&P 500 return, and the actual capital gains taxes paid to GDP ratio during the same period. What is striking in the figure is the large amount of carryover losses realized in the early 70's. As explained in Section 6, these losses are a result of past declines in the S&P 500. The accumulated losses imply that, when the capital gains tax rate starts dropping in the mid 70's, the capital gains taxes paid in the simulated model are very small or zero. Capital gains taxes paid only pick up in the early-to-mid 80's. Following an increase in the capital gains tax rate, and positive returns in the late 80's, the simulated population builds large, unrealized, embedded gains, which are realized in the 90's when the capital gains tax rate drops.

According to our simulation, the aftermath of the dot-com boom period and the Great Recession of 2008 also produce large carryover losses across investors that persist for more than a decade. What differentiates these events from the episode in the early 70's is that they do not coincide with a decline in the capital gains tax rate. Consequently, when the capital gains tax rate declines in 1997 and 2003, and when many investors do not have carryover losses, the simulation reproduces the spikes in the collected aggregate capital gains taxes in the subsequent years.

Overall, we see that realized gains vary positively with realized returns on equity, negatively with the capital gains tax rate, and positively with capital losses carried over from previous periods. The capital gains tax collection varies positively with realized returns on equity and negatively with the capital gains tax rate and capital losses.

Figure 8: **Model Capital Gains Tax and Carryover Loss vs. Tax Rate and S&P 500 Return.** The figure shows the time series simulated values of the aggregate capital gains tax to wealth ratio and carryover loss to wealth ratio over the period 1954-2014 vs. the capital gains tax rate, the S&P 500 return, and the actual capital gains taxes paid to GDP ratio.



8 Robustness

Our analysis includes 24 cases. We have two ways to construct the initial cross-section in 1953, namely, assuming no-transition (*) or uniform transition for the evolution of capital gains tax rate.¹⁷ Over the period of 1954 to 2014, we consider that the investor assumes that the capital gains tax rate evolves following a uniform, a historical (*), or a no-transition regime. We use four ways to construct the wealth distribution across age, namely, a model-based one assuming that the initial wealth of each investor is 100 U.S. dollars (*), the 2013 empirical wealth distribution, a uniform wealth distribution, and another model-based wealth distribution where we adjust the initial wealth for inflation (untabulated). For each case, we produce correlation coefficients between realized capital gains in the model and the data and capital gains taxes paid in the model and the data.

Table 1 presents the contemporaneous correlations between model-simulated quantities, where investors use the historical transition regime to rebalance their portfolios, and the quantities from actual data. Panel A presents the correlations when the initial cross-section is constructed assuming that investors follow a strategy that considers the current capital gains tax rate to be permanent, while Panel B presents the correlations when the initial cross-section is constructed assuming that investors follow a strategy that considers that next year's capital gains tax rate is equally likely to take a value from 10% to 40%. Each panel shows the correlation between realized capital gains in the data versus the simulated realized capital gains (CG/GDP vs. CG/Wealth) and the correlation between aggregate capital gains taxes paid versus the simulated taxes paid (CG Tax/GDP vs. CG Tax/Wealth) using the following wealth distributions across age: model-based, 2013 empirical, and uniform wealth distribution.

From Table 1, we see that the correlations between simulated realized capital gains and the data range between 37.8% and 39.4%. The correlations between simulated capital gains taxes paid and the data are larger as they vary between 40.7% and 43.6%. These correlations

¹⁷The (*) denotes the assumptions used in the base case.

Table 1: Correlations between Model Based Realized Capital Gains and the Data and Correlations between Model Based Capital Gains Taxes Paid and the Data. The table presents Pearson correlations between model simulated quantities at time t and data quantities at time t over the period of 1954 to 2014 assuming investors follow the historical transition strategy. Panel A presents the correlations for the case when the initial cross-section is constructed assuming that investors follow the no-transition strategy before 1953, that is, they do not consider the possibility of other tax rates. Panel B presents the correlations for the case when the initial cross-section is constructed assuming that investors follow the uniform transition strategy before 1953. The columns of the table correspond to different aggregation schemes for the individual simulated results: based on the simulated wealth of investors of different ages, based on the 2013 empirical wealth distribution across investors of different ages, and based on a uniform wealth distribution across different ages. CG/GDP is the ratio of realized capital gains over GDP in the data. CG Tax/GDP is the ratio of capital gains taxes paid over GDP in the data. CG/Wealth is the ratio of simulated realized capital gains over aggregate wealth in the model. CG Tax/Wealth is the ratio of simulated capital gains taxes paid over aggregate wealth in the model. All the correlation coefficients have a p-value below one percent (untabulated).

Panel A: Using No-Transition Strategy to Construct Initial Cross-section

Model Wealth Distribution		2013 Wealth Distribution		Uniform Wealth Distribution	
CG/GDP vs.	CG Tax/GDP vs.	CG/GDP vs.	CG Tax/GDP vs.	CG/GDP vs.	CG Tax/GDP vs.
CG/Wealth	CG Tax/Wealth	CG/Wealth	CG Tax/Wealth	CG/Wealth	CG Tax/Wealth
38.1%	41.6%	37.8%	40.7%	38.0%	41.6%

Panel B: Using Uniform Transition Strategy to Construct Initial Cross-section

Model Wealth Distribution		2013 Wealth Distribution		Uniform Wealth Distribution	
CG/GDP vs.	CG Tax/GDP vs.	CG/GDP vs.	CG Tax/GDP vs.	CG/GDP vs.	CG Tax/GDP vs.
CG/Wealth	CG Tax/Wealth	CG/Wealth	CG Tax/Wealth	CG/Wealth	CG Tax/Wealth
39.4%	43.6%	39.2%	42.7%	39.2%	43.4%

are statically significant as all the Pearson correlation coefficients in Table 1 have a p-value below one percent (untabulated) and appear large, that is, they are also economically significant. Further, the correlations are not sensitive to the assumptions on the initial cross-section and the wealth distribution. While correlations do change with the assumptions, their variations are small relative to their levels. Besides, computing correlations as nonparametric Spearman correlation leads mostly to larger coefficients as can be seen from Tables 4 to 6 in the Internet Appendix.

We have shown that the transition probabilities of the tax rate evolution can have large effects on the conditional optimal equity holdings. Yet, in Table 2 and 3 of the Internet Appendix the contemporaneous correlation between the total amount of realized capital gains as a percentage of wealth in the simulation and the total amount of realized capital gains as a percentage of GDP range between 36.0% and 46.5% and exhibit p-values of at most one percent. Hence, our simulation results of aggregate realized capital gains and aggregate capital gains taxes paid are robust in that their correlations with the data do not depend crucially on the assumed or perceived transition probabilities for the capital gains tax rate. The intuition for this is that perceived transition probabilities shift the level of an optimal equity-to-wealth ratio through altering the no-trade region but do not alter the dynamics of the optimal equity-to-wealth ratio around the no-trade region.

9 Conclusion

We have presented a framework for determining the dynamics of aggregate capital gains taxes paid by a population of investors as well as their dependence on variables such as the return on equity, the distribution of embedded gains and carryover losses, and the cross-sectional distribution of wealth. Our framework is based on solving the problem of optimal asset allocation and capital gains realization for an individual investor that faces a changing capital gains tax rate. We aggregate the decisions made by individual investors using various

assumptions on investor wealth. Using historical returns for the S&P 500 and historical US tax rates on capital gains, we calculate aggregate capital gains realizations and capital gains taxes paid in a calibrated simulation model, and show that our results correlate with actual data. This result, which is robust to several modifications of our assumptions, suggests that there is promise in using our model to determine the potential response by investors to changes to the capital gains tax rate.

Our framework can be improved. For example, our model predicts very low capital gains taxes paid when carryover losses are high, or when the capital gains tax rate is high by historical standards, while, in reality, there appears to be a base level of revenue to the government from actual capital gains taxes paid. Our analysis could also improve if we had a better source of data for the population distribution of embedded gains, carryover losses, and portfolio holdings across different age cohorts. Nonetheless, our results speak to the data and can form the basis for further research.

A strong motivation for our work derives from the policy debate on how responsive realizations of capital gains and capital gains tax collection are to the level of the capital gains tax rate. Taking prices and tax rates as given, we see from our work that a decline in the capital gains tax rate leads to a large temporary increase in the realization of capital gains. The collected capital gains taxes can, but do not have to, show a temporary increase, it depends on the extent of capital losses in investors portfolios. While, without an equilibrium analysis, we cannot speak to the sign of the permanent effect, given our partial equilibrium analysis we expect its absolute value to be significantly smaller than that of the temporary effect. This is because, after accounting for the temporary effect of a decline in the tax rate, the optimal equity-to-wealth level drops but its dynamics around the new upper boundary of the no-trade region remain largely the same as before the decline.

A The Consumption-Portfolio Choice Problem

The assumptions and notation of our consumption-portfolio choice problem closely follows Dammon, Spatt, and Zhang (2001), Gallmeyer, Kaniel, and Tompaidis (2006), and Ehling, Gallmeyer, Srivastava, Tompaidis, and Yang (2018).

Time is discrete, with $t = 0, \dots, T$. The riskless money market pays a continuously compounded time-varying interest rate $r(t)$, and the risky stock pays out a time-varying dividend $\delta(t) S(t)$ with an ex-dividend price of $S(t)$, where $\delta(t)$ denotes the time-varying dividend yield. An investor enters time t with wealth $W(t)$, where $\underline{\pi}(t)$ is invested in the stock and $1 - \underline{\pi}(t)$ is invested in the money market account. At time t , the investor rebalances his portfolio, consumes, and pays taxes on interest and dividend income and, if he sells equity with embedded gains, on realized capital gains. The investor exits time t with an exiting stock portfolio, $\bar{\pi}(t)$, which will be held until time $t + 1$.

A.1 Dividend and Interest Income Taxation

Dividend and interest income are taxed at constant rates τ_D and τ_I , implying that the tax-adjusted excess return of the stock and the tax-adjusted risk-free rate, respectively, are

$$\begin{aligned} R(t) &= \frac{S(t+1)}{S(t)} [1 + \delta(t+1)(1 - \tau_D)] - R_f(t), \\ R_f(t) &= (1 - \tau_I) \exp(r(t)) + \tau_I. \end{aligned} \tag{A.1}$$

A.2 Capital Gains and Losses

The basis-to-price ratio, $b(t)$, is the relative tax basis of the equity position in dollar terms divided by the current level of the stock price, $b(t) = B(t-1)/S(t)$. The absolute tax basis $B(t)$ is the weighted-average purchase price of a stock after rebalancing at time t . With an embedded capital gain, the basis-to-price ratio is smaller than one, while with an embedded capital loss the basis-to-price ratio is larger than one. With an embedded capital gain, selling results in a realization of the embedded capital gain and its taxation. With an embedded capital loss, selling occurs without taxation. Any capital loss is realized immediately and entirely. Thus, the current realized capital gain or loss, expressed as a proportion of wealth, is

$$g(t) = (1 - b(t))^+ (\underline{\pi}(t) - \bar{\pi}(t))^+ - (b(t) - 1)^+ \underline{\pi}(t), \tag{A.2}$$

where $x^+ \triangleq \max(x, 0)$.

A.3 The Capital Gains Tax

We allow the capital gains tax rate, τ_C , to vary over time. The capital gains tax paid at time t , expressed as a proportion of wealth, is

$$\phi(t) = \tau_C(t) (g(t) - l(t))^+, \tag{A.3}$$

where $l(t) \geq 0$ denotes the carryover loss, scaled by wealth. Capital losses are carried forward to the next period according to

$$l(t+1) = \left(\frac{W(t+1)}{W(t)} \right)^{-1} (l(t) - g(t))^+, \quad (\text{A.4})$$

where the wealth growth rate appears since $l(t)$ and $l(t+1)$ are scaled by $W(t)$ and $W(t+1)$, respectively.

A.4 Trading

Wealth evolves according to

$$\frac{W(t+1)}{W(t)} = \bar{\pi}(t) R(t) + [1 - c(t) - \phi(t)] R_f(t), \quad (\text{A.5})$$

where an admissible trading strategy satisfies a short-selling constraint and a margin constraint

$$\bar{\pi}(t) \geq 0, \quad \text{and} \quad (1 - m_+) \bar{\pi}(t) \leq 1 - c(t) - \phi(t), \quad (\text{A.6})$$

respectively, and where $1 - m_+$ denotes the marginable fraction of equity. Holding a portfolio from time t to $t+1$, leads to an entering portfolio

$$\underline{\pi}(t+1) = \bar{\pi}(t) \frac{S(t+1)}{S(t)} \left(\frac{W(t+1)}{W(t)} \right)^{-1}, \quad (\text{A.7})$$

where the change in the portfolio weight is due to the change in stock price and wealth.

A.5 The Basis-to-Price Ratio

The evolution of the basis-to-price ratio depends on the evolution of the stock price and the investor's trading. With an embedded capital loss in the stock at time t , the investor immediately liquidates the stock. After liquidation, the time- t basis-to-price ratio resets to $b(t) = 1$ because any new purchase of the stock is at the current market price. The basis-to-price ratio next period simply reflects the return of the stock from time t to $t+1$; i.e., $b(t+1) = S(t)/S(t+1)$.

If an investor reduces an equity position with an embedded capital gain at time t , then the tax basis is unchanged and the return of the stock drives the basis-to-price ratio next period; i.e., $b(t+1) = [S(t)/S(t+1)] b(t)$. If the investor decides to purchase additional shares of the stock at time t while facing an embedded capital gain in the stock, the resulting position is made up from past-purchased shares and newly-purchased shares. The past-purchased shares have weight $\underline{\pi}(t)$ and basis-to-price ratio $b(t)$ and the newly-purchased shares have weight $\bar{\pi}(t) - \underline{\pi}(t)$ and basis-to-price ratio of 1. The combined position of past-purchased and newly-purchased shares has a basis-to-price ratio that is the weighted-average of $b(t)$ and 1 adjusting for the stock return $S(t)/S(t+1)$: $b(t+1) = [S(t)/S(t+1)] [\underline{\pi}(t) b(t) + (\bar{\pi}(t) - \underline{\pi}(t))] / \bar{\pi}(t)$.

Combining the cases, the basis-to-price ratio evolves as

$$b(t+1) = \begin{cases} \frac{S(t)}{S(t+1)}, & \text{if } \pi(t) = 0 \text{ or } b(t) > 1, \\ \frac{S(t)}{S(t+1)} \left[\frac{\pi(t)b(t) + (\bar{\pi}(t) - \pi(t))^+}{\pi(t) + (\bar{\pi}(t) - \pi(t))^+} \right], & \text{otherwise.} \end{cases} \quad (\text{A.8})$$

A.6 Investor Utility

The investor derives utility from real consumption and from terminal wealth at the time of death through a bequest to his heirs. The investor's utility over consumption is given by a function with constant relative risk aversion γ . His time preference is determined by a time preference parameter β . When the investor dies, all assets are liquidated and distributed to his heirs as an equal amount paid each year in real terms.

When the investor dies, the assets totaling $W(t)$ are liquidated and distributed to his heirs as an equal amount of payment, p , each year in real terms for η periods starting from time t . We assume that the heirs have the same preferences as the investor and that the investor places the same weight on utility from own consumption and utility from heirs' consumption, where changing weights on the investor's own consumption and his heirs' consumption is equivalent to changing the number of bequest years.

The annual payment to the investor's heirs, p , is calculated as

$$W(t) = p + e^{\bar{i} - \bar{r}} p + e^{2(\bar{i} - \bar{r})} p + \dots + e^{(\eta-1)(\bar{i} - \bar{r})} p \Rightarrow p = \frac{1 - e^{\bar{i} - \bar{r}}}{1 - e^{(\bar{i} - \bar{r})\eta}} W(t), \quad (\text{A.9})$$

where $\bar{i}$ and $\bar{r}$ are the inflation rate and the after-tax interest rate that apply to the η periods of bequest.¹⁸ Thus, the total bequest utility of those η payments is

$$u(p) + \beta u(p) + \beta^2 u(p) + \dots + \beta^{\eta-1} u(p) = \alpha u(W(t)), \quad (\text{A.10})$$

where the bequest factor α is

$$\alpha = \frac{1 - \beta^\eta}{1 - \beta} \left(\frac{1 - e^{\bar{i} - \bar{r}}}{1 - e^{(\bar{i} - \bar{r})\eta}} \right)^{1-\gamma}. \quad (\text{A.11})$$

A.7 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 wealth. According to the principle of dynamic programming, the investor's lifetime

¹⁸We set the inflation rate and the nominal interest rate equal to their historical averages, see Table 1 of the Internet Appendix.

consumption and portfolio choice problem is described by a Bellman equation

$$V(t, \pi(t), b(t), l(t), \tau_C(t)) = \max_{c(t), \pi(t)} e^{-\lambda_t} \frac{c(t)^{1-\gamma}}{1-\gamma} + (1 - e^{-\lambda_t}) \frac{\alpha}{1-\gamma} \\ + e^{-\lambda_t} \beta E_t \left[\left(\frac{e^{-i(t)} W(t+1)}{W(t)} \right)^{1-\gamma} V(t+1, \pi(t+1), b(t+1), l(t+1), \tau_C(t+1)) \right], \quad (\text{A.12})$$

for $t = 0, 1, \dots, T-1$ subject to the wealth evolution equation (A.5), the stock proportion dynamics (A.7), the basis-to-price ratio evolution equation (A.8), the carryover loss evolution equation (A.4), the no-short selling and margin constraints (A.6), the realized capital gains equation (A.2), and the capital gains taxation equation (A.3). In equation (A.12), the expectation is taken with respect to the exogenous distribution of the stock returns and the predictive distribution of $\tau_C(t+1)$ given $\tau_C(t)$.

If the investor survives at time t , which he does 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 , and the wealth growth rate $W(t+1)/W(t)$. Otherwise, if the investor dies, he receives utility from his bequest to his heirs.

We numerically solve the individual investor's problem using the methodology described in Ehling, Gallmeyer, Srivastava, Tompaidis, and Yang (2018).

B Wealth Distribution in the Model

We denote the population wealth distribution at time t by $\{\Omega_j(t)\}_{j=20}^J$, where $\Omega_j(t)$ is the total wealth of all age- j individuals at time t scaled by the total population wealth at time t . Thus, $\sum_{j=20}^J \Omega_j(t) = 1$ at any time t . Given the cross-section of individuals' paid capital gains taxes-to-wealth ratio at time t , $\{\phi_j(t)\}_{j=1}^J$, we compute the population aggregate paid capital gains taxes scaled by population wealth as a weighted sum, $\sum_{j=20}^J \phi_j(t) \Omega_j(t)$. In a similar way, we can compute the population aggregate realized capital gains scaled by population wealth.

The model-driven wealth distribution is obtained by multiplying two terms at each age j , namely the individual investor's wealth at age j , $W_j(t)$, and the population density at age j , μ_j . Scaling by the sum across all ages leads to $\Omega_j(t) = W_j(t) \mu_j / \sum_{j=20}^J W_j(t) \mu_j$. We obtain individual investor wealth at each age from our simulation of individual investors in Step 1 and Step 2; thus, this is a model-driven quantity that depends on investor's initial wealth at age 20 and his portfolio rebalancing strategy. We obtain the population density at each age using the age-dependent hazard rate in the 1990 US life table and a constant population growth rate of 1 percent. Given the conditional survival probability from age $j-1$ to j , ψ_j , and the constant population growth rate, ρ , the population density is determined through $\mu_{j+1} = \frac{\psi_{j+1}}{1+\rho} \mu_j$ and $\sum_{j=20}^J \mu_j = 1$. The population density is time-independent while the cross-section of investors' wealth is time-dependent. Hence, the model-driven wealth distribution is time-dependent.

References

- Auerbach, Alan J, 1988, Capital gains taxation in the United States: Realizations, revenue, and rhetoric, *Brookings Papers on Economic Activity* 19, 595–638.
- Auten, Gerald E., and Charles T. Clotfelter, 1982, Permanent versus transitory tax effects and the realization of capital gains, *Quarterly Journal of Economics* 97, 613–632.
- Bogart, W. T., and William M. Gentry, 1995, Capital gains taxes and realizations: Evidence from interstate comparisons, *Review of Economics and Statistics* 77, 267–282.
- Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud, 2005, A simulation approach to dynamic portfolio choice with an application to learning about return predictability, *Review of Financial Studies* 18, 831–873.
- Brandt, Michael W., and Pedro Santa-Clara, 2006, Dynamic portfolio selection by augmenting the asset space, *Journal of Finance* 61, 2187–2217.
- Burman, Leonard E., and William C. Randolph, 1994, Measuring permanent responses to capital-gains tax changes in panel data, *American Economic Review* 84, 794–809.
- Campbell, John Y., João Cocco, Francisco Gomes, Pascal J. Maenhout, and Luis M. Viceira, 2001, Stock market mean reversion and the optimal equity allocation of a long-lived investor, *Review of Finance* 5, 269–292.
- Campbell, John Y., and Luis M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics* 114, 433–495.
- , 2001, Who should buy long-term bonds?, *American Economic Review* 91, 99–127.
- Dammon, R., C. Spatt, and H. Zhang, 2001, Optimal consumption and investment with capital gains tax, *Review of Financial Studies* 14, 583–616.
- Dowd, Tim, Robert McClelland, and Athiphat Muthitacharoen, 2015, New evidence on the tax elasticity of capital gains, *National Tax Journal* 68, 511–544.
- Dumas, Bernard, and Elisa Luciano, 1991, An exact solution to a dynamic portfolio choice problem under transaction costs, *Journal of Finance* 46, 577–595.
- Ehling, Paul, Michael Gallmeyer, Sanjay Srivastava, Stathis Tompaidis, and Chunyu Yang, 2018, Portfolio tax trading with carryover losses, *Management Science* 64, 3971–4470.
- Feldstein, Martin, Joel Slemrod, and Shlomo Yitzhaki, 1980, The effects of taxation on the selling of corporate stock and the realization of capital gains, *Quarterly Journal of Economics* 94, 777–791.
- Gallmeyer, Michael, Ron Kaniel, and Stathis Tompaidis, 2006, Tax management strategies with multiple risky assets, *Journal of Financial Economics* 80, 243–291.

- Garlappi, Lorenzo, and Georgios Skoulakis, 2010, Solving consumption and portfolio choice problems: The state variable decomposition method, *Review of Financial Studies* 23, 3346–3400.
- Gârleanu, Nicolae, and Lasse Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.
- Merton, Robert C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.
- Poterba, James M., 2002, Taxation, risk-taking, and household portfolio behavior, in A. J. Auerbach, and M. Feldstein, ed.: *Handbook of Public Economics*, vol. 3 of *Handbook of Public Economics* . chap. 17, pp. 1109–1171 (Elsevier).
- Sialm, Clemens, 2009, Tax changes and asset pricing, *American Economic Review* 99, 1356–1383.