Economics 230a, Fall 2013
Lecture Note 11: Tax-Favored Retirement Saving, Capital Gains and Estate Taxation

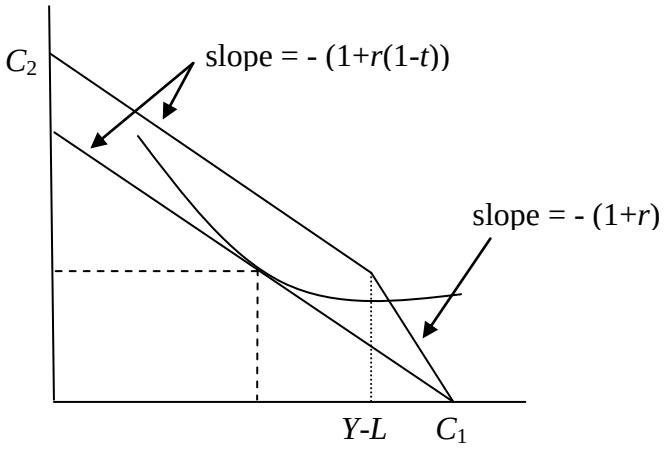
Three important elements of the tax treatment of asset accumulation are (1) the special treatment afforded retirement saving; (2) capital gains taxation; and (3) estate taxation.

Tax-Favored Retirement Saving

As already discussed, one way of implementing a consumption tax is to allow a deduction for saving from an income tax. Although the United States does not treat savings in general this way, it does provide such treatment for specially designated retirement saving, through employer-based accounts (known as 401(k) plans) and individual accounts, known as Individual Retirement Accounts (IRAs). The same is true in many other countries. Although the traditional tax treatment of such accounts involves the "consumption tax" approach – deduction of contributions to accounts and taxation of withdrawals – an alternative approach (in the United States, called Roth IRAs and Roth 401(k)s) is to ignore contributions and withdrawals as well as not taxing accruing income – as capital income would be treated under a labor income tax. These alternative approaches are often summarized as "EET" and "TEE" methods of taxation (T for taxable, E for tax-exempt), where the first letter corresponds to the treatment of contributions, the second to accruing income, and the third to withdrawals. As we discussed earlier, these two approaches both eliminate the tax on new saving if tax rates are constant over time. (As we also discussed, under a comprehensive tax that applies to all saving, the two approaches differ as well in the treatment of existing assets, if withdrawals of preexisting assets are taxed under the consumption-tax approach, but this distinction does not arise here, for the special tax treatment applies only to assets in the specified accounts, not to other assets.)

Tax-favored retirement accounts also typically have annual caps on the level of contributions, which would constrain individuals who wish to contribute large amounts. Also, contributions do not have to represent new saving – individuals can contribute by transferring existing assets from regular taxable accounts to retirement saving accounts, either directly (in the case of IRAs) or indirectly (in the case of 401(k)s) by reducing balances in regular taxable accounts in order to finance current consumption while at the same time having contributions to 401(k) plans deducted from their salaries by employers.

Thus, tax-favored saving accounts differ from the tax treatment of savings under a consumption tax or a labor income tax in three important respects: (1) they apply only to saving for retirement, and have various other provisions (such as early withdrawal penalties) that make them less attractive for saving for pre-retirement consumption; (2) they have annual contribution limits; and (3) they apply to balances in specified accounts, rather than all assets. In light of these differences, what should we expect the impact on saving to be? Consider first a simple two-period life-cycle example, in which individuals work in the first period and save what is not initially consumed for retirement consumption. Without special retirement saving incentives, the individual faces the original budget line, with slope $-(1+r(1-t))$, where t is the capital income tax rate, and chooses the point of tangency shown.



When the accounts are introduced, the individual can save up to limit L in tax-favored form, receiving a rate of return of r; further saving still faces capital income taxation. Thus, there will be only an income effect if desired saving exceeds L. The same logic applies in the case where individuals have pre-existing assets, in which case they will hit the contribution constraint as long as their existing wealth plus new saving exceeds L.

On the other hand, some factors may make it more likely that contributions to accounts have more than income effects. First, to the extent that withdrawal restrictions make retirement saving an imperfect substitute for other saving, individuals may not wish to transfer all assets into retirement saving accounts; thus, they may be on the steeper slope of the outer budget line in the figure. Also, behavior beyond the standard rational choice assumptions may lead individuals to save more if there is a specific reason for saving in these accounts or if the saving is coordinated by employers. Empirical evidence is somewhat mixed, in part because it is difficult to come up with clean natural experiments to assess the treatment of saving incentives, because "treatment" has typically differed across individuals only because of differences in characteristics that may also be associated with unobserved differences in saving propensities. For example, individuals who work for employers offering 401(k)s may differ from those who work for employers not offering 401(k)s. See Bernheim's Handbook chapter, section 4. Gelber's paper is a recent attempt to control for unobserved differences among individuals, by considering individuals who had to wait to qualify for 401(k) plans. In general, the evidence does suggest that retirement saving responds positively to favorable tax treatment. There is also considerable evidence that individuals are influenced by employer education efforts (as discussed by Bernheim, section 5.4). Recent research also suggests that individuals are strongly influenced by "default" policies of employers – that they are much more likely to participate in retirement saving if that is the default employers present, rather than if they must make the decision to enroll. Another strategy for inducing employees to participate is to require an active decision whether or not they wish to participate rather than giving them the option of passive non-participation. Carroll et al. compare active decisions and defaults. While both increase participation, they argue that requiring an active decision is preferable to a specific default if tastes for saving are heterogeneous (but not if financial illiteracy provides an argument for overriding individual preferences).

Another interesting decision is what investments to hold in tax-favored saving accounts. In particular, if individuals have both tax-favored accounts and taxable accounts, how should their portfolio allocations vary across the accounts? One's first intuition might be that individuals should base each account's composition on the account's tax treatment and the individual's risk tolerance. For example, an individual who is not too risk averse might choose to hold a lot of equity in a taxable account, and less equity in a tax-favored account, since (following the logic of the discussion in Lecture 10), the tax-favored account does not derive any additional benefit from the favorable tax treatment that equity receives. But this logic misses an important point: the same person is holding the two accounts, and should optimize simultaneously while choosing

their composition. This leads to a simple arbitrage argument that individuals should concentrate lightly taxable assets in taxable accounts and more fully taxable assets in tax-favored accounts. The demonstration of this result proceeds by contradiction. Suppose that there are two assets, for example fully taxed debt and partially taxed equity, and that the investor has both a tax-favored account and a taxable account, each with some equity and some debt. Then the individual can increase debt in the tax-favored account while reducing debt in the taxable account, offsetting the changes in each account with equal changes in equity, in such a way that the overall portfolio risk is unchanged but the expected after-tax rate of return is increased. See Poterba's Handbook chapter, section 4.3. This arbitrage argument does not hold exactly because of the withdrawal restrictions on tax-favored retirement accounts, but allowing explicitly for this factor does not have an important impact; see Dammon, Spatt and Zhang. Nevertheless, as Poterba discusses, evidence suggests that individuals adopt similar asset allocations in their tax-favored and taxable accounts. This behavior might be due to a tendency of individuals to maintain different "mental accounts" for different types of saving, but this is an open question.

Capital Gains Taxation

Capital gains are taxed upon *realization* rather than on accrual. This factor makes capital gains taxation complex and subject to a variety of potential taxpayer responses, which is one reason why capital gains taxation is an interesting topic. Another reason to be interested in capital gains taxation is that capital gains are very concentrated among high-income individuals.

What does realization-based taxation do? Consider a two-period model in which an investor has an asset purchased in an earlier period for $1, which has already appreciated in value by an amount g. The investor can either hold the asset for another period, earning an additional return r, or sell and earn the market rate of return i. Suppose all income is taxed at rate t, but only when assets are sold. Also suppose that the investor's objective is to maximize terminal wealth.

If the investor sells the asset and reinvests, terminal wealth is:

$$W_R = (1+g(1-t))(1+i(1-t)) = (1+g)(1+i) - t[g(1+i(1-t) + (1+g)i]$$

If the investor holds the asset until the end of the second period, terminal wealth is:

$$W_H = (1+g)(1+r) - t[(1+g)(1+r) - 1] = (1+g)(1+r) - t[g + (1+g)r]$$

Comparing the terms in brackets in the second version of each expression, we can see that the "hold" strategy enjoys a tax advantage over the "realize" strategy – first period gains, g, are taxed one period earlier under the latter, and hence the tax liability has a higher accumulated value at the end of the second period because it is multiplied by $1+i(1-t)$. It follows that if $i = r$, the investor will choose to hold rather than to realize, and indeed that there is a range of values of $r < i$ for which it will still be optimal to hold rather than to sell. This phenomenon is known at the lock-in effect – in order to defer tax on previously accumulated gains, individuals have an incentive not to sell assets even when, for non-tax reasons, they would prefer to sell. In this example, the lock-in effect is associated with the investor's willingness to accept a lower before-tax rate of return, but in a realistic setting the major distortion comes from an inefficient

allocation of assets across investors. That is, when an individual realizes a capital gain by disposing of an asset, that asset does not typically disappear, but instead ends up in someone else's portfolio. Thus, it is unlikely simply to have a below-market rate of return. Rather, in a setting with risky assets, other investors may be willing to pay more for the asset than the individual currently holding it. For example, suppose that there are two investors, one holding appreciated stock in Apple and the other holding appreciated stock in General Electric. As these two assets are not perfectly correlated, a combined portfolio would offer a better risk-return trade-off than either specialized position. Absent taxation, each investor could be made better off by trading, but if each faces the capital gains tax, the gains from trade may not be realized.

The lock-in effect is exacerbated by two other provisions found in the US tax system and typical of others as well. First, gains on assets held for at least one year are taxed at a lower rate (in United States at present, a maximum of 20% vs. a maximum of 39.6%). Second, gains held until death are not taxed at all. On the other hand, the lock-in effect is reversed when an asset has gone down in value ($g < 0$ in the above example), since deferral of tax in this case means deferring a tax *refund*. Thus, individuals have an incentive to hold gains and realize losses, meaning that those with large numbers of distinct positions in different assets could, on a regular basis, achieve liquidity by "harvesting" losses without having to realize gains. This possibility, in turn, is largely responsible for another tax provision, which limits the annual value of deductible losses (in excess of realized gains) to $3,000. Unfortunately, as discussed in Lecture 10, a limit on the deductibility of losses also discourages risk-taking.

Empirical Evidence on Responses to Capital Gains Taxation

There has been a substantial literature relating capital gains realizations to capital gains tax rates. One of the key issues is the need to distinguish between short-run and long-run responses. We would expect that a change in tax rates could have a large impact on the timing of realizations, because individuals can adjust the timing of their asset sales. For example, after the October, 1986 passage of the Tax Reform Act of 1986, which increased the capital gains tax rate on high-income individuals from 20% to 28% effective January 1, 1987, there was such a surge in realizations in the remainder of 1986 that realizations for that year were approximately twice as much as those in 1985 and 1987. But that doesn't mean that we would expect realizations to be permanently twice as high under a 20% tax rate than under a 28% tax rate. One standard approach, developed using panel data by Burman and Randolph (1994; hereafter B-R), and discussed in Poterba, section 3.2, estimates the specification:

$$(1) \qquad \ln g_{it} = X_{it}\beta + \gamma_1 \tau_{it}^p + \gamma_2\left(\tau_{it} - \tau_{it}^p\right) + \gamma_3(\tau_{it} - \tau_{it-1}) + \varepsilon_{it}$$

where g is capital gains, X is a vector of individual attributes, τ is the individual's capital gains tax rate, and τ^p is a measure of the individual's "permanent" tax rate. There are three econometric issues that must be confronted in estimating (1). The first is that realized gains may be zero; a Tobit estimator is used to deal with this. The second issue is that the capital gains tax rate may depend on the level of gains realized, since tax rates rise with income. To deal with this, B-R uses as an instrument for τ a so-called "first-dollar" tax rate – the capital gains tax rate the individual would face on the first-dollar of capital gains realized. The third issue is how to define the individual's "permanent" tax rate. B-R represent this by regressing τ on the maximum federal plus state capital gains tax rate in year t in the state where individual i lives, as well as

other individual attributes *X* (but not the first-dollar tax rate in year *t*). The rationale is that if an individual's tax rate fluctuates over time due to changes in individual circumstances, such as other income or deductions, then this will affect τ but not τ^p. From their estimates, B-R find a long-term elasticity (based on the coefficient γ_1), taking account of both extensive and intensive responses in the Tobit (i.e., to realize gains and how many gains to realize), of close to zero, and a short-term elasticity (based on the sum of the coefficients $\gamma_1 + \gamma_2 + \gamma_3$) of larger than 6 in absolute value. They conclude that virtually all observed responses of capital gains to tax rates involve timing of realizations, rather than changes in the underlying frequency of realizations. This has important implications for considerations of policy changes, for it means that even though revenues may increase in the short run in response to a reduction in capital gains tax rates (since the short-run elasticity exceeds 1 in absolute value), the opposite is true in the long run.

One critique of the B-R specification is that the use of the maximum federal plus state tax rate to identify τ^p does not correctly distinguish timing and permanent responses. On the one hand, the maximum state and federal tax rates change over time during the sample period, so some of the responses to τ^p may be timing responses, which would tend to overstate the estimate of the long-run response. On the other hand, individual tax rates may persistently vary from τ^p, meaning that some of the response classified as temporary may actually be permanent, which would tend to understate the estimated long-run response. To deal with this issue, Auerbach and Siegel (A-S) replace τ_{it}^p with τ_{it+1}, the tax rate the individual will face the following year, which is generally known at time *t*, and add the maximum federal plus state tax rate from year *t*+1 as an instrument, also including the first-dollar tax rate as an instrument for τ_{it+1}. The notion here is that next period's tax rate for the individual is a better measure of the individual's "long-run" tax rate than the current year's maximum tax rate. A-S also find a temporary elasticity that is much higher than the permanent elasticity, but their permanent elasticity is substantially greater than 1 in absolute value – much higher than that found using the B-R methodology for the same sample. One other finding by A-S is that individuals who are wealthier or more sophisticated (based on the nature of their transactions) have a higher temporary elasticity and a lower permanent elasticity. The first of these results, especially, may indicate more careful tax planning.

A further empirical finding of interest is by Ivković, Poterba and Weisbenner, who consider differences in capital gains realizations by individuals who hold both tax-favored and taxable accounts. According to standard theory, there should be no lock-in effect for assets in tax-favored accounts, so that gains should be realized sooner, and losses later, than in taxable accounts. Indeed, the authors find that, *relative to assets in their tax-favored accounts*, investors are less likely to realize gains and more likely to realize losses in their taxable accounts (Figure 3B**)**. However, they also find that investors are more likely to realize taxable gains than taxable losses (Figure 1). There are a variety of possible explanations for this latter finding, including a belief that stock prices are mean-reverting (so that those with gains are expected to fall and those with losses are expected to rise), a need to rebalance portfolios (and hence to shed those stocks that have gained and as a result occupy a larger portfolio share), and the presence of a "disposition effect," by which individuals perceive losses more fully if they are realized.

Reforming the Capital Gains Tax

Some changes in the capital gains tax (such as taxing capital gains at death) could serve to reduce the lock-in effect, but other problems remain as long as the basic approach to taxing

capital gains upon realization is followed. What other alternatives exist? One simple idea would be to tax capital gains as they accrue, rather than upon realization (perhaps combined with a reduced rate to offset the increased present value of taxes). But there are two problems with this approach: (1) taxpayers may lack liquidity to pay taxes until assets are actually sold; and (2) the government may not know the value of some assets until they are actually sold. One proposal for dealing with the liquidity problem, by Vickrey (1939), amounts to keeping an account of accruing gains and the associated tax liability and charging interest on this accruing unpaid balance until asset sale. That is, the tax liability as of date s would evolve according to:

$$(2) \qquad T_{s+1} = [1+i(1-t)]T_s + tr_sA_s$$

where r_s is the rate of return at date s, A_s is the value of the asset at date s, i is the safe rate of interest and t is the tax rate. Vickrey's approach did not deal with the problem that r_s and A_s are unobservable, but Auerbach (AER 1991) argued that one can generalize Vickrey's approach to:

$$(3) \qquad T_{s+1} = [1+i(1-t)]T_s + tiA_s + t^*(r_s-i)\,A_s$$

where t^* can take on any value, since (as discussed in Lecture 10), a tax rate on a risky asset's return in excess of the safe rate has no effect on the investor's opportunities. Auerbach then showed that a tax liability of the form:

$$(4) \qquad T_{s+1} = \left[1 - \left(\frac{1+i(1-t)}{1+i}\right)^s\right] A_s$$

satisfies (3). Note that the only information needed to assess the tax in (4) is the sale price, A_s, the holding period, s, the safe rate of interest, i, and the tax rate, t, all observable. Auerbach and Bradford generalize this result and show how it can be implemented using a tax system based exclusively on observed cash flows, without having to keep track of individual assets.

Estate Taxation

Taxation of estates (or inheritances, if levied on the recipients rather than those leaving the estate) is interesting for many of the same reasons that capital gains taxes are. Estate taxation hits only individuals near the top of the income and wealth distribution (typically 1-2% of decedents each year), and is also subject to tax planning that can make the tax base responsive to tax rates. But, in addition, as mentioned in Lecture 7, there is the issue of why individuals leave estates. One explanation is simply that individuals preserve assets in old age to avoid running out of wealth should they live longer than expected (which in the absence of a perfect annuity market they deal with through precautionary saving). Some sort of bequest motive is typically thought to be important as well, particularly for very large estates where the precautionary saving motive is not plausible. Except for the precautionary saving motive, estate sizes should respond to the magnitude of the estate tax. Kopczuk surveys empirical research on the various types of responses to estate taxation, including charitable bequests (which are tax-deductible) and lifetime gifts (which are tax-favored relative to estates). One interesting question is why estates subject to tax are as large as they are, in light of various options available to transfer wealth to heirs with little or no tax cost during one's lifetime. Some evidence suggests that part of the explanation is a desire to hold wealth while alive (beyond what is needed for precautionary saving).