Internet Appendix to "Should the government be paying investment fees on $3 trillion of tax-deferred retirement assets?"

Mattia Landoni* and Stephen P. Zeldes†

This version: January 12, 2020

Paper available at `https://ssrn.com/abstract=3046077`
Internet Appendix available at `https://ssrn.com/abstract=3138303`

Abstract – This Internet Appendix contains supplementary materials for Landoni and Zeldes (2020). Section 1 contains certain derivations for our general equilibrium model. Section 2 examines how, in the partial equilibrium model in Section 2 of the paper, alternative arrangements regarding the deductibility of fees influences the incidence of the higher fees that occur under Traditional. Section 3 provides additional background information on fees. Section 4 calibrates the size of the implicit government account and total fees paid on it for six additional countries, similar to what is done in Section 4 of the main text for the US.

Keywords: mutual fund fees, taxes, retirement savings
JEL: D14, G11, G23, G28, G51, H21, J26, J32

*Assistant Professor of Finance, Edwin L. Cox School of Business, Southern Methodist University
†Benjamin M. Rosen Professor of Finance and Economics, Columbia Business School, Columbia University; and NBER

1 Proofs and derivations

Please note that this section is not meant to be read sequentially. Each subsection of this section is a stand-alone derivation.

1.1 Derivation of the planner's simplified objective function

The planner maximizes

$$U = \max_{\{C_{0,i}\},\{C_{1,i}\},N} \int_0^1 \ln C_{0,i} + \delta \ln C_{1,i} - \gamma \ln d_{i,j} \; di. \tag{1}$$

Note that it is optimal for the planner to give equal consumption to all individuals ($C_{0,i} = C_0$ and $C_{1,i} = C_1$) because the utility function is concave and separable in its arguments. Then, using the assumption that the N firms are located equidistantly along the circle, the planner's objective function simplifies to

$$U = \max_{C_0,N} \ln C_0 + \delta \ln C_1 - \gamma \cdot 2N \int_0^{1/(2N)} \ln i \; di, \tag{2}$$

or

$$U = \max_{C_0,N} \ln C_0 + \delta \ln C_1 + \gamma \ln N - \gamma \left(1 + \ln 2\right). \tag{3}$$

1.2 Equilibrium fees

This subsection closely follows the solution of Salop (1979) and Tirole (1988, Ch. 7). Firms choose fees to maximize profits:

$$\max_{F_j, f_j} \pi_j = \left[F_j + \left(f_j^v - c\right) S_j\right] q_j - \phi, \tag{4}$$

In this setting, there exists a symmetric equilibrium such that all firms charge the same fees:

$$F_j = F \qquad \forall j \in \{1,\, 2, \ldots\, N\}\,; \tag{5}$$

$$f_j^v = f^v \qquad \forall j \in \{1,\, 2, \ldots\, N\}\,. \tag{6}$$

The equilibrium fee structure (F, f^v) is such that no firm has an incentive to switch to a different structure given that every other firm is also charging fees with the same structure (F, f^v). Consider the situation of the marginal investor i living between firms j and $j+1$

who is indifferent between the two firms.[1] For this marginal investor, the distance from the two adjacent firms is such that

$$\ln C_{0,i,j} + \delta \ln C_{1,i,j} - \gamma \ln d_{i,j} = \ln C_{0,i,j+1} + \delta \ln C_{1,i,j+1} - \gamma \ln d_{i,j+1}, \tag{7}$$

where $C_{t,i,j}$ is i's time-t consumption conditional on choosing firm j.

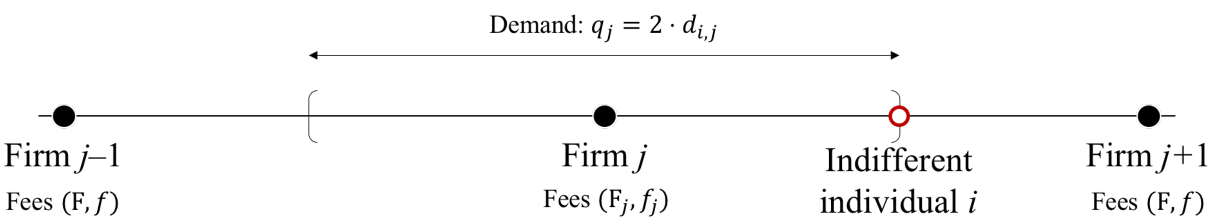

Figure 1: Geometric intuition for the calculation of firm-level demand.

Firm j chooses $\left(F_j, f_j^v\right)$ taking all other firms' fees as given, and therefore $\left(F_{j+1}, f_{j+1}^v\right)$ is simply the equilibrium fee structure, (F, f^v) . Moreover, since the distance between firms is $1/N$, the distance of the investor from firm $j+1$ is

$$d_{i,j+1} = d_{j,j+1} - d_{i,j} = \frac{1}{N} - d_{i,j}. \tag{8}$$

Thus, (7) simplifies to

$$\ln C_{0,i,j} + \delta \ln C_{1,i,j} - \gamma \ln d_{i,j} = \ln C_{0,i} + \delta \ln C_{1,i} - \gamma \ln \left(\frac{1}{N} - d_{i,j} \right). \tag{9}$$

Finally, as shown in Figure 1, the demand faced by firm j is equal to twice the number of individuals living between the firm and the indifferent individual, because there are individuals living to the left and to the right of the firm. Thus, solving (9) for $d_{i,j}$, we obtain the equilibrium demand function faced by firm j:

$$q_j = 2d_{i,j} = \frac{2}{N} \frac{\tilde{C}_j}{1 + \tilde{C}_j} \text{ where } \tilde{C}_j \equiv \left(\frac{C_{0,i,j}}{C_{0,i}} \right)^{\frac{1}{\gamma}} \cdot \left(\frac{C_{1,i,j}}{C_{1,i}} \right)^{\frac{\delta}{\gamma}}. \tag{10}$$

The optimal level of fees is found by deriving the first-order conditions of the firm's objective

[1] Salop (1979) also considers a "supercompetitive" equilibrium in which firm j's fees (f_j^v) are low enough to make it competitive for individuals living *beyond* firm $j+1$, and it captures all of its competitor's market. To simplify the analysis, we rule out this equilibrium by assuming that each individual only considers the two nearest firms.

Costs	Fees		
	Both fixed and variable	Fixed only ($f^v = 0$)	Variable only ($F = 0$)
Both fixed and variable	1. Main result: $f^v = c$. Too many firms due to business stealing.	2. $f^v < c$. Too many firms due to oversaving *and* business stealing.	3. $f^v > c$. Undersaving *vs.* business stealing. Uncertain result.
Fixed only ($c = 0$)	4. $f^v = c = 0$. Special case of #1.	5. $f^v = c = 0$. Special case of #1.	6. $f^v > c$. Special case of #3.
Variable only ($\phi = 0$)*			
Equilibrium fees ($\Pi > 0$)	7. $f^v = c$. Special case of #1.	8. $f^v < c$, special case of #2.	9. $f^v > c$. Special case of #3.
Zero profits ($\Pi = 0$)	10. $f^v = c$, $F = 0$.	11. $f^v < c$, F from eq. (20).	12. $f^v = c$. Special case of #1.

* As discussed in Section 1.2.4, the variable costs-only scenario ($\phi = 0$) features no barriers to entry and thus the size of the asset management industry is not well-defined. Moreover, even though the profits of each individual firm are infinitesimal, aggregate profits do not vanish. This implies that market equilibrium fees result in positive profits. Alternatively, enforcing zero profits results in a different level of fees. This fact does not affect any of our results, however.

Table 1: Model results under different assumptions for cost and fee structures.

function (4) with respect to F_j and f_j^v.

$$\frac{\partial \pi_j}{\partial f_j^v} = S_j \cdot q_j + \left[F_j + \left(f_j^v - c\right) S_j\right] \cdot \frac{\partial q_j}{\partial f_j^v} + \left(f_j^v - c\right) \cdot \frac{\partial S_j}{\partial f_j^v} \cdot q_j = 0. \tag{11}$$

$$\frac{\partial \pi_j}{\partial F_j} = q_j + \left[F_j + \left(f_j^v - c\right) S_j\right] \cdot \frac{\partial q_j}{\partial F_j} + \left(f_j^v - c\right) \frac{\partial S_j}{\partial F_j} \cdot q_j = 0. \tag{12}$$

To solve, we impose a symmetric equilibrium in which $F_j = F$ and $f_j^v = f^v$ (and $q_j = 1/N$). We examine one general case and three possible special cases covering all combinations of cost structure and fee structure. The results are summarized in Table 1.

1.2.1 General case: F, $f^v > 0$

This is the main case examined in the paper. Solving (11) and (12) yields

$$f^v = c, \tag{13}$$

$$F = \frac{2\gamma}{1+2\gamma+\delta}(1-c)(\omega+\Pi)\frac{1-\tau_L}{1-\tau_S}. \tag{14}$$

1.2.2 Only fixed fees ($F > 0$, $f^v = 0$)

In this case, (12) becomes the only optimality condition and simplifies to:

$$\frac{\partial \pi_j}{\partial F_j} = q_j + (F_j - cS_j) \cdot \frac{\partial q_j}{\partial F_j} - c\frac{\partial S_j}{\partial F_j} \cdot q_j = 0. \tag{15}$$

Solving, we obtain

$$F = \frac{2\gamma + c\delta\frac{1+\delta}{1+\delta-c}}{1+2\gamma+\delta}(\omega+\Pi)\frac{1-\tau_L}{1-\tau_S}. \tag{16}$$

1.2.3 Only variable fees ($F = 0$, $f^v > 0$)

In this case, (11) becomes the only optimality condition and simplifies to:

$$\frac{\partial \pi_j}{\partial f_j^v} = S_j \cdot q_j + \left(f_j^v - c\right) S_j \cdot \frac{\partial q_j}{\partial f_j} + \left(f_j^v - c\right) \cdot \frac{\partial S_j}{\partial f_j} \cdot q_j = 0. \tag{17}$$

Solving, we obtain

$$f^v = \frac{2\gamma + c\delta}{2\gamma+\delta}(> c). \tag{18}$$

1.2.4 Only variable costs ($\phi = 0$)

If $\phi = 0$, in all three cases discussed above, the equilibrium fee structure described by equations (11) and (12) would result in positive equilibrium profits in the absence of fixed costs. This is not problematic per se, but if instead one desires to enforce zero profits, the fee structure must be deduced from equation (11) together with the zero-profit condition, which yields

$$F = 0 \text{ and } f^v = c, \tag{19}$$

if variable fees are permitted, and otherwise

$$F = cS = \frac{\delta c}{1+\delta-c}\omega\frac{1-\tau_L}{1-\tau_S} \text{ and } f^v = 0. \tag{20}$$

1.3 Proof that $\partial\Pi/\partial N < 0$, $\partial S/\partial N < 0$ and $\partial F/\partial N < 0$ for given N

From the main text, we have the following equations given N:

$$F_j = F = \frac{\omega - \phi N}{\frac{1+2\gamma+\delta}{2\gamma} \cdot \frac{1}{1-c} \cdot \frac{1-\tau_S}{1-\tau_L} - 1}, \tag{21}$$

$$\Pi = \sum_j \pi_j = F - \phi N = \frac{\omega - \phi N}{1 - \frac{2\gamma}{1+2\gamma+\delta} (1-c) \frac{1-\tau_L}{1-\tau_S}} - \omega, \tag{22}$$

$$S = \frac{(2\gamma + \delta)(\omega - \phi N)}{(1 + 2\gamma + \delta) \frac{1-\tau_S}{1-\tau_L} - 2\gamma (1-c)}. \tag{23}$$

For Traditional, the proof is trivial. Eqs. (21)–(23) are of the form $a - bN$ because tax rates simplify out. For Roth, it is enough to substitute $\tau_L = G/(\omega + \Pi)$ and $\tau_S = 0$ in the same equations and rearrange.

- $\partial\Pi/\partial N < 0$: Eq. (22) simplifies to

$$\Pi = \frac{\omega - \phi N}{1 - \tilde{K}\left(1 - \frac{G}{\omega+\Pi}\right)} - \omega \text{ with } \tilde{K} \equiv \frac{2\gamma}{1 + 2\gamma + \delta}(1-c) \in (0,1) \tag{24}$$

  or

$$\Pi = \frac{(\omega - G)\tilde{K} - \phi N}{1 - \tilde{K}}, \tag{25}$$

  a decreasing function of N.

- $\partial S/\partial N < 0$: Eq. (23) simplifies to:

$$S = \frac{(2\gamma + \delta)(\omega - \phi N)}{(1 + 2\gamma + \delta)\frac{1}{1-\frac{G}{\omega+\Pi}} - 2\gamma(1-c)} \tag{26}$$

  As N grows, the numerator shrinks. Π also shrinks, so τ_L $(= G/(\omega + \Pi))$ grows, so $1 - \tau_L$ shrinks, so $1/(1 - \tau_L)$ grows, so the denominator grows. Overall, S shrinks.

- $\partial F/\partial N < 0$ follows trivially because F is a fixed proportion of S.

1.4 Welfare analysis

In this subsection we show that in our model $U^{Roth} \geq U^{Trad}$, both under the assumption of arbitrary N constant across Roth and Traditional, and under the assumption of free entry and endogenous N determined by market competition.

1.4.1 N given

With N given, notice that

$$U^{Roth} \geq U^{Trad} \iff e^{U^{Roth}}/e^{U^{Trad}} \geq 1 \tag{27}$$

Simplifying, we obtain

$$\frac{(1+\delta(1-c))(1-N\phi/\omega - G/\omega)}{(1+\delta(1-c))(1-N\phi/\omega) - G/\omega\,(1+\delta+2c\gamma)} \geq 1. \tag{28}$$

Since the left-hand side expression is the ratio of two exponentials, it must be positive. Since the numerator is clearly positive, it follows that the denominator must be positive too. Then, (28) simplifies to

$$1+\delta(1-c) \leq 1+\delta+2c\gamma, \tag{29}$$

which always holds, and holds with equality when $c = 0$.

1.4.2 Endogenous N

Given the optimal N, we have

$$U^{Roth} - U^{Trad} = (1+\gamma+\delta)\log(1 - G/\omega) - (1+\delta)\log\left(1 - G/\omega \cdot \frac{1+2\gamma+\delta}{1+\delta(1-c)}\right). \tag{30}$$

The argument of the logarithm in the second term can in principle be negative; we just assume it is positive. To understand why this assumption is innocuous, note that

$$G/\omega \cdot \frac{1+2\gamma+\delta}{1+\delta(1-c)} = \tau_L^{Trad}, \tag{31}$$

i.e., we have simply assumed that the tax rate under Traditional is less than 100%. If there are no frictions ($\gamma = 0$, $c = 0$) this simplifies to the trivial condition $G/\omega < 1$, i.e., public expenditure cannot be more than output. However, there are frictions in our model ($\gamma > 0$, $c \geq 0$), and $\tau_L^{Trad} < 1$ is a stricter condition than $G/\omega < 1$. If γ or c are large enough, under this policy there are so many firms in equilibrium that society cannot afford both asset management and public expenditure at current levels. The assumption that $\tau_L^{Trad} < 1$ simply rules out this pathological case.

Rearranging, we obtain

$$U^{Roth} - U^{Trad} > 0 \iff 1 - G/\omega > \left(1 - G/\omega \cdot \frac{1+2\gamma+\delta}{1+\delta(1-c)}\right)^{\frac{1+\delta}{1+\gamma+\delta}}. \tag{32}$$

To show that this inequality always holds, set $x \equiv \frac{1+\delta}{1+\gamma+\delta}$. Then,

$$1 - {}^{G}\!/\!_{\omega} > \left(1 - {}^{G}\!/\!_{\omega} \cdot \frac{1}{x}\right)^{x} > \left(1 - {}^{G}\!/\!_{\omega} \cdot \frac{1+2\gamma+\delta}{1+\delta\left(1-c\right)}\right)^{x} \tag{33}$$

where the first inequality is true for every $0 < {}^{G}\!/\!_{\omega} < x < 1$ and the second one is true because $x > \left(1+\delta\left(1-c\right)\right) / \left(1+2\gamma+\delta\right)$.

1.4.3 Alternative specifications of costs and fees

So far we have discussed welfare when firms face both fixed and variable costs and are free to set both fixed and variable fees (with zero variable costs a special case), and showed that firms choose to set variable fees equal to marginal costs ($f^v = c$). What if firms are instead restricted to charge either variable fees or fixed fees, but not both? Above in Section 1.2 we have solved the model under these restrictions and other possible combinations of fixed and variable costs and fixed and variable fees.

The business-stealing effect exists independent of fee and cost structure, and in each case it steers society in the model towards an overly large asset management sector ($N^{Roth} > N^{*}$). However, if there is a mismatch between the fee structure and the cost structure, another friction arises. For example, if there are variable costs ($c > 0$) but fees are restricted to be fixed-only ($f^v = 0 < c$), individuals do not internalize the cost of managing assets in their saving decision, so that the Euler equation is now different from the planner's ($1 - f^v \neq 1 - c$), i.e., the optimal C_1/C_0 under Roth is higher than what the planner would choose. The additional fixed fees F that firms charge to cover their variable costs imply that, compared to the baseline scenario with $f^v = c$, initial consumption is lower ($C^{Roth}_{0,f^v=0} < C^{Roth}_{0,f^v=c}$) and saving is larger ($S^{Roth}_{f^v=0} > S^{Roth}_{f^v=c}$), so that $N^{Roth}_{f^v=0} > N^{Roth}_{f^v=c}$ and $U^{Roth}_{f^v=0} < U^{Roth}_{f^v=c}$. A switch from Roth to Traditional in this scenario will, as in the baseline scenario, increase the equilibrium number of firms and variable costs cS, and cause a welfare loss in the model.

In contrast, if there are fixed costs ($\phi > 0$) but fees are restricted to be variable-only ($F = 0$), firms must set $f^v > c$ in order to cover their fixed costs. Individuals face marginal fees higher than the marginal cost of managing assets, so that the optimal C_1/C_0 under Roth is *lower* than what the planner would choose. The additional resources released by setting $F = 0$ imply that compared to the baseline scenario with $F > 0$, initial consumption is higher ($C^{Roth}_{0,F=0} > C^{Roth}_{0,F>0}$) and saving is less ($S^{Roth}_{F=0} < S^{Roth}_{F>0}$), starving the asset management sector of assets and offsetting the business-stealing effect, so that the net effect on the number of firms and welfare is unclear ($N^{Roth}_{F=0} \gtrless N^{Roth}_{F>0}$ and $U^{Roth}_{F=0} \gtrless U^{Roth}_{F>0}$). If the decrease in saving is large enough that $N^{Roth}_{F=0} < N^* < N^{Roth}_{F>0}$, Traditional in the model may yield higher welfare than Roth by increasing aggregate saving and thereby the number of firms. However, under reasonable calibrations, a switch from Traditional to Roth in the model is still welfare-enhancing.[2] The intuition is that Traditional does not create a bigger saving subsidy than Roth in our model, and therefore it does not solve the undersaving problem; it merely exacerbates the individuals' price insensitivity because it subsidizes fees while leaving the individual's consumption/saving tradeoff intact.

1.5 Optimal Taxation

1.5.1 Optimal taxation under Traditional

Can the government in the model exercise its one degree of freedom under Traditional to obtain the same outcomes as the planner, or at least better outcomes than under Roth? Here we show that the tax policy used in the paper ($\tau_R = \tau_L$) could be improved upon unless $c = 0$. However, the improvement is not enough in the model to make Traditional better than Roth.

In a market equilibrium under Traditional, maximized utility simplifies to

$$U = K + \ln(1 - \tau_L) + \delta \ln(1 - \tau_R) \tag{34}$$

for some constant K that does not depend on government policy. Thus the government chooses τ_L and τ_R to maximize (34) subject to its budget constraint

$$\tau_R = \frac{{}^{G}\!/_{(\omega+\Pi)}\,(1 + 2\gamma + \delta) - \tau_L}{\delta\,(1 - c)}, \tag{35}$$

which yields the following tax rate path:

$$\tau_L^* = {}^{G}\!/_{(\omega+\Pi)} \frac{1 + 2\gamma + \delta}{1 + \delta} + \frac{c}{1 + \delta} \cdot \delta, \tag{36}$$

[2] By "reasonable calibrations" we mean the following. Even assuming no variable costs ($c = 0$, the most favorable scenario for Traditional), maximized utility under Roth is higher than under Traditional unless γ is larger than a certain threshold. If, as in Section 5 of the paper, we assume $\delta = 0.55$ and ${}^{G}\!/_{\omega} = 0.2$, the threshold for γ is roughly 0.57, implying that about 40% of total resources in the economy are devoted to asset management. In general, for reasonable choices of ${}^{G}\!/_{\omega}$ (between 0.1 and 0.5), the threshold for γ is of the same order of magnitude of δ, implying that individuals in the model are willing to devote as many resources to asset management services as they do to retirement consumption itself. We find this unreasonable. This result is independent of the size of firm-level fixed costs ϕ.

and

$$\tau_R^* = \left({}^{G}\!/\!_{(\omega+\Pi)} \frac{1+2\gamma+\delta}{1+\delta} + \frac{c}{1+\delta} \right) \cdot \frac{1}{1-c}. \tag{37}$$

This path can be upward or downward sloping, depending on the specific values of $G/(\omega+\Pi)$ (or equivalently G/ω when $\Pi = 0$), δ, γ, and c. However, even with these constrained-optimal tax rates, utility in the model cannot be higher under Traditional than under Roth. To see this, recall the expressions for consumption and number of asset management firms. The planner chooses:

$$C_0^* = \frac{1}{1+\delta+\gamma} (\omega - G), \tag{38}$$

$$C_1^* = \frac{\delta}{1+\delta+\gamma} (\omega - G)(1-c)(1+r), \tag{39}$$

$$N^* = \frac{\gamma}{1+\delta+\gamma} (\omega - G)(1-c)\frac{1}{\phi}. \tag{40}$$

And the market equilibrium quantities are (assuming zero profits):

$$C_{0,i}^{Mkt} = \frac{1}{1+2\gamma+\delta} \omega (1-\tau_L), \tag{41}$$

$$C_{1,i}^{Mkt} = \frac{\delta}{1+2\gamma+\delta} \omega (1-\tau_L)(1-c)(1+r)(1+\tau_M), \tag{42}$$

$$N^{Mkt} = \frac{1}{\phi} \cdot \frac{2\gamma}{1+2\gamma+\delta} (1-c)\,\omega \frac{1-\tau_L}{1-\tau_S}. \tag{43}$$

Because under Traditional $\tau_S = \tau_L$, N^{Trad} *does not depend on tax rates at all,* and therefore $N^{Trad} > N^{Roth} > N^*$ regardless of τ_L^{Trad}. In turn, a higher N implies $\tau_L^{Trad} > \tau_L^{Roth} = G/\omega$ for the government budget constraint to be satisfied, which in turn implies fewer resources available for consumption. Moreover, since in equilibrium $f^v = c$, the intertemporal consumption choice in the model is not distorted under Roth. With fewer resources and no distortions to correct, even with the best possible tax rates, Traditional in the model cannot be as good as Roth.

1.5.2 Comparing Roth to Taxable; Optimal taxation with TTE accounts

In order to compare Taxable with Roth, we derive the optimal tax rate on investment income in our model. If this rate is zero (or negative), Roth in the model is optimal (or constrained-optimal).

Aggregate utility under a taxable (TTE) system is

$$U^{TTE} = \ln C_0^{TTE} + \delta \ln C_1^{TTE} + \gamma \ln N^{TTE} + \gamma (1 + \ln 2) \tag{44}$$

Using the results from the paper with a slight modification (using the aftertax return $r(1-\tau_I)$ instead of just r), and setting $\tau_S = 0$, we obtain

$$C_0 = \frac{1}{1+\delta+2\gamma} \omega (1-\tau_L), \tag{45}$$

$$C_1 = \frac{\delta}{1+\delta+2\gamma} \omega (1-\tau_L)(1-c)(1+r(1-\tau_I)), \tag{46}$$

$$N = \frac{2\gamma}{1+\delta+2\gamma} \frac{\omega}{\phi} (1-c)(1-\tau_L). \tag{47}$$

Substituting in and rearranging we obtain utility as a function of tax rates:

$$U^{TTE} = \tilde{K} + (1+\delta+\gamma) \ln (1-\tau_L) + \delta \ln (1 + r(1-\tau_I)), \tag{48}$$

where $\tilde{K}$ is a constant that does not depend on tax rates. Next, we use the government's budget constraint to pin down τ_L as a function of τ_I, and solve for the τ_I that maximizes welfare.

Taxes on returns are collected at time 1. To compute tax revenue, consider that the final account balance in retirement is $(S(1-f^v) - F)(1+r)$ or, succinctly, $S(1-f)(1+r)$, where

$$f \equiv f^v + F/S = \frac{2\gamma + c\delta}{\delta + 2\gamma} \tag{49}$$

and

$$S = \frac{\delta + 2\gamma}{1+\delta+2\gamma} \omega (1-\tau_L). \tag{50}$$

We assume fees are nondeductible, reflecting the current U.S. tax environment.[3] Then, the tax basis of the investment is $S(1-f)$ and the tax revenue is $S(1-f) r\tau_I$.

The government's intertemporal budget constraint is

[3] Prior to 2018, in the U.S., investment management fees and financial planning fees were deductible if they exceeded 2% of AGI.

$$G = \omega \tau_L + S(1-f)\frac{r}{1+r}\tau_I. \tag{51}$$

Note that

$$S(1-f) = \frac{\delta + 2\gamma}{1+\delta+2\gamma}\omega(1-\tau_L)\left(1 - \frac{2\gamma + c\delta}{\delta + 2\gamma}\right)$$

$$= \frac{\delta}{1+\delta+2\gamma}(1-c)\,\omega(1-\tau_L) \equiv s\cdot\omega(1-\tau_L), \tag{52}$$

where $s \in (0,1)$, a constant, is defined for notational convenience. Substituting this expression into the budget constraint, and rearranging, we obtain $1-\tau_L$ as a function of τ_I:

$$1 - \tau_L = \frac{1 - G/\omega}{1 - s\frac{r}{1+r}\tau_I}. \tag{53}$$

Then, the utility function becomes

$$U^{TTE} = \tilde{K}_2 - (1+\delta+\gamma)\ln\left(1 - s\frac{r}{1+r}\tau_I\right) + \delta\ln(1 + r(1-\tau_I)) \tag{54}$$

where $\tilde{K}_2$ is another constant that does not depend on tax rates. The first-order condition is:

$$\frac{\partial U^{TTE}}{\partial \tau_I} = -(1+\delta+\gamma)\frac{-s\frac{r}{1+r}}{1 - s\frac{r}{1+r}\tau_I} + \delta\frac{-r}{1+r(1-\tau_I)} = 0 \tag{55}$$

Simplifying, we obtain

$$\tau_I^* = -\frac{1}{1-c}\cdot\frac{\gamma + c(1+\delta+\gamma)}{1+\gamma}\cdot\frac{1+r}{r} < 0 \tag{56}$$

Thus, in this case, $\tau_I^* < 0$ which means Roth in the model is better than Taxable, and negative tax rates on investment returns in the model would be even more welfare-enhancing. The intuition is the following:

- First, note that with logarithmic utility, regardless of the after-tax rate of return $r(1-\tau_I)$, dollar saving S and number of firms N are the same.

- Second, a negative τ_I means higher τ_L to balance the budget. A higher τ_L means lower S and lower N, which is good, because the equilibrium with $\tau_I = 0$ results in too many firms. Thus, at $\tau_I = 0$, $\partial U^{TTE}/\partial \tau_I < 0$.

- Finally, as τ_I moves away from zero, the Euler equation gets more and more distorted. At some point the damage from the distortion balances out the benefit from fewer firms, and an optimum is reached.

The calibrated optimal τ_I^* is somewhere between -1% and -3% based on $\gamma = 0.005$ to 0.2, $r = 150\%$ and $c = 20\%$.

2 Extension: Accounting for the deductibility of fees

This section examines how, in the partial equilibrium model in Section 2 of the paper (holding constant both percentage fees f and tax rates), alternative arrangements regarding the deductibility of fees influence the incidence of the higher fees that occur under Traditional. We examine two related questions that arose often when explaining our argument. First, does Roth result in higher present-value government revenue in the model than Traditional simply because under Traditional fees are paid with pretax money (i.e., implicitly tax-deductible), whereas under Roth they are paid with after-tax money? Second, our argument assumes that fees are paid with money from within the retirement account ("account money"). Would our argument still hold if fees were paid with money from outside of the retirement account ("outside money")?

2.1 Background on deductibility of fees

Under the U.S. tax code, some retirement account investment costs paid using "outside money" (i.e., money from an ordinary taxable account) are or have been tax-deductible. For an employer paying the expenses of a plan, "ordinary and necessary" plan-related expenses are deductible business expenses under U.S.C. 26 §162. For an individual contributing to an IRA, prior to the 2017 US tax reform, fees were deductible as "miscellaneous itemized deductions" under U.S.C. 26 §212 relating to expenses for production of income (Dold and Levine, 2011).

At the time of writing, the tax code does not explicitly discuss the treatment of IRA and retirement plan fees, leaving the matter in the hands of the Internal Revenue Service (IRS). The IRS has not issued detailed public guidance, but it has repeatedly upheld in private that the payment of certain retirement account fees using outside money is not considered a contribution to the account (and therefore it is presumably a deductible expense under Sections 162 or 212, although the IRS has not pronounced itself on deductibility).

In recent taxpayer guidance (Private Letter Ruling 201104061), the IRS appears to make a distinction between account-level fees and asset-level fees (without using this terminology). Account-level fees, such as wrap fees and advisory fees, are considered akin to overhead expenses, and therefore payable with outside money; whereas asset-level fees, such as brokerage commissions, are "intrinsic to the value" of the assets and should be capitalized in the asset values, in essence requiring that they be paid with account money. For instance, if the wrap fee covers brokerage commissions but does not depend on the number of trades, it is considered an account-level fee and it becomes payable with outside money.

Importantly, the IRS has never made a distinction between Traditional and Roth IRAs with regard to fee deductibility. In practice, however, there is a very relevant distinction. Assume for simplicity that contribution limits are not binding, and that the tax rate at the time of contributions is the same as the tax rate on distributions ($\tau_L = \tau_R = \tau$). Under these assumptions, regardless of account type, one dollar of fees could be paid with outside money, and then deducted as an expense, for a total after-tax cost of $1 - \tau$. An individual paying for fees with money from a Traditional account should be indifferent about the source of the fee payment, because one dollar of Traditional account money is worth only $1 - \tau$ in after-tax terms. In contrast, an individual paying one dollar of fees from a Roth account would prefer to use outside money because the account money is already after-tax. Under these assumptions, paying fees with outside money is only beneficial for Roth investors.[4]

2.2 The sources of the revenue difference between Roth and Traditional

The above discussion of fee deductibility raises important questions. Would alternative assumptions about the deductibility of fees shift some or all of the burden to individuals (unlike in Table 2 of the paper, where all of the added fees under Traditional are borne by the government)? In particular, if the U.S. government were to switch to a Roth-only

[4] In reality, contributions to Traditional and Roth accounts are subject to limits. Since these limits are set at the same nominal amount (currently $6,000) for both accounts, the Traditional limit is more likely to be binding. If the limit is binding, the shadow cost of using Traditional account money to pay fees could be greater than the cost of using outside money, and therefore there are taxpayers for whom it is advantageous to pay Traditional fees with outside money regardless of deductibility considerations. Nonetheless, for taxpayers that currently use Traditional account money to pay fees, a switch to a system in which Roth is the only option and fees paid with outside money are deductible would create an incentive to use outside money.

system, and fees were made deductible again, would part or all of the additional revenue of Roth be offset by individuals' increased fee deductions?

To examine these questions, we extend our results from Section 2 of the paper, obtained under the assumption that the level of fees (f) is not affected by the amount of assets under management. We decompose the difference in tax revenue between Traditional and Roth into two components: (i) fee deductibility and (ii) the sheer existence of additional assets. Assuming that all labor, retirement and investment income is taxed at the same flat rate ($\tau_L = \tau_R = \tau_I = \tau$), a fraction $1 - \tau$ of the difference in tax revenue is due to the fact that fees are implicitly deductible under Traditional and nondeductible under Roth. However, the remainder (a fraction τ) is due to the fact that under Traditional there are more assets and more fees are paid.

To see this, consider an individual allocating \$1 of pretax income to retirement savings at time 0. Depending on the taxation scheme, tax may be deferred or not, and therefore the initial account balance may be 1 or $1 - \tau$. The account grows at a rate r, and the asset manager charges fees proportional to the account balance at a rate f. At time 0, if the taxation scheme is Traditional, the government takes a fraction τ of the account balance, and the remainder is paid out to the individual.

Panel (a) of Table 2 shows the different present value outcomes in terms of retirement wealth, fee revenue and tax revenue under each of four account types. Because of the absence of frictions other than fees, the present value of these three quantities must sum to the initial contribution, i.e., to one. The four account types are obtained by combining two taxation schemes (Roth or Traditional) and two deductibility rules (deductible or nondeductible fees, represented as "Ded" and "NDed"). The first two accounts are $\text{Roth}_{\text{NDed}}$ (or simply "Roth", since fees in standard Roth accounts are non-deductible) and Trad_{Ded} (or simply "Trad", since fees in standard Traditional accounts are effectively deductible). The other two accounts are hypothetical. A fee-deductible Roth (Roth_{Ded}) is a Roth account in which the individual is able to get a deduction for fees paid. We assume that the future value of these deductions is added to the individual's retirement wealth.[5] A fee-nondeductible Traditional ($\text{Trad}_{\text{NDed}}$) is a Traditional account in which the individual is taxed on the gross-of-fees balance $(1 + r)^T$, i.e., fees are explicitly made nondeductible.

[5]Alternative assumptions (e.g., that the value of the current-period deduction is immediately added to the account) yield the same qualitative result but with less-tractable expressions.

	Retirement wealth	Fee Revenue	Tax Revenue
a. Present value			
Roth	$(1-\tau)(1-f)^T$	$(1-\tau)\left(1-(1-f)^T\right)$	τ
Trad	$(1-f)^T(1-\tau)$	$1-(1-f)^T$	$\tau(1-f)^T$
Roth_{Ded}	$(1-\tau)\left[(1-f)^T(1-\tau)+\tau\right]$	$(1-\tau)\left(1-(1-f)^T\right)$	$\tau\left[\tau+(1-\tau)(1-f)^T\right]$
$\text{Trad}_{\text{NDed}}$	$(1-f)^T-\tau$	$1-(1-f)^T$	τ
b. Decomposition 1 (Practical)			
Roth_{Ded} – Roth	$\tau(1-\tau)\cdot\left(1-(1-f)^T\right)$	0	$-\tau(1-\tau)\left(1-(1-f)^T\right)$
Trad – Roth_{Ded}	$-\tau(1-\tau)\cdot\left(1-(1-f)^T\right)$	$\tau\left(1-(1-f)^T\right)$	$-\tau^2\left(1-(1-f)^T\right)$
Trad – Roth	0	$\tau\left(1-(1-f)^T\right)$	$-\tau\left(1-(1-f)^T\right)$
c. Decomposition 2 (Policy)			
$\text{Trad}_{\text{NDed}}$ – Roth	$-\tau\left(1-(1-f)^T\right)$	$\tau\left(1-(1-f)^T\right)$	0
Trad – $\text{Trad}_{\text{NDed}}$	$\tau\left(1-(1-f)^T\right)$	0	$-\tau\left(1-(1-f)^T\right)$
Trad – Roth	0	$\tau\left(1-(1-f)^T\right)$	$-\tau\left(1-(1-f)^T\right)$

Table 2: Present value of retirement wealth, fee revenue, and tax revenue under different combinations of taxation scheme and fee deductibility. Roth is an account with front-loaded taxation, such as a U.S. Roth IRA. Roth is an account with back-loaded taxation, such as a U.S. Traditional IRA. "Ded" and "NDed" indicate that fees are, respectively, deductible and nondeductible. Since in a Traditional account fees are implicitly deductible, there is no Trad_{Ded}, but just Trad. Similarly, since in a Roth account fees are implicitly nondeductible, there is no $\text{Roth}_{\text{NDed}}$, but just Roth.

Our goal is to understand the sources of present-value differences in fee revenue, retirement wealth, and tax revenue between account types. To do this, we calculate the difference between different account types with respect to these three quantities. The three differences must sum to zero.

Panel (b) of Table 2 decomposes the difference between Traditional and Roth based on the following identity:

$$\text{Trad} - \text{Roth} = \left(\text{Roth}_{\text{Ded}} - \text{Roth}\right) + \left(\text{Trad} - \text{Roth}_{\text{Ded}}\right).$$

The first term $(\text{Roth}_{\text{Ded}} - \text{Roth})$ is the effect of starting from a pure Roth and making fees deductible. Doing so increases retirement wealth at the expense of the government. The second term $(\text{Trad} - \text{Roth}_{\text{Ded}})$ is the effect of switching from a fee-deductible Roth to a Traditional. Doing so directly increases assets under management, and thus fees, this time at the expense of both the individual and the government.

This decomposition is of practical importance because, although a fee-deductible Roth account does not exist per se, as discussed above, individuals were able in the past to deduct some fees paid with outside money and might be able again in the future. A fee-deductible Roth represents the case in which individuals are able to deduct *all* fees. In this extreme case, a switch from Traditional to Roth fails to realize a fraction $(\text{Roth}_{\text{Ded}} - \text{Roth}) / (\text{Trad} - \text{Roth}) = 1 - \tau$ of the expected improvement in tax revenue. Thus, a fraction $1 - \tau$ of the expected revenue benefit of switching from Traditional to Roth is attributable to the implicit nondeductibility of fees under Roth, and the remainder (τ) is attributable to the lower total fee revenue, i.e., to the existence of additional assets under Traditional.

Panel (c) of Table 2 decomposes the difference between Traditional and Roth based on a different identity:

$$\text{Trad} - \text{Roth} = \left(\text{Trad}_{\text{NDed}} - \text{Roth}\right) + \left(\text{Trad} - \text{Trad}_{\text{NDed}}\right).$$

The first term $(\text{Trad}_{\text{NDed}} - \text{Roth})$ is the effect of starting from Roth and switching to a fee-nondeductible Traditional. Doing so leaves tax revenue intact, but it still increases fee revenue by an amount $\tau\left(1 - (1-f)^T\right)$ at the expense of retirement wealth. The second term $(\text{Trad} - \text{Trad}_{\text{NDed}})$ is the effect of going from a fee-nondeductible Traditional to a pure Traditional. Fee revenue is unvaried, but the burden of the excess fees is transferred from the government to the individual.

This decomposition is of policy importance because it shows that making the individual fully responsible for fees obviously solves the government's revenue problem, but only by shifting the burden of fees onto individuals. Thus, if fees are deductible, individuals are indifferent and the government prefers Roth; if fees are nondeductible, the government is indifferent and individuals prefers Roth.

2.3 Practical importance

Above we showed that a switch to Roth combined with a deduction allowance for Roth fees may cause individuals to start paying fees with money from outside the account. In this scenario, individuals would capture a fraction $1 - \tau$ of the total fee savings from the switch. Assuming $\tau = 25\%$, this would amount to a three-quarters reduction in the expected government savings. In reality, however, the extent to which individuals would be able to take advantage of this opportunity is limited by several factors.

For IRA owners, even when fees were deductible, there were significant limits to the deductibility of expenses incurred in the production of income under U.S.C. 26 §212. First, miscellaneous itemized deductions were subject to a floor of 2% of adjusted gross income. Second, they were a preference item for purposes of the alternative minimum tax. Third, they were itemized deductions and therefore worthless for taxpayers taking the standard deduction (Dold and Levine, 2011). Taken together, these restrictions would have prevented a large number of IRA owners from taking deductions for Roth fees.

For participants in employer-sponsored retirement plans, there is no obvious way to cover investment costs with outside money. Some employers already cover some of the costs, but it is not obvious that upon a switch to Roth employers would have an incentive to do more than they already do. According to a Deloitte study (Rosshirt et al., 2014), employers cover roughly one-tenth of account costs, or 6 bps—a small fraction of the total.

Finally, our estimated investment costs largely consists of trading costs and expense ratios of mutual funds and similar investment product, two types of expenses for which using outside money is usually impractical or impossible. IRA owners might be able to cover their explicit trading costs and advisory services with a wrap fee, which could be paid with outside money, but they would still be have to use account money for implicit trading costs such as bid-ask spreads, as well as for mutual fund expense ratios. Retirement plan participants are largely invested in mutual funds and other collective investment products, and many employers would find it difficult to cover more costs than they already do.

3 Further detail on investment fees

3.1 Additional detail on fees and costs

3.1.1 Distribution fees and complimentary advice provided with the account

Often, asset management accounts come with some level of complimentary advice. Part of the cost of these services is financed by "distribution fees" that are charged not by the account provider, but rather by the managers of the products available in the account. These fees are then rebated to the account provider in the same way that salespeople receive a commission from the manufacturer of the products they sell. For instance, in the case of mutual funds, there are "load" fees, e.g., one-time fees paid upon purchase or redemption of shares, as well as ongoing fees called variously "level load fees", "service fees", or "12b-1" fees that are included in the ongoing expense ratio.[6] 12b-1 fees cover two types of expense: distribution costs, i.e., commissions to the sales force (capped at 75 bps), and shareholder servicing costs, e.g., cost of providing internet access to fund filings, etc. (capped at 25 bps). All these fees are distribution fees and are rebated to the account provider.

For instance, with a 5% front load, an investor giving $100 to the broker is only investing $95. If the fund has 12b-1 fees in addition to loads, these fees will be levied every year upon the $95. The same fund may have multiple classes of shares. According to Morningstar's Glossary, "In a typical multi-class situation, the class A fund has a front-end load and either a 0.25% distribution fee or a 0.25% service fee. Class B shares usually have a contingent deferred sales charge and a corresponding 0.75% 12b-1 fee, plus a maximum 0.25% service fee. [...] Class C shares customarily charge a level load with the same fee structure found in a class B share."

An account provider may derive revenue from explicit account fees, ongoing distribution fees like 12b-1 fees, and one-time distribution fees like loads. In the presence of explicit account fees, investors typically have access to "no-load" funds, although no-load funds can

[6] 12b-1 fees are so called after SEC Rule 12b-1 under the Investment Company Act of 1940. FINRA regulations from 1993 establish the caps on these fees. See SEC > Mutual Funds Fees and Expenses (https://www.sec.gov/answers/mffees.htm).

still charge up to 0.25% of level load fees. Overall, advisory and distribution fees (excluding 12b-1 fees, which are already included in the net expense ratio) average about 50 bps (Bogle, 2014).

3.1.2 Advisory fees charged by mutual fund managers

A mutual fund's net expense ratio includes three types of costs. First, paperwork costs: custodial fees, legal fees, record-keeping fees, etc. These fees typically cover the cost of inevitable services provided by third parties unaffiliated with the mutual fund. Second, distribution and service fees, discussed above. Third, asset management advisory fees, i.e., the actual revenue of the money management company that sponsors the fund in the first place.

Typically, advisory fees are not set at arm's length because the fund is a captive customer of the management company. This unusual price determination process has been very controversial. Proponents of the status quo argue that market forces curb excessive advisory fees, because of the threat of investors withdrawing their money and taking it to a different fund (e.g., Coates and Hubbard, 2007). However, others contend that market forces are not sufficient to keep fees in check because no fund's fees are set at arm's length; even if a fund's fees appear "reasonable" with respect to the competition, they need not be reasonable overall (Freeman et al., 2008). The Supreme Court (*Jones et al. v. Harris Associates L.P.*, 2010) rejects the "market" argument, in part because it is conscious of the lack of arm's length prices, arguing instead in favor of the "workable standard" set in the *Gartenberg* case, i.e., that in order for high fees to be evidence of breach of fiduciary duty, they must be so disproportionately high that they bear no resemblance to the services provided and could not be the result of arm's length bargaining. Evidence of breach of fiduciary duty must otherwise be found in the process by which the mutual fund board has reviewed the advisor's fees.

3.1.3 Trading costs

Trading costs include explicit commissions and implicit costs like bid-ask spreads and market impact. Quantifying trading costs is challenging. We are not aware of any peer-reviewed, asset-weighted estimates of trading costs for U.S. equity mutual funds, and of any estimates (asset-weighted or equal-weighted) for bond funds. Equal-weighted estimates are useful to discuss the average *fund*, but the government is interested in the average *dollar* invested in a fund, because a fraction of each dollar will eventually generate tax revenue. Often, equal-weighted estimates are driven by many inefficient small funds, whereas most of the dollars are in a few efficient, large funds. Here we report some equal-weighted estimates for completeness' sake.

Livingston and Zhou (2015) estimates that equal-weighted average explicit portfolio commissions alone are in the order of 18 bps. Wealthfront (2016) finds a very similar number (20 bps). The literature on implicit trading costs reports a wide range of estimates, perhaps because of the difficulty of quantifying these costs. Wermers (2000) estimates that commissions, transaction costs and cash drag due to liquidity cause a 230 bps wedge between the average equity mutual fund's returns and the return of the stocks they hold. Edelen et al. (2013) estimate average total trading costs of 144 bps using a sample of over 3,000 U.S. domestic equity funds. In this sample, implicit costs exceed the average expense ratio (119 bps).

3.2 Summary of the literature on performance of actively managed mutual funds

Measuring mutual fund performance is difficult. First, actual performance net of the benchmark has a large random component, and a reliable estimate of performance requires a long time series. Second, unlike direct estimates of fees, every benchmark-based estimate implies and depends on an asset-pricing model. As a result, the literature on mutual fund performance contains numerous estimates done using different methodologies and benchmarks, a few of which are summarized Table 3. Some of these are in the main paper, and the remainder are described next.

The literature begins with classics such as Fama (1965) and Jensen (1968). Both studies show no evidence of managers predictably beating the market on a net-of-fee basis; on average, mutual funds show a small underperformance with respect to the market benchmark. Consistent with market efficiency, this underperformance is of the same magnitude of fees and cash drag. More recently, Carhart (1997) compiles a mutual fund database that is comprehensive and free of survivorship bias, and uses it to replicate the basic result that there

	Performance (bps)		
Source	Net	Gross	Benchmark
Berk and van Binsbergen (2015)	-12	*	Investable Vanguard funds
Malkiel (2013)	-64		Large cap active vs. SP500 Index
Malkiel (2013)	-82		Bond funds vs. Barclays US Agg
Fama and French (2010)	-100	-5	3- and 4-factor benchmarks
Wermers (2000)	-100	130	Own stock holdings
Carhart (1997)	-102	**	1-, 3- and 4-factor benchmarks
Jensen (1968)	-40	~0	1-factor benchmark (CAPM)
Fama (1965)	-60	+20	Market

Table 3: **Estimates of average equity mutual fund underperformance.** "Net" and "Gross" refers to expenses. The definition of "expenses" is typically the expense ratio, but in the case of Wermers (2000) it includes everything including cash drag and trading costs (see text) — Footnotes:
[*] Underperformance with respect to the Vanguard benchmark, which charges fees of 18 bps
[**] 100 bps of expense ratio are associated with underperformance of 154 bps. Using our asset-weighted estimate of 66bps, $102 = 154 \times 66/100$.

is no evidence of skilled or informed mutual fund managers.[7] Using four-factor and three-factor benchmarks, Carhart finds that there is manager-specific persistence in performance that is not explained by fees, but only for the worst-performing funds. He estimates that 100 bps of expense ratio are associated with a 154 bps underperformance with respect to the market. Wermers (2000) decomposes mutual fund returns into stock-picking talent, style, transaction costs, and expenses, concluding that mutual funds hold stocks that beat the market by 1.3%, but the funds' returns underperform the market by 1%. He attributes the large discrepancy to cash drag (0.7%) and expenses and transaction costs (1.6%).

Based on a CAPM benchmark, Fama and French (2010) estimate net-of-fees underperformance of about 1% per year. Malkiel (2013) compares several categories of funds with their indices, finding that active large-cap equity funds underperform the S&P 500 Index by 64 bps, and bond funds underperform the Barclay US Aggregate Bond Index by about 84 bps.

[7] Malkiel (1995) also addresses survivorship bias and extends the sample period of previous studies which claimed to find persistence in returns. Carhart also addresses those studies, explaining their findings as the result of momentum investing.

Country	Data Year	Retirement Assets $b	Retirement Assets % Deferred	Gov. Acct. τ_R	Gov. Acct. $b	Fees	τ_C	Subsidy $b	Subsidy % GDP
United States	2018	16,464	94%	20%	3,084	0.80%	21%	19.5	0.10%
Canada	2015	1,003	86%	15%	129	2.06%	15%	2.3	0.15%
United Kingdom	2015	950	32%	20%	41	1.45%	20%	0.7	0.02%
Netherlands	2015	108	100%	39%	41	1.41%	25%	0.4	0.06%
Switzerland	2015	945	100%	4.0%	38	1.29%	18%	0.4	0.06%
Australia	2015	1,797	55%	3.4%	34	1.10%	30%	0.3	0.02%
Japan	2015	112	100%	2.6%	3	1.47%	30%	0.0	0.00%

Table 4: **Estimated subsidy to the asset management industry in seven countries with the largest Traditional retirement assets.** Fees are the asset-weighted average of money market, equity and fixed-income mutual fund fees based on overall (not retirement-only) asset allocation in that country. For each country, τ_R (the tax rate on retirement income, and therefore the fraction of Traditional assets that implicitly belong to the government) is calculated as the average tax rate faced by a person earning the average retirement income with no other income. τ_C, the corporate tax rate, is simply the top statutory tax rate. Sources: see text.

Some recent studies have focused on investable benchmarks. French (2008) estimates a broad measure of the annual cost of active management, including not only costs faced by individual investors but also costs faced by institutions and market-making gains by financial intermediaries over 1980-2006. The cost of active management is 0.67% of the aggregate value of the market, in addition to the approximately 0.10% cost of passive management. As a passive benchmark, French uses the Vanguard Total Stock Market Index. Berk and van Binsbergen (2015) compare active funds' dollar returns (as opposed to percent returns) against the relevant Vanguard benchmarks. They estimate a value weighted net alpha of -12 bps (not statistically different from zero) in addition to the fees on the Vanguard benchmark (18 bps), implying a total cost of active money management of at least 30 bps, not including the benchmark's implicit trading costs, and any account-level fees for recordkeeping and advice.

4 International fee calibration

In Table 4, we carry out a back-of-the envelope calibration of the annual subsidy to asset managers for the seven countries with the largest dollar amounts of tax-deferred assets (the rows are ordered by the size in dollars of the implicit government account). As in the main text, the subsidy is calculated as

$$\text{Annual subsidy} = S \cdot \tau_R \cdot f \cdot (1 - \tau_C) . \qquad (57)$$

We use data on all existing types of tax-advantaged retirement plans and their tax treatment from the Organization for Economic Cooperation and Development (OECD, 2015a,b), estimates of their magnitudes from various sources (2015-2018), average retirement income from each country's statistical office, information on basic deductions, personal tax brackets, and the corporate tax rate from each country's tax authority, and fee estimates from Morningstar (Alpert et al., 2013) and other sources.[8]

For consistency with our U.S. estimates, we exclude defined benefit (DB) pension plans from the calculation. With or without DB plans, the U.S. has the world's largest retirement assets, and therefore leads the list. However, other countries have substantial amounts of DB retirement assets (United Kingdom, Netherlands and Japan), and omitting DB leads to an important underestimate of the size of the implicit government account. In the case of United Kingdom and Netherlands, this underestimate meaningfully affects the estimated subsidy.

Each of the components of the subsidy has substantial variation across countries. For instance, although Switzerland, Australia and Japan have significant tax-deferred assets, the estimated subsidy is small simply because under current tax law retirement payouts are lightly taxed. Canada has the second-largest subsidy in dollar terms ($2.3 billion) and the largest as a fraction of GDP (0.15%), driven by the surprisingly large fees charged by Canadian funds (2.06%).

[8] Our estimate of τ_R is a rough lower bound, equal to the average tax rate faced by a person earning the average retirement income with no other income.

References

Alpert, B.N., Rekenthaler, J., Suh, S., 2013. Global Fund Investor Experience 2013 Report. Technical Report. Morningstar Fund Research.

Berk, J., van Binsbergen, J., 2015. Measuring skill in the mutual fund industry. Journal of Financial Economics .

Bogle, J.C., 2014. The arithmetic of 'all-in' investment expenses. Financial Analysts Journal 70, 13–21.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Coates, J.C., Hubbard, R.G., 2007. Competition in the mutual fund industry: evidence and implications for policy. Harvard John M. Olin Discussion Paper No. 592, available at `http://ssrn.com/abstract=1005426`.

Dold, E.T., Levine, D.N., 2011. Tax treatment on the payment of IRA fees gets a new but familiar private letter ruling. Taxes—The Tax Magazine , 17–20.

Edelen, R., Evans, R., Kadlec, G., 2013. Shedding light on "invisible" costs: Trading costs and mutual fund performance. Financial Analysts Journal 69, 33–44.

Fama, E.F., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.

Fama, E.F., French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. The Journal of Finance 65, 1915–1947.

Freeman, J.P., Brown, S.L., Pomerantz, S., 2008. Mutual fund advisory fees: new evidence and a fair fiduciary duty test. Oklahoma Law Review 61, 83–153.

French, K.R., 2008. Presidential address: The cost of investing. The Journal of Finance 53, 1537–1573.

Jensen, M.C., 1968. Problems in selection of security portfolios. The Journal of Finance 23, 389–416.

Landoni, M., Zeldes, S., 2020. Should the government be paying investment fees on $3 trillion of tax-deferred retirement assets? Working Paper.

Livingston, M., Zhou, L., 2015. Brokerage commissions and mutual fund performance. Journal of Financial Research 38, 283–303.

Malkiel, B.G., 1995. Returns from investing in equity mutual funds 1971 to 1991. Journal of Finance 50, 549–572.

Malkiel, B.G., 2013. Asset management fees and the growth of finance. Journal of Economic Perspectives 27, 97–108.

OECD, 2015a. Stocktaking of the tax treatment of funded private pension plans in OECD and EU countries.

OECD, 2015b. The tax treatment of funded private pension plans - OECD and EU country profiles.

Rosshirt, D.E., Parker, S.A., Pitts, D.A., 2014. Inside the Structure of Defined Contribution/401(k) Plan Fees, 2013: A study assessing the mechanics of the 'all-in' fee. Technical Report. Deloitte Consulting LLP.

Salop, S.C., 1979. Monopolistic competition with outside goods. Bell Journal of Economics 10, 141–156.

Tirole, J., 1988. The Theory of Industrial Organization. MIT Press.

Wealthfront, 2016. Mutual fund fees. URL: `\url{https://www.wealthfront.com/research/mutual-fund-fees}`. online document. Visited on 11/6/2016.

Wermers, R., 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses. Journal of Finance 55, 1655–1703.