Online Appendix for "A Simpler Theory of Optimal Capital Taxation" by Emmanuel Saez and Stefanie Stantcheva

A.1 Anticipated Reforms – Additional Results

Optimal tax with anticipated reform and heterogeneous discount rates

With heterogeneous discount rates, the change in social welfare is:

$$dSWF \propto \int_i \omega_i e^{-\delta_i T} \cdot \Big[u_{ic}(c_i, k_i, z_i) - u_{ic}(c_i, k_i, z_i) \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \int_0^\infty e^a_K(t) \cdot e^{-\delta_i (t-T)} dt$$
$$- \frac{\tau_L}{1-\tau_K} \delta_i u_{ic}(c_i, k_i, z_i) \frac{z^m}{rk^m} \int_0^\infty e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i (t-T)} dt \Big]$$

Define the normalized social welfare weight $g_i(T) \equiv \frac{\omega_i u_{ci} e^{-\delta_i T}}{\int_i \omega_i u_{ci} e^{-\delta_i T}}$ which depends on the time of the reform and rewrite:

$$dSWF \propto 1 - \int_i g_i(T) \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K} \int_i \delta_i g_i(T) \int_0^\infty e^a_K(t) \cdot e^{-\delta_i (t-T)} dt$$
$$- \frac{\tau_L}{1-\tau_K} \frac{z^m}{rk^m} \int_i \delta_i g_i(T) \int_0^\infty e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i (t-T)} dt \Big]$$

which yields the same optimal tax formula as in the text, but with

$$\bar{g}_K = \int_i g_i(T) \cdot k_i / k^m, \quad e^a_K = \int_i \delta_i g_i(T) \int_0^\infty e^a_K(t) \cdot e^{-\delta_i (t-T)} dt$$
$$\text{and} \quad e^a_{L,1-\tau_K} = \int_i \delta_i g_i(T) \int_0^\infty e^a_{L,1-\tau_K}(t) \cdot e^{-\delta_i (t-T)} dt$$

Relative to the case with homogeneous discount rates, the social welfare weights $g_i(T)$ now depend on the time of the reform T, and on each person's discount rate δ_i. Agents who are more impatient (larger δ_i) get discounted more heavily in the anticipated social welfare objective. This effect is starker the longer in advance the reform is announced (the larger T). In the limit, only the most patient agents with the lowest δ_i will be counted in the social welfare objective.

Finite and infinite anticipation elasticities

We now show that anticipation elasticities are infinite under full certainty with wealth in the utility and certainty. We also show that they are finite with uncertainty (with or without wealth in the utility).

First, the anticipation elasticity to a reform $d\tau_K$ for $t \geq T$ is infinite when there is full certainty, even with wealth in the utility. The proof is as in Piketty and Saez (2013) for the Chamley-Judd model (without wealth in the utility).

To see this, note that under full certainty the first-order condition of the agent with respect to capital always holds:

$$u_{ci,t} = (1+\bar{r})/(1+\delta_i)u_{ci,t+1} + 1/(1+\delta_i)u_{ki,t+1}$$

Suppose we start from a situation in a well-defined steady state: $(\delta_i - \bar{r})u_{ci} = u_{ki}$ where we have perfect consumption smoothing.

The intertemporal budget constraint is:

$$\sum_{t\geq 0}\left(\frac{1}{1+r}\right)^t c_{ti} + \lim_{t\to\infty} k_{ti} = \sum_{t\geq 0}\left(\frac{1}{1+r}\right)^t z_{ti} + k_{0i}$$

Consumption smoothing implies:

$$u_{ci}(\bar{r}k_i + z_i, k_i) = \lambda$$

for the multiplier λ on the budget constraint. Hence, $k_i^\infty = \lim_{t\to\infty} k_{ti} > 0$. Given that there is perfect consumption smoothing, using the budget constraint to solve for consumption yields:

$$c = \left(1 - \frac{1}{1+r}\right)\left(\sum_{t\geq 0}\left(\frac{1}{1+r}\right)^t z_{ti} + k_{0i} - k_i^\infty\right) \tag{A1}$$

Consider what happens if the capital tax rate increases by $d\tau_K > 0$ for $t \geq T$. The present discounted value of all resources, denoted by Y_i for agent i is:

$$Y_i = k_{i0} + \sum_{t=1}^{T}\left(\frac{1}{1+r}\right)^t z_{ti} + \sum_{t\geq T}\left(\frac{1}{1+\bar{r}}\right)^t z_{ti}$$

The change in resources evaluated at $\tau_K = 0$ is:

$$dY_i = \left(\frac{1}{(1+r)}\right)^T \sum_{t\geq T} t\left(\frac{1}{(1+r)}\right)^{t-T+1} z_{ti}d\tau_K \propto \left(\frac{1}{(1+r)}\right)^T d\tau_K$$

Hence, consumption pre-reform will shift down by a factor proportional to $\left(\frac{1}{(1+r)}\right)^T d\tau_K$. From the aggregated budget constraint we have that:

$$k_t^m = (1+r)^t k_0^m - c_0^m(1 + (1+r) + (1+r)^2 + ... + (1+r)^{t-1}) + (z_{t-1}^m + .. + (1+r)^{t-1} z_0^m)$$

Therefore, the change in the aggregate capital stock is:

$$dk_t^m = -dc_0^m \left(\frac{(1+r)^{t-1} - 1}{r}\right)$$

Recall that the change in consumption (from (A1)) is proportional to $\left(\frac{1}{(1+r)}\right)^T d\tau_K$. Hence:

$$dk_t^m \propto -\left(\frac{1}{(1+r)}\right)^T \left(\frac{(1+r)^{t-1} - 1}{r}\right) d\tau_K = -(1+r)^{-T} \left(\frac{(1+r)^{t-1} - 1}{r}\right) d\tau_K$$

and:

$$e_{Kt} \propto k_t^m (1+r)^{-T} \left(\frac{(1+r)^{t-1} - 1}{r}\right) d\tau_K$$

The anticipation elasticity e_K^{ante} is defined as:

$$e_K^{ante} = \frac{\delta}{1+\delta} \sum_{t<T} \left(\frac{1}{1+\delta}\right)^{t-T} e_{Kt} \propto \frac{\delta}{1+\delta} \sum_{t<T} \left(\frac{1}{1+\delta}\right)^{t-T} k_t^m (1+r)^{-T} \left(\frac{(1+r)^{t-1} - 1}{r}\right) d\tau_K$$

Since we have $\delta > r$, $\lim_{T\to\infty} \left(\frac{1+\delta}{1+r}\right)^T = \infty$, which makes the sum above (to which the anticipation elasticity is proportional) converge to infinity when T goes to infinity.

Optimal tax formula starting away from the steady state:

If the economy is not in steady state, the spirit of the formula still holds, but it is no longer possible to treat marginal utilities and aggregate variables as constant. In that case, not just the elasticities, but also the weighting factors multiplying them and the distributional characteristic take into account the full transition path. For conciseness, let $u_{ci}(t) \equiv u_{ic}(c_i(t), k_i(t), z_i(t))$. The change in social welfare is (with each term of the formula marked in underbraces):

$$dSWF \propto 1 - \underbrace{\int_i \omega_i \delta_i \int_T^\infty \frac{u_{ic}(t)}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) rk^m(t) \cdot e^{-\delta_i t} dt} rk_i(t) \cdot e^{-\delta_i t} dt}_{\bar{g}_K}$$

$$-\frac{\tau_L}{1-\tau_K} \underbrace{\int_i \omega_i \delta_i \int_0^\infty u_{ic}(t) e^a_{L,1-\tau_K}(t) \frac{z^m(t)}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) z^m(t) \cdot e^{-\delta_i t} dt} e^{-\delta_i t} dt}_{e^a_{L,(1-\tau_K)}} \cdot \underbrace{\frac{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) z^m(t) \cdot e^{-\delta_i t} dt}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) rk^m(t) \cdot e^{-\delta_i t} dt}}_{z^m/(rk^m)}$$

$$-\frac{\tau_K}{1-\tau_K} \underbrace{\int_i \omega_i \delta_i \int_0^\infty \frac{u_{ic}(t)}{\int_i \omega_i \delta_i \int_T^\infty u_{ic}(t) rk^m(t) \cdot e^{-\delta_i t} dt} rk^m(t) e^a_K(t) \cdot e^{-\delta_i t} dt}_{e^a_K}$$

For an unanticipated reform starting from an arbitrary tax system and away from the steady state, set $T = 0$ in the above expression for $dSWF$ and replace the anticipated elasticities by their unanticipated counterparts $e^u_{L,(1-\tau_K)}$ and e^u_K.

A.2 Comparison to Other Models

This section formally compares our model to earlier model described in Section 5.

A.2.1 Judd (1985) Model

In the Judd (1985) model, individual utility is:

$$V_i(\{c_i(t), z_i(t), k_i(t)\}_{t \geq 0}) = \int_0^\infty u_i(c_i(t), k_i(t), z_i(t)) e^{-\int_0^t \delta_i(c_i(s)) ds} dt.$$

The effect on V_i from a small change in the capital tax $d\tau_K$ is now:

$$dV_i = d\tau_K \left[\int_{t=0}^\infty \left(u_{ic}(c_i(t), k_i(t), z_i(t)) e^{-\int_0^t \delta_i(c_i(s)) ds} + \delta_i'(c_i(t)) \int_t^\infty u_i(s) e^{-\int_0^s \delta_i(c_i(m)) dm} ds \right) \right.$$
$$\left. \times \left(rk^m(t) - rk_i(t) - \frac{\tau_K}{1-\tau_K} rk^m(t) e_K(t) \right) dt \right].$$

In the steady state, we can hence write dV_i as:

$$d\tau_K r\left[\int_0^\infty \left(u_{ic}e^{-\delta_i(c_i)t} + \delta_i'(c_i)u_i e^{-\delta_i(c_i)t}\int_t^\infty e^{-\delta_i(c_i)(s-t)}ds\right)\left(k^m(t) - k_i(t) - \frac{\tau_K}{1-\tau_K}k^m(t)e_K(t)\right)dt\right]$$

$$= d\tau_K r\left[\left(u_{ic}\int_0^\infty e^{-\delta_i(c_i)t}dt + \delta_i'(c_i)u_i\int_0^\infty e^{-\delta_i(c_i)t}\frac{1}{\delta_i(c_i)}\right)\times[k^m(t) - k_i(t)]\right.$$

$$\left. - \int_0^\infty \left(u_{ic}e^{-\delta_i(c_i)t} + \delta_i'(c_i)u_i e^{-\delta_i(c_i)t}\int_0^\infty e^{-\delta_i(c_i)t}ds\right)\frac{\tau_K}{1-\tau_K}k^m(t)e_K(t)\right]$$

$$= d\tau_K r k^m \frac{1}{\delta_i(c_i)}\left(u_{ic} + \frac{\delta_i'(c_i)}{\delta_i(c_i)}u_i\right)\left[1 - \frac{k_i}{k^m} - \frac{\tau_K}{1-\tau_K}\delta_i(c_i)\int_0^\infty e^{-\delta_i(c_i)t}e_K(t)\right].$$

We can hence see that the formulas from our model apply but with g_i and e_K as redefined in the text.

A.2.2 Aiyagari (1995) Model

Note that all proofs below would be exactly the same as the proofs for wealth-in-the-utility if we reformulated it in discrete time, replacing the standard utility without wealth in the utility, $u_{ti}(c_{ti})$, by $u_{ti}(c_{ti}, k_{ti})$. This is done by letting u_{ti}' denote $\frac{\partial u_{ti}(c_{ti},k_{ti})}{\partial c_{ti}}$ instead of $\frac{\partial u_{ti}(c_{ti})}{\partial c_{ti}}$.

We apply the envelope theorem, which states that the changes in the capital tax rate $d\tau_K$ only has a direct impact on utility through the direct reduction in consumption that it causes. Using this, and taking the derivative of the social welfare SWF with respect to $d\tau_K$ yields:

$$dSWF = \sum_{t<T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u_{ti}' \cdot (\tau_K r dk_t^m) + \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u_{ti}' \cdot (r d\tau_K(k_t^m - k_{ti}) + \tau_K r dk_t^m)$$

$$= -d\tau_K\left(\frac{\tau_K}{1-\tau_K}\left[\sum_{t<T}\left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt}\int_i \omega_i u_{ti}' + \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt}\int_i \omega_i u_{ti}'\right]\right.$$

$$\left. + \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u_{ti}' \cdot r(k_t^m - k_{ti})\right)$$

$$= -d\tau_K\left(\frac{\tau_K}{1-\tau_K}\left[\sum_{t\geq 0}\left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt}\int_i \omega_i u_{ti}'\right] - \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u_{ti}' \cdot r(k_t^m - k_{ti})\right).$$

If variables have already converged to their ergodic paths when the anticipation responses start: then all terms in e_{Kt} are zero before the steady state has been reached and hence, we can divide through by $\int_i \omega_i u_{ti}' k_t^m = \int_i g_i k_t^m$ which is constant across t. Thus:

$$dSWF \propto \frac{\tau_K}{(1-\tau_K)}\left(\frac{\delta}{1+\delta}\sum_{t<T}\left(\frac{1}{1+\delta}\right)^{t-T} e_{Kt} + \frac{\delta}{1+\delta}\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^{t-T} e_{Kt}\right) - 1 + \frac{\int_i g_i k_{ti}}{\int_i g_i k_t^m}.$$

Let the distributional factor $\bar{g}_K = \frac{\int_i g_i k_{ti}}{\int_i g_i k_t^m}$. The optimal capital tax in the Aiyagari (1995) model is given by:

$$\tau_K = \frac{1-\bar{g}_K}{1-\bar{g}_K + e_K}.$$

with $e_K = \frac{\delta}{1+\delta}\sum_{t>0}\left(\frac{1}{1+\delta}\right)^{t-T} e_{Kt}$. For an unanticipated reform, the formula applies with $T=0$ when the economy is already in the steady state as of time 0.

If variables have not converged to their ergodic paths when the anticipation responses start: we have to take into account the transition of the marginal utilities and the capital stock across time.

$$dSWF = -d\tau_K\left(\frac{\tau_K}{(1-\tau_K)}\left[\sum_{t<T}\left(\frac{1}{1+\delta}\right)^t r k_t^m e_{Kt}\int_i \omega_i u'_{ti}\right] - \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti}\cdot r(k_t^m - k_{ti})\right).$$

Dividing by $\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \int_i \omega_i u'_{ti}\cdot k_t^m$ yields:

$$\begin{aligned} dSWF \propto \frac{\tau_K}{(1-\tau_K)}&\left[\sum_{t<T}\left(\frac{1}{1+\delta}\right)^t k_t^m e_{Kt}\frac{\int_i \omega_i u'_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t\int_i \omega_i u'_{ti}\cdot k_t^m}\right] \\ &-1 + \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \frac{\int_i \omega_i u'_{ti}\cdot k_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t\int_i \omega_i u'_{ti}\cdot k_t^m}. \end{aligned}$$

Now we have to redefine the average welfare weight as:

$$\bar{g}_K \equiv \sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t \frac{\int_i u'_{ti}\cdot k_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t\int_i u'_{ti}\cdot k_t^m},$$

and the total elasticity as:

$$e_K = \sum_{t\geq 0}\left(\frac{1}{1+\delta}\right)^t k_t^m e_{Kt}\frac{\int_i u'_{ti}}{\sum_{t\geq T}\left(\frac{1}{1+\delta}\right)^t\int_i u'_{ti}\cdot k_t^m}.$$

With these redefined variables, the same formula holds.

A.3 Optimal Nonlinear Taxes in the Generalized Model

Let $e_K^{top}(t)$ be the average elasticity of total capital income of those individuals with capital income above threshold rk^{top}. It is measured at time t following a small reform of the top bracket tax rate $d\tau_K$ taking place at time 0. The elasticity is weighted by capital income. Let $e_{L,1-\tau_K}(t)$ be the elasticity of labor income of those individuals with capital income above threshold rk^{top}.

Proposition A1. ***Optimal top capital tax rate in the steady state.***

Suppose there is a linear tax on labor income τ_L. The optimal top capital tax rate above capital income level rk^{top} takes the form:

$$\tau_K^{top} = \frac{1 - \bar{g}_K^{top} - \tau_L \cdot \frac{z^m}{r(k^{m,top} - k^{top})} \cdot \bar{e}_{L,(1-\tau_K)}}{1 - \bar{g}_K^{top} + a_K^{top} \cdot \bar{e}_K^{top}},$$

with $\bar{e}_K^{top} \equiv \int_i g_i \delta_i \int_{t=0}^{\infty} e_K^{top}(t) \cdot e^{-\delta_i t} dt$. $\bar{g}_K^{top} = \frac{\int_{i:k_i \geq k^{top}} g_i \cdot (k_i - k^{top})}{\int_{i:k_i \geq k^{top}} (k_i - k^{top})}$ is the average capital income weighted welfare weight in the top capital tax bracket, and $a_K^{top} = \frac{k^{m,top}}{k^{m,top} - k^{top}}$ is the Pareto parameter of the capital income distribution. $\bar{e}_{L,(1-\tau_K)} \equiv \int_i g_i \delta_i \int_{t=0}^{\infty} e_{L,(1-\tau_K)}(t) \cdot e^{-\delta_i t} dt$.

Proof of Proposition A1: We consider the top tax rate τ_K on capital above threshold k^{top}. As r is uniform, this is equivalent to a top tax rate applying above capital income threshold rk^{top}. Let N denote the fraction of individuals above k^{top}. We again use the notation $k^{m,top}$ to denote the average wealth above the top threshold, i.e.:

$$k^{m,top} = \frac{\int_{i:k_i(t) \geq k^{top}} rk_i}{N},$$

Suppose we change the top tax rate on capital by $d\tau_K$. As defined in the text, let $e_K^{top}(t)$ be the elasticity of capital holding of top capital earners (the wealth elasticity of total wealth to the tax rate of those with capital income above rk^{top}). For all individuals above the cutoff, the change in utility is:

$$dV_i = d\tau_K \delta_i \Big[\int_0^{\infty} u_{ic}(c_i(t), k_i(t)) Nr(k^{m,top}(t) - k^{top}) e^{-\delta_i t} - \int_0^{\infty} u_{ic}(c_i(t), k_i(t)) r(k_i(t) - k^{top}) e^{-\delta_i t} - \frac{\tau_K}{1-\tau_K} \int_0^{\infty} u_{ic}(c_i(t), k_i(t)) Nrk^{m,top}(t) e_K^{top}(t) \cdot e^{-\delta_i t} dt \Big].$$

Starting from the steady state, capital levels are constant so that:

$$dV_i = u_{ic} r(k^{m,top} - k^{top}) N d\tau_K \left[1 - \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt \right],$$

where $a_K^{top} = \frac{k^{m,top}}{(k^{m,top} - k^{top})}$.

For individuals below the cutoff, the change in utility is:

$$dV_i = u_{ic} r(k^{m,top} - k^{top}) N d\tau_K \left[1 - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt \right].$$

The change in social welfare is such that:

$$dSWF \propto 1 - \int_{i:k_i \geq k^{top}} g_i \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_i g_i \delta_i \int_0^\infty e_K^{top}(t) \cdot e^{-\delta_i t} dt.$$

Let

$$\bar{g}_K^{top} \equiv \int_{i:k_i \geq k^{top}} g_i \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N} \quad \text{and} \quad e_K^{top} \equiv \int_i g_i \delta_i \int_0^\infty e_K^{top}(t) \cdot e^{-\delta_i t} dt.$$

Then, we obtain the optimal tax rate τ_K such that $dSWF = 0$:

$$\tau_K = \frac{1 - \bar{g}_K^{top}}{1 - \bar{g}_K^{top} + a_K^{top} e_K^{top}}.$$

With endogenous labor, let

$$e_{L,(1-\tau_K)}(t) = \frac{dz^m(t)}{d(1-\tau_K)} \frac{(1-\tau_K)}{z^m(t)} = \frac{dz^m(t)}{d\bar{r}} \frac{\bar{r}}{N z^m(t)}.$$

be the elasticity of aggregate (average) labor income z^m with respect to the top capital tax rate, normalized by N, in the two bracket tax system.

For all individuals with capital income above the cutoff:

$$\begin{aligned} dV_i = d\tau_K \cdot \delta_i \Bigg[& \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) N r (k^{m,top}(t) - k^{top}) \cdot e^{-\delta_i t} \\ & - \frac{\tau_L}{1-\tau_K} \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) z^m(t) N e_{L,(1-\tau_K)}(t) \cdot e^{-\delta_i t} \\ & - \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) r (k_i(t) - k^{top}) \cdot e^{-\delta_i t} \\ & - \frac{\tau_K}{1-\tau_K} \int_0^\infty u_{ic}(c_i(t), k_i(t), z_i(t)) N r k^{m,top}(t) e_K^{top}(t) \cdot e^{-\delta_i t} dt \Bigg]. \end{aligned}$$

Starting from the steady state, capital and labor income are constant over time:

$$dV_i = u_{ic} N r (k^{m,top} - k^{top}) d\tau_K \cdot \Bigg[1 - \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N}$$

$$- \frac{\tau_L}{1-\tau_K} \frac{z^m}{r(k^{m,top} - k^{top})} \int_0^\infty \delta_i e_{L,(1-\tau_K)}(t) \cdot e^{-\delta_i t} dt - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt \Bigg].$$

The change in social welfare is:

$$dSWF = \int_i \omega_i dV_i \propto 1 - \int_{i: rk_i \geq rk^{top}} g_i \frac{(k_i - k^{top})}{(k^{m,top} - k^{top})N}$$

$$- \frac{\tau_L}{1-\tau_K} \frac{z^m}{r(k^{m,top} - k^{top})} \int_i g_i \int_0^\infty \delta_i e_{L,(1-\tau_K)}(t) \cdot e^{-\delta_i t} dt - \frac{\tau_K}{1-\tau_K} a_K^{top} \int_i g_i \int_0^\infty \delta_i e_K^{top}(t) \cdot e^{-\delta_i t} dt.$$

Define e_K^{top}, $e_{L,(1-\tau_K)}$, and $\bar{g}_K^{top}$ as in the text. The optimal formula in the text is then obtained by rearranging the previous condition.

It is straightforward to generalize this to the case of an anticipated reform by discounting the elasticities using $e^{-\delta(t-T)}$ and using the anticipated elasticities rather than the unanticipated ones.

A.4 Optimal Taxation with Horizontal Equity Concerns.

In this section, we formally consider optimal capital and labor taxation under horizontal equity concerns.

As derived in Section 3.2.1, the optimal revenue-maximizing rates are: $\tau_L^R = \frac{1}{1+e_L}$ and $\tau_K^R = \frac{1}{1+e_K}$. Without loss of generality, we suppose that capital is more elastic so that $\tau_K^R < \tau_L^R$. The optimal linear comprehensive tax on income is, as derived in (16):

$$\tau_Y = \frac{1 - \bar{g}_Y}{1 - \bar{g}_Y + e_Y} \quad \text{with} \quad \bar{g}_Y = \frac{\int_i g_i \cdot y_i}{\int_i y_i}$$

Suppose that the distribution of capital and labor income is dense enough, so that at every total income level $y = rk + z$, there are agents with $y = rk$ (capital income only) and $y = z$ (labor income only).

Generalized social welfare weights that capture horizontal equity concerns are such that:

(i) If $\tau_L = \tau_K$, then the social welfare weights g_i are standard. For instance, we can set $g_i = u_{ci}$ for all agents. Any reform that changes taxes should put zero weight on those who after the reform are such that $\tau_L z_i + \tau_K r k_i < \max_j \{\tau_L z_j + \tau_K r k_j | z_j + r k_j = z_i + r k_i\}$, i.e., on those who pay less taxes at a given total income $y = rk_i + z_i$, or, equivalently, have the highest disposable income and consumption at any income. This means that if labor taxes are increased, $g_i = 0$ for those with any positive capital income at each total income level. Conversely, increasing capital taxes will yield $g_i = 0$ for those individuals with some labor income at each total income level.

(ii) If $\tau_L > \tau_K$, then all the social welfare weights are concentrated on those with $\tau_L z_i + \tau_K r k_i > \max_j\{\tau_L z_j + \tau_K r k_j | z_j + rk_j = z_i + rk_i\}$, i.e., on those agents with only labor income. Conversely, if $\tau_L < \tau_K$, all the social welfare weights are on agents with only capital income.

Suppose that, starting from a situation with $\tau_L = \tau_K$ we introduce a small tax break on capital income, $d\tau_K < 0$. Capital income earners now get an unfair advantage and all the weight is concentrated on those with no capital income (equivalently, everyone with $k_i > 0$ receives a weight $g_i = 0$). As a result, a small tax break on capital can only be optimal if it raises tax revenue and, hence, allows to lower the tax rate on labor income as well. This can only occur if $\tau_Y > \tau_K^R$, i.e., the optimal comprehensive tax rate is above the revenue-maximizing rate on capital income.

Proposition A2. ***Optimal labor and capital taxation with horizontal equity concerns.***

(i) If $\tau_Y \leq \tau_K^R$, taxing labor and capital income at the same comprehensive rate $\tau_L = \tau_K = \tau_Y$ is the unique optimum.

(ii) If $\tau_Y > \tau_K^R$, a differential tax system with the capital tax rate set to the revenue maximizing rate $\tau_K = \tau_K^R < \tau_L$ (with both τ_K and τ_L smaller than τ_Y) is the unique optimum.

Proof. Let us consider the two cases in turn.

(i) If $\tau_Y \leq \tau_K^R$.

To see why $\tau_L = \tau_K = \tau^*$ is an equilibrium, suppose that we tried to lower the tax rate on capital income. Then, all the weight will concentrate on people with only labor income, which will then in turn make it optimal to increase the tax on capital again.

This equilibrium is unique. There is no other equilibrium with equal taxes on capital and labor that can raise more revenue with a lower tax rate, by definition of τ_Y as the optimal rate on comprehensive income. There is also no equilibrium with non-equal tax rates on capital and labor. Suppose that we tried to set (without loss of generality) $\tau_K < \tau_L$. Then to raise enough revenue we would require that $\tau_K < \tau_Y < \tau_L$. Since capital owners are now advantaged, all the social welfare weight concentrates on people with only labor income. Since then a fortiori $\tau_K < \tau_K^R$, increasing τ_K would mean that more revenue would be raised, which would allow us to lower τ_L, which is good since all weight is on people with only labor income.

(ii) If $\tau_Y > \tau_K^R$.

In this case, the equilibrium has $\tau_K = \tau_K^R < \tau_Y$ and $\tau_Y > \tau_L > \tau_K^R$. Clearly this is an equilibrium since we cannot decrease τ_L without losing revenue and we cannot raise more revenue through τ_K (since it is already set at the revenue-maximizing rate for the capital tax base). In addition, we cannot decrease τ_K further without increasing τ_L, which is not desirable since it would benefit people capital income earners, who already receive a weight of zero.

This equilibrium is also unique. If we set $\tau_L = \tau_K$ equal, we should set them equal to τ_Y which is the optimal tax rate on comprehensive income. But then, since τ_K is now above its revenue maximizing rate, we could lower both τ_K and τ_L without losing revenues, so this would not be an equilibrium. On the other hand, as long as we set $\tau_K < \tau_L$, capital income earners get zero weight and the only possibility is to go all the way to $\tau_K = \tau_K^R$ since only people with only labor income have a non-zero weight.

□

As a result, horizontal equity concerns will be a force pushing towards the comprehensive income tax system derived in Section 3.2.1. In the text, we provided an efficiency argument in favor of a tax on comprehensive income (based on income shifting opportunities) while the argument here is based on equity considerations. With horizontal equity preferences, deviations from a comprehensive income tax system can only be justified if they raise more revenue and generate a Pareto-improvement, which drastically reduces the scope for them. In Saez and Stantcheva (2016) we argue that this is akin to a generalized Rawlsian principle whereby discrimination against some groups (e.g., capital owners versus labor providers) is only permissible if it makes the group discriminated against better off, i.e., if it generates a Pareto improvement.

A.4.1 Horizontal Equity with Nonlinear Taxation

The same reasoning as for linear taxation with horizontal equity also applies to nonlinear taxes. Starting from a comprehensive tax system $T_Y(z + rk)$ as derived in Section 3.2.1, lowering the tax rate on capital income, conditional on a given total income level, will generate a horizontal inequity and concentrate all social weight on those with no capital income conditional on that total income level. Such a preferential tax break for capital income earners will only be acceptable if it generates more revenue and allows to lower the tax rate on labor income as well. We show this below.

Formally, suppose that we start from the optimal tax on comprehensive income, $T_Y(rk + z)$, as derived in Section 3.2.1 which does not discriminate between capital and labor income conditional on total income. We say that a tax system unambigously favors capital at income level $y = rk + z$, if for any (rk, z) such that $y = rk + z$, and any $\varepsilon \in]0, z]$, $T_Y(rk, z) > T(rk + \varepsilon, z - \varepsilon)$ (having more capital income, conditional on a given total income leads to lower taxes). (Note that it may be the case that a tax system favors capital only at some y levels or only at some rk, z ranges.)

Denote a change in the tax by $\delta T(rk, z)$.

A deviation $\delta T(rk, z)$ is said to introduce horizontal inequity, if, starting from a comprehensive tax system $T_Y(rk+z)$, the resulting tax system $T_Y(z+rk)+\delta T(rk, z)$ cannot be expressed as $\tilde{T}_Y(rk+z)$ for some function $\tilde{T}_Y$.

With nonlinear taxes, we can again define the generalized social welfare weights as follows.

i) If there is a comprehensive tax $T_Y(z+rk)$, then everybody has standard weights, such as, for instance, $g_i = u_{ci}$. For any deviation $\delta T(rk, z)$ that introduces horizontal inequity, the weights concentrate on the agents who pay the highest tax at a given total income level, i.e., on those with $T_Y(z_i+rk_i)+\delta T(rk_i, z_i) = \max_j \{T_Y(z_j+rk_j)+\delta T(rk_j, z_j) | z_j+rk_j = rk_i+z_i\}$ (which is equivalent to putting all the weight on the agent(s) with lowest disposable income at any total income level).

Hence, the weights also need to depend on $\delta T(z, rk)$, the direction of the tax reform.

ii) If the tax is such that $T(rk, z)$ cannot be expressed as $\tilde{T}_Y(rk+z)$ for some function $\tilde{T}_Y$, then the weights concentrate on those with
$T(z_i, rk_i) = \max_j \{T(z_j, rk_j) | z_j+rk_j = rk_i+z_i\}$, i.e., on the agents which pay the highest tax (equivalently, have the lowest disposable income) conditional on total income.

Equilibria:

Suppose that, at the comprehensive tax rate, no small reform $\delta T(rk, z)$ that introduces horizontal equity and favors capital (according to our definitions above) can increase total tax revenues, i.e., for all $\delta T(rk, z)$ that favor capital and introduce horizontal inequity, the alternative tax system $\tilde{T}_Y(rk, z) = T_Y(rk+z)+\delta T(rk, z)$ is such that:

$$\int_i T_Y(rk_i(T_Y)+z_i(T_Y))di > \int_i \tilde{T}_Y(rk_i(\tilde{T}_Y), z_i(\tilde{T}_Y))di$$

where naturally, the choices $z_i(T)$ and $k_i(T)$ depend on the tax system T. Then the unique equilibrium has the comprehensive tax system in place, as derived in 3.2.1. No horizontal inequity can be an equilibrium unless it introduces a Pareto improvement.

Suppose on the other hand that if the revenue maximizing tax rate on capital, $T_K^R(rk)$ were implemented, and a labor income tax $T_L(z)$ was used to complement it, more revenue could be raised than with the tax on comprehensive income $T_Y(rk+z)$ and the tax burden on all agents would be lower than under the comprehensive income tax. Then, the optimum is to set differential taxes on capital and labor income, with the capital tax at its optimal revenue-maximizing schedule. Horizontal inequity is an equilibrium because it generates a Pareto improvement.

A.5 Progressive Consumption Taxes

The progressive consumption tax is defined on an exclusive basis as $t_C(.)$ such that

$$\dot{k} = \bar{r}k + z - [c + t_c(c)]$$

Equivalently, we can again define the inclusive consumption tax $T_C(y)$ on pre-tax resources y devoted to consumption such that $c + t_c(c) = y$ is equivalent to $c = y - T_C(y)$, i.e., $y \to y - T_C(y)$ is the inverse function of $c \to c + t_c(c)$ and hence $1 + t'_C = 1/(1 - T'_C)$.

The case of a progressive consumption tax is most easily explained with inelastic labor income (possibly heterogenous across individuals). Real wealth k^r in the presence of the progressive consumption tax is:

$$k^r(k) = k - \frac{T_C(\bar{r}k + z) - T_C(z)}{\bar{r}}$$

Recall that real wealth is defined as nominal wealth adjusted for the price of consumption. There are two ways to see that the above is the right expression. First, wealth k provides an income stream $\bar{r}k$ which translates into extra permanent consumption equal to the income minus the tax paid on the extra consumption $\bar{r}k - [T_C(\bar{r}k + z) - T_C(z)]$ which can be capitalized into wealth k^r by dividing by $\bar{r}$. If labor income is heterogeneous across agents, then $k^r(k, z)$ should also be indexed by z. Another way to see this is to ask what the capital k^r would be that would yield the same disposable income as the nominal capital under the consumption tax. Disposable income in terms of real capital k^r is $\bar{r}k^r - T_C(z)$. Disposable income expressed in terms of nominal capital is: $\bar{r}k - T_C(\bar{r}k + z)$. These two must be equal, which yields the expression for k^r above. k^r has three natural properties: with no consumption tax, real and nominal wealth are equal, $dk^r/dk = 1 - T'_C$, i.e., and extra dollar of nominal wealth is worth $1 - T'_C$ in real terms, and $k^r(0) = 0$.

In that case, we have in steady-state

$$c = \bar{r}k + z - T_C(\bar{r}k + z) = \bar{r}k^r + z - T_C(z)$$

and the first order condition for utility maximization is $a'_i(k^r) = \delta - \bar{r}$. Hence, real capital is chosen to satisfy the same condition as nominal capital when there is no consumption tax. Put differently, any consumption tax will be undone by agents in terms of their savings and will have no effect on the real value of their wealth held (and, hence, by definition of the real wealth, on their purchasing power). Hence, the consumption tax is equivalent to a tax on labor income only.

The equivalence is not exact with elastic labor supply, as in that case, the marginal consumption tax depends on the labor choice and the first-order condition for labor income is $h_i'(z) = 1 - T_C'(\bar{r}k + z) + a_i'(k^r)[T_C'(\bar{r}k + z) - T_C'(z)]/\bar{r}$.

References

Aiyagari, R. (1995). Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting. *Journal of Political Economy 103*(6), 1158–1175.

Judd, K. L. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of public Economics 28*(1), 59–83.

Piketty, T. and E. Saez (2013). A theory of optimal inheritance taxation. *Econometrica 82*(4), 1241–1272.

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*(1), 24–45.