Taxation under Learning-by-Doing: Supplementary Material

Miltiadis Makris
University of Southampton

Alessandro Pavan
Northwestern University

October 2017

Abstract

Section 1 in this supplement shows how the allocations in the two-period economy in the main text relate to their counterparts in the 40-period economy considered in the quantitative analysis in the last section of the paper. Section 2 shows that the wedges derived through the allocation approach coincide with the wedges derived through the perturbation approach.

1 The 40-period Model

Suppose each worker works for $T = 2\widehat{T}$ periods and discounts the future with the discount factor β. Productivity is constant within each of the two blocks of a worker's life, with each block comprising $\widehat{T}$ periods. In the quantitative analysis in the main text $\widehat{T} = 20$ with each period corresponding to a year. Let θ_1 denote productivity in the first $\widehat{T}$ periods and θ_2

denote productivity in the second $\widehat{T}$ periods. Moreover, assume that

$$\theta_2 = h_2 \theta_1^{\rho} \left(\frac{\sum_{s=1}^{\widehat{T}} \hat{\beta}_s y_s}{\sum_{s=1}^{\widehat{T}} \hat{\beta}_s} \right)^{\zeta} \varepsilon_2.$$

Note that LBD is active in each of the first $\widehat{T}$ periods. Also note that the above representation implies that LBD is stronger in earlier years than in later ones. Assume that $(\widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_{\widehat{T}})$ is proportional to $(1, \beta, ..., \beta^{\widehat{T}-1})$, in which case $\frac{\widehat{\beta}_t}{\sum_{s=1}^{\widehat{T}} \widehat{\beta}_s} = \frac{\beta^{t-1}}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}}$.

Next let

$$\overline{y}(\theta_1) \equiv \frac{\sum_{s=1}^{\widehat{T}} \beta^{s-1} y_s(\theta_1)}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}}$$

denote the average income generated by an agent of initial productivity θ_1 over the first $\widehat{T}$ periods. The expected lifetime utility of a worker of period-1 productivity equal to θ_1 is given by

$$V_1(\theta_1) \equiv \textstyle\sum_{t=1}^{\widehat{T}} \beta^{t-1} [v(c_t(\theta_1)) - \psi(y_t(\theta_1), \theta_1)]$$

$$+\beta^{\widehat{T}} \int \left\{ \textstyle\sum_{t=1}^{\widehat{T}} \beta^{t-1} [v(c_{\widehat{T}+t}((\theta_1, \theta_2))) - \psi(y_{\widehat{T}+t}((\theta_1, \theta_2)), \theta_2)] \right\} dF_2(\theta_2 | \theta_1, \overline{y}(\theta_1)).$$

With an abuse of notation, hereafter, let $\theta^s = \theta_1$ and $\theta_s = \theta_1$ for all $s = 1, ..., \widehat{T}$, and $\theta^s = (\theta_1, \theta_2)$ and $\theta_s = \theta_2$ for all $s = \widehat{T} + 1, ..., 2\widehat{T}$.

1.1 Optimal allocations

We now show that the allocations under the optimal tax code in the 40-period economy coincide with those in the two-period model in the main text.

For any $t = 1, ..., \widehat{T}$, let

$$V_t(\theta^t) \equiv \sum_{s=t}^{\widehat{T}} \beta^{s-t} [v(c_s(\theta^s)) - \psi(y_s(\theta^s), \theta_s)] +$$

$$\beta^{\widehat{T}+1-t}\int\left\{\sum_{s=1}^{\widehat{T}}\beta^{s-1}[v(c_{\widehat{T}+s}(\theta_1,\theta_2))-\psi(y_{\widehat{T}+s}((\theta_1,\theta_2)),\theta_2)]\right\}dF_2(\theta_2|\theta_1,\overline{y}(\theta_1))$$

denote the continuation expected utility of each worker whose productivity in the first block is equal to θ_1. Observe that

$$V_t(\theta^t)=v(c_t(\theta_1))-\psi(y_t(\theta_1),\theta_1)+\beta\Pi_{t+1}(\theta^t),$$

where, for any $t=1,...,\widehat{T}$,

$$\Pi_{t+1}(\theta^t)\equiv\sum_{s=t+1}^{\widehat{T}}\beta^{s-t-1}[v(c_s(\theta_1))-\psi(y_s(\theta_1),\theta_1)]+$$

$$\beta^{\widehat{T}-t}\int\left\{\sum_{s=1}^{\widehat{T}}\beta^{s-1}[v(c_{\widehat{T}+s}(\theta_1,\theta_2))-\psi(y_{\widehat{T}+s}(\theta_1,\theta_2),\theta_2)]\right\}dF_2(\theta_2|\theta_1,\overline{y}(\theta_1))$$

denotes the utility expected from period $t+1$ onwards.

Likewise, for any $t=\widehat{T}+1,...,2\widehat{T}$, let

$$V_t(\theta^t)\equiv\sum_{s=t}^{2\widehat{T}}\beta^{s-t}[v(c_t(\theta^t))-\psi(y_t(\theta^t),\theta_t)]$$

denote the continuation expected utility of each worker whose productivity in the first block is equal to θ_1 and whose productivity in the second block is equal to θ_2. Observe that, for any $t=\widehat{T}+1,...,2\widehat{T}$,

$$V_t(\theta^t)=v(c_t(\theta_1,\theta_2))-\psi(y_t(\theta_1,\theta_2),\theta_2)+\beta\Pi_{t+1}(\theta^t),$$

where, for any $t = \widehat{T}+1, ..., 2\widehat{T}-1$,

$$\Pi_{t+1}(\theta^t) \equiv \sum_{s=t+1}^{2\widehat{T}} \beta^{s-t-1}[v(c_t((\theta^t))) - \psi(y_t(\theta^t)), \theta_t)] = V_{t+1}(\theta^{t+1}),$$

whereas, for $t = 2\widehat{T}$,

$$\Pi_{2\widehat{T}+1}(\theta^{2\widehat{T}}) \equiv 0.$$

Also observe that, for any $t = 1, ..., \widehat{T}-1$,

$$\Pi_{t+1}(\theta^t) = V_{t+1}(\theta^{t+1}),$$

whereas, for $t = \widehat{T}$,

$$\Pi_{\widehat{T}+1}(\theta^t) = \int V_{\widehat{T}+1}(\theta_1, \theta_2) dF_2(\theta_2|\theta_1, \overline{y}(\theta_1)).$$

The local IC conditions for this 40-period economy are

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_s(\theta_1), \theta_1) - \beta^{\widehat{T}}\mathbb{E}^{\lambda[\chi]|\theta_1}\left[I_1^2(\tilde{\theta}, \overline{y}(\theta_1)) \sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\theta_1, \tilde{\theta}_2), \tilde{\theta}_2)\right]$$

almost all θ_1, and

$$\frac{\partial V_{\widehat{T}+1}(\theta_1, \theta_2)}{\partial \theta_2} = -\sum_{s=\widehat{T}+1}^{2\widehat{T}} \beta^{s-\widehat{T}-1}\psi_\theta(y_t(\theta), \theta_2) = -\sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\theta_1, \theta_2), \theta_2)$$

all θ_1, almost all $\theta_2 \in Supp[F_2(\cdot|\theta_1, \overline{y}(\theta_1)]$. Note that $\lambda[\chi]|\theta_1$ is the probability measure over $\theta \equiv (\theta_1, \theta_2)$ given θ_1 and $\overline{y}(\theta_1)$.

Denoting by $C(x) \equiv v^{-1}(x)$, we have that he planner's relaxed problem can be described as follows:

$$\max_{\substack{y_t(\cdot), V_t(\cdot), \Pi_{\widehat{T}+1}(\cdot), \\ Z_{\widehat{T}+1}(\cdot), t=1,...,\widehat{T}}} \int \sum_{s=1}^{\widehat{T}-1} \beta^{s-1}\{[y_s(\theta_1) - C\left(V_s(\theta_1) + \psi(y_s(\theta_1), \theta_1) - \beta V_{s+1}(\theta_1)]\right)$$

$$+\beta^{\widehat{T}-1}[y_{\widehat{T}}(\theta_1) - C\left(V_{\widehat{T}}(\theta_1) + \psi(y_{\widehat{T}}(\theta_1), \theta_1) - \beta \Pi_{\widehat{T}+1}(\theta_1)]\right) + \beta^{\widehat{T}} Q_2(\theta_1, \overline{y}(\theta_1), \Pi_{\widehat{T}+1}(\theta_1), Z_{\widehat{T}+1}(\theta_1))\} dF_1(\theta_1)$$

subject to

$$(1-r)V_1(\underline{\theta}_1) + r\int V_1(s)dF_1(s) - \kappa = 0, \tag{1}$$

and

$$\frac{\partial V_1(\theta_1)}{\partial \theta_1} = -\sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_s(\theta_1), \theta_1) + \beta^{\widehat{T}} Z_{\widehat{T}+1}(\theta_1), \tag{2}$$

where

$$Q_2(\theta_1, \overline{y}(\theta_1), \Pi_{\widehat{T}+1}(\theta_1), Z_{\widehat{T}+1}(\theta_1)) \equiv$$

$$\max_{\substack{y_{\widehat{T}+t}(\theta_1, \cdot), \\ V_{\widehat{T}+t}(\theta_1, \cdot), \\ t=1,...,\widehat{T}}} \int \sum_{s=1}^{\widehat{T}} \beta^{s-1}\left\{y_{\widehat{T}+s}(\theta) - C(V_{\widehat{T}+s}(\theta) + \psi(y_{\widehat{T}+s}(\theta), \theta_2) - \beta V_{\widehat{T}+s+1}(\theta)\right\} dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1))$$

subject to

$$V_{2\widehat{T}+1}(\theta) = 0,$$

$$\Pi_{\widehat{T}+1}(\theta_1) = \int V_{\widehat{T}+1}(\theta) dF_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)), \tag{3}$$

$$Z_{\widehat{T}+1}(\theta_1) = -\mathbb{E}^{\lambda[\chi]|\theta_1}\left[I_1^2(\tilde{\theta}, \overline{y}(\theta_1)) \sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\theta_1, \tilde{\theta}_2), \tilde{\theta}_2)\right], \tag{4}$$

and

$$\frac{\partial V_{\widehat{T}+1}(\theta_1,\theta_2)}{\partial\theta_2} = -\sum_{s=1}^{\widehat{T}} \beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\theta_1,\theta_2),\theta_2). \tag{5}$$

The above is a two-stage optimal control problem. In the first problem, the controls are $(y_1(\theta_1),...,y_{\widehat{T}}(\theta_1),V_2(\theta_1),...,V_{\widehat{T}}(\theta_1),\Pi_{\widehat{T}+1}(\theta_1),Z_{\widehat{T}+1}(\theta_1))$, while the state variable is $V_1(\theta_1)$. In the second problem, the controls are $(y_{\widehat{T}+1}(\theta),...,y_{2\widehat{T}}(\theta),V_{\widehat{T}+2}(\theta),...,V_{2\widehat{T}}(\theta))$ while the state variable is $V_{\widehat{T}+1}(\theta)$.

Also note that the first-best allocations solve a similar problem but without the local IC constraints. Thus, the FB allocations can be read from the SB allocations characterized below by setting the costate variables to zero.

1.1.1 Solution to the Relaxed Program: Utilitarian preferences and no moving supports

Using the property that

$$\begin{aligned} &-\int I_1^2(\theta,\overline{y}(\theta_1))\textstyle\sum_{s=1}^{\widehat{T}}\beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\theta_1,\theta_2),\theta_2)f_2(\theta_2\mid\theta_1,\overline{y}(\theta_1))d\theta_2 \\ &= \int I_1^2(\theta,\overline{y}(\theta_1))\tfrac{\partial V_{\widehat{T}+1}(\theta_1,\theta_2)}{\partial\theta_2}f_2(\theta_2\mid\theta_1,\overline{y}(\theta_1))d\theta_2 = \int V_{\widehat{T}+1}(\theta_1,\theta_2)\tfrac{\partial}{\partial\theta_1}f_2(\theta_2\mid\theta_1,\overline{y}(\theta_1))d\theta_2, \end{aligned} \tag{6}$$

we have that the integral constraint (4) can be conveniently rewritten as

$$Z_{\widehat{T}+1}(\theta_1) = \int V_{\widehat{T}+1}(\theta_1,\theta_2)\frac{\partial}{\partial\theta_1}f_2(\theta_2\mid\theta_1,\overline{y}(\theta_1))d\theta_2. \tag{7}$$

As usual, we proceed backwards, by solving first for the policies that correspond to the second block. Let $\pi_2(\theta_1)$ and $\xi_2(\theta_1)$ be the multipliers associated with the two integral constraints (3) and (7) and $\mu_2(\theta_1,\theta_2)$ the costate variable for the law of motion of $V_{\widehat{T}+1}(\theta_1,\theta_2)$.

Along with (3), (5), and (7), the following are necessary optimality conditions:

$$\frac{1}{v'(c_{\widehat{T}+s}(\theta))} = \frac{1}{v'(c_{\widehat{T}+s+1}(\theta))}, \text{ for all } s = 1, ..., \widehat{T} - 1, \tag{8}$$

$$1 - \frac{\psi_y(y_{\widehat{T}+s}(\theta), \theta_2)}{v'(c_{\widehat{T}+s}(\theta))} - \mu_2(\theta)\frac{\psi_{\theta y}(y_{\widehat{T}+s}(\theta), \theta_2)}{f_2(\theta_2 \mid \theta_2, \overline{y}(\theta_1))} = 0, \text{ for all } s = 1, ..., \widehat{T}, \tag{9}$$

$$\frac{\partial \mu_2(\theta)}{\partial \theta_2} = f_2(\theta_2 \mid \theta_1, \bar{y}(\theta_1)) \cdot \left\{ \frac{1}{v'(c_{\widehat{T}+1}(\theta))} + \pi_2(\theta_1) + \xi_2(\theta_1)\frac{\partial f_2(\theta_2 \mid \theta_2, \overline{y}(\theta_1))/\partial \theta_1}{f_2(\theta_2 \mid \theta_2, \overline{y}(\theta_1))} \right\}, \tag{10}$$

along with the boundary conditions

$$\mu_2(\theta_1, \underline{\theta}_2) = 0, \tag{11}$$

$$\mu_2(\theta_1, \overline{\theta}_2) = 0, \tag{12}$$

where

$$c_{\widehat{T}+1}(\theta) = C(V_{\widehat{T}+1}(\theta) + \psi(y_{\widehat{T}+1}(\theta), \theta_2) - \beta V_{\widehat{T}+2}(\theta)).$$

Conditions (8) and (9) imply that $c_{\widehat{T}+1} = c_{\widehat{T}+2} = ... = c_{2\widehat{T}}$ and, $y_{\widehat{T}+1} = y_{\widehat{T}+2} = ... = y_{2\widehat{T}}$. It is then immediate to see that the policies that solve the above conditions coincide with the period-2 policies that solve the relaxed program in the two-period model. That is, for any $s = 1, ..., \hat{T}$, $(c_{\widehat{T}+s}(\theta_1, \theta_2), y_{\widehat{T}+s}(\theta_1, \theta_2)) = (c_2^{2pm}(\theta_1, \theta_2), y_2^{2pm}(\theta_1, \theta_2))$, where $c_2^{2pm}(\theta_1, \theta_2), y_2^{2pm}(\theta_1, \theta_2))$ are the policies that solve the relaxed program in the two-period model. Furthermore, the continuation utility at the beginning of period $\widehat{T} + 1$ satisfies

$$V_{\widehat{T}+1}(\theta) = (1 + \beta + ... + \beta^{\widehat{T}-1})V_2^{2pm}(\theta),$$

where $V_2^{2pm}(\theta)$ is the continuation utility in the two-period model.

Next, consider the choice of the policies for the first block. Let $\mu_1(\theta_1)$ be the costate variable associated with the constraint (2) and π_1 the multiplier associated with the redistribution constraint

$$\int V_1(\theta_1)dF_1(\theta_1) = \kappa. \tag{13}$$

In addition to (2) and (13), the following optimality conditions must hold:

$$\frac{1}{v'(c_s(\theta_1))} = \frac{1}{v'(c_{s+1}(\theta_1))}, \text{ for all } s = 1, ..., \widehat{T} - 1, \tag{14}$$

$$1 - \tfrac{\psi_y(y_s(\theta_1),\theta_1)}{v'(c_s(\theta_1))} + \beta^{\widehat{T}} \int \left\{ y_{\widehat{T}+1}(\theta) - c_{\widehat{T}+1}(\theta) - \pi_2(\theta_1) \tfrac{V_{\widehat{T}+1}(\theta)}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}} \right\} \tfrac{\partial}{\partial \overline{y}} f_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) d\theta_2$$

$$+\beta^{\widehat{T}} \xi_2(\theta_1) \tfrac{\partial}{\partial \overline{y}} \int I_1^2(\theta, \overline{y}(\theta_1)) \psi_\theta(y_{\widehat{T}+1}(\theta), \theta_2) f_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) d\theta_2 - \mu_1(\theta_1) \tfrac{\psi_{\theta y}(y_s(\theta_1),\theta_1)}{f_1(\theta_1)} = 0, \text{ for } s = 1, ..., \widehat{T}, \tag{15}$$

$$\frac{\partial \mu_1(\theta_1)}{\partial \theta_1} = f_1(\theta_1) \cdot \left\{ \frac{1}{v'(c_1(\theta_1))} + \pi_1 \right\}, \tag{16}$$

$$\frac{1}{v'(c_1(\theta_1))} + \pi_2(\theta_1) = 0, \tag{17}$$

$$\mu_1(\theta_1) + \xi_2(\theta_1) f_1(\theta_1) = 0, \tag{18}$$

along with the boundary conditions

$$\mu_1(\underline{\theta}_1) = 0, \tag{19}$$

and

$$\mu_1(\overline{\theta}_1) = 0, \tag{20}$$

where

$$c_1(\theta_1) = C\left(V_1(\theta_1) + \psi(y_1(\theta_1), \theta_1) - \beta V_2(\theta_1)\right).$$

Note that, when writing the FOCs with respect to $\Pi_{\widehat{T}+1}(\theta_1)$, $Z_{\widehat{T}+1}(\theta_1)$ and $y_s(\theta_1)$, we have used the properties that $\frac{\partial Q_2}{\partial \Pi_{\widehat{T}+1}} = \pi_2(\theta_1)$, $\frac{\partial Q_2}{\partial Z_{\widehat{T}+1}} = \xi_2(\theta_1)$, and (6), along with the fact that

$$\begin{gathered}\tfrac{\partial Q_2}{\partial \overline{y}} = \int \sum_{s=1}^{\widehat{T}} \beta^{s-1} \left\{ y_{\widehat{T}+s}(\theta) - c_{\widehat{T}+s}(\theta) \right\} \tfrac{\partial}{\partial \overline{y}} f_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) d\theta_2 \\ -\pi_2(\theta_1) \int V_{\widehat{T}+1}(\theta) \tfrac{\partial}{\partial \overline{y}} f_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) d\theta_2 \\ +\xi_2(\theta_1) \tfrac{\partial}{\partial \overline{y}} \int I_1^2(\theta, \overline{y}(\theta_1)) \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta), \theta_2) f_2(\theta_2 \mid \theta_1, \overline{y}(\theta_1)) d\theta_2\end{gathered}$$

and

$$\frac{\partial \overline{y}}{\partial y_s} = \frac{\beta^{s-1}}{1 + \beta + ... + \beta^{\widehat{T}-1}}.$$

We also used the property that consumption and earnings allocations are constant over the second block.

Clearly, (14) and (15) imply that $c_1 = c_2 = ... = c_{\widehat{T}}$ and $y_1 = y_2 = ... = y_{\widehat{T}}$. Given this property, we have that the necessary conditions in the above program reduce to the same conditions for the period-1 policies in the relaxed program in the two-period model, with $\delta = \beta^{\widehat{T}}$. We conclude that consumption and earnings in the first block of the 40-period economy are given by the period-1 consumption and earnings polices in the two-period model. That is, for any $s = 1, ..., \hat{T}$, $(c_s(\theta_1), y_s(\theta_1)) = (c_1^{2pm}(\theta_1), y_1^{2pm}(\theta_1))$, where $c_1^{2pm}(\theta_1), y_1^{2pm}(\theta_1))$ are the optimal policies in the two-period model. Furthermore, the lifetime expected utility of each worker with productivity equal to θ_1 in the first block is given by

$$V_1(\theta_1) = (1 + \beta + ... + \beta^{\widehat{T}-1}) V_1^{2pm}(\theta_1),$$

where $V_1^{2pm}(\theta_1)$ is the lifetime expected utility in the two-period model.

1.1.2 Solution to the Relaxed Program: Rawlsian preferences and/or moving supports

Following steps similar to the ones above, one can see that the optimal policies under Rawlsian preferences for redistribution and/or moving supports in the 40-period economy are constant over each of the two blocks, with consumptions and earnings coinciding with their counterparts in the two-period model.

1.1.3 Sufficiency

We conclude by showing that, when the solution to the relaxed program in the two-period model satisfies all the integral monotonicity constraints of the two-period model, then the solution to the relaxed program in the 40-period model (which, by virtue of the results above, consists in the repetition over each of the two blocks of the corresponding policies in the two-period model) also satisfies all the corresponding integral-monotonicity conditions in the 40-period model.

Let $(c_t^{2pm}(\theta), y_t^{2pm}(\theta))_{t=1,2}$ denote the solution to the relax program in the two-period model. Assume these policies satisfy the following integral monotonicity constraints: for any pair $\theta_2, \hat{\theta}_2 \in \Theta_2$ and any pair $\theta_1, \hat{\theta}_1 \in \Theta_1$,

$$\int_{\hat{\theta}_2}^{\theta_2} \psi_\theta(y_2^{2pm}(\theta_1, r), r)dr \leq \int_{\hat{\theta}_2}^{\theta_2} \psi_\theta(y_2^{2pm}(\theta_1, \hat{\theta}_2), r)dr \tag{21}$$

and

$$\begin{aligned} & \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1^{2pm}(r), r) + \delta \int \left[I_1^2((r, \theta_2), y_1^{2pm}(r)) \psi_\theta(y_2^{2pm}(r, \theta_2), \theta_2) \right] dF_2(\theta_2 | r, y_1^{2pm}(r)) \right\} dr \\ \leq & \int_{\hat{\theta}_1}^{\theta_1} \left\{ \psi_\theta(y_1^{2pm}(\hat{\theta}_1), r) + \delta \int \left[I_1^2((r, \theta_2), y_1^{2pm}(\hat{\theta}_1)) \psi_\theta(y_2^{2pm}(\hat{\theta}_1, \theta_2), \theta_2) \right] dF_2(\theta_2 | r, y_1^{2pm}(\hat{\theta}_1)) \right\} dr. \end{aligned} \tag{22}$$

Next observe that the integral monotonicity conditions in the 40-period model require that, for any pair $\theta_2, \hat{\theta}_2 \in \Theta_2$ and any pair $\theta_1, \hat{\theta}_1 \in \Theta_1$,

$$\int_{\hat{\theta}_2}^{\theta_2} \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta_1, r), r)dr \leq \int_{\hat{\theta}_2}^{\theta_2} \sum_{s=1}^{\widehat{T}} \beta^{s-1} \psi_\theta(y_{\widehat{T}+s}(\theta_1, \hat{\theta}_2), r)dr,$$

and

$$\int_{\hat{\theta}_1}^{\theta_1}\left\{\sum_{s=1}^{\widehat{T}}\beta^{s-1}\psi_\theta(y_s(r),r)+\beta^{\widehat{T}}\int I_1^2((r,\theta_2),\overline{y}(r))\sum_{s=1}^{\widehat{T}}\beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(r,\theta_2),\theta_2)dF_2(\theta_2|r,\overline{y}(r))\right\}dr$$

$$\leq\int_{\hat{\theta}_1}^{\theta_1}\left\{\sum_{s=1}^{\widehat{T}}\beta^{s-1}\psi_\theta(y_s(\hat{\theta}_1),r)+\beta^{\widehat{T}}\int I_1^2((r,\theta_2),\overline{y}(\hat{\theta}_1))\sum_{s=1}^{\widehat{T}}\beta^{s-1}\psi_\theta(y_{\widehat{T}+s}(\hat{\theta}_1,\theta_2),\theta_2)dF_2(\theta_2|r,\overline{y}(\hat{\theta}_1))\right\}dr.$$

It is easy to see that when, for any $s=1,...,\hat{T}$, $(c_s(\theta_1),y_s(\theta_1))=(c_1^{2pm}(\theta_1),y_1^{2pm}(\theta_1))$ and $(c_{\widehat{T}+s}(\theta_1,\theta_2),y_{\widehat{T}+s}(\theta_1,\theta_2))=(c_2^{2pm}(\theta_1,\theta_2),y_2^{2pm}(\theta_1,\theta_2))$, the above integral monotonicity conditions reduce to their counterparts in the two-period economy after observing that $\overline{y}(\theta_1)=y_1^{2pm}(\theta_1)$ and $\delta=\beta^{\widehat{T}}$. Hence, when the solution to the relaxed program also solves the full program in the two-period economy, the same is true in the 40-period economy.

1.2 Allocations under Past-Income-Invariant Taxation

Finally we show that, when agents face tax schedules which are independent of past income levels (which is the case when the tax code is the one that approximates the current US code, the quasi-optimal code, or the linear code discussed in the main body), consumption and labor supply are constant over each of the two blocks and coincide with the corresponding levels in the two-period version of the same economy when the discount factor is given by $\delta=\beta^{\hat{T}}$.

To see this, suppose that, in each period $s=1,...,\widehat{T}$, workers face a tax schedule $\mathcal{T}_1(y_s)$, whereas in periods $s=\widehat{T}+1,...,2\widehat{T}$, they face a tax schedule $\mathcal{T}_2(y_s)$. They then choose consumption and earnings in each period to maximize their expected continuation utility subject to the budget constraint

$$c_s=y_s-\mathcal{T}_t(y_s)+\frac{S_s}{\beta}-S_{s+1},$$

where $\mathcal{T}_t(y_s)=\mathcal{T}_1(y_s)$ if $s\leq\widehat{T}$, and $\mathcal{T}_t(y_s)=\mathcal{T}_2(y_s)$ if $\widehat{T}<s\leq 2\widehat{T}$. The variable S_{s+1}

represents the balance in the worker's savings account at the end of period s, with S_0 predetermined and

$$S_{2\widehat{T}+1} = 0.$$

Importantly, note that the (after-tax) return on savings is equal to the inverse of the annual discount factor, that is, $(1 + r(1 - \tau_{capital})) = 1/\beta$, where $\tau_{capital}$ denotes the capital tax rate.

The optimal allocations then solve the following necessary conditions

$$v'(c_s(\theta^s)) = v'(c_{s+1}(\theta^s)), \text{ for all } s = 1, ..., \widehat{T} - 1, \widehat{T} + 1, ..., 2\widehat{T} - 1,$$

$$v'(c_{\widehat{T}}(\theta_1)) = \int v'(c_{\widehat{T}+1}(\theta_1, \theta_2)) dF_2(\theta_2|\theta_1, \overline{y}(\theta_1)), \tag{23}$$

$$1 - \mathcal{T}_2'(y_s(\theta^s)) = \frac{\psi_y(y_s(\theta^s), \theta_s)}{v'(c_s(\theta^s))}, \text{ for all } s = \widehat{T} + 1, ..., 2\widehat{T}, \tag{24}$$

$$\begin{gathered} v'(c_s(\theta^s))[1 - \mathcal{T}_1'(y_s(\theta^s))] \\ + \frac{\beta^{\widehat{T}}}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}} \int \left\{ \sum_{t=1}^{\widehat{T}} \beta^{t-1} [v(c_{\widehat{T}+t}((\theta_1, \theta_2))) - \psi(y_{\widehat{T}+t}((\theta_1, \theta_2)), \theta_2)] \right\} \frac{\partial f_2(\theta_2|\theta_1, \overline{y}(\theta_1))}{\partial \overline{y}(\theta_1)} d\theta_2 \\ = \psi_y(y_s(\theta^s), \theta_s), \text{ for all } s = 1, ..., \widehat{T}. \end{gathered} \tag{25}$$

Note that Condition (25) uses the fact that

$$\frac{\partial \overline{y}(\theta_1)}{\partial y_s} = \frac{\beta^{s-1}}{\sum_{s=1}^{\widehat{T}} \beta^{s-1}}.$$

Clearly, workers choose the same consumption within each block of their working life, with consumption across the two blocks satisfying the standard Euler condition (23). Given this, Condition (24) implies that output is constant in the second block. In turn, Condition (25) implies that output is also constant over the first block. Given the above observations, it is then immediate to see that consumption and output decisions in this multi-period economy coincide with their counterparts in the two-period version of the same economy in which $\delta = \beta^{\hat{T}}$.

2 Equivalence between allocation and perturbation approach

2.1 Period-1 wedges

Changing the variables of integration by letting $y_1 = y_1(\theta_1)$, and noting that, when earnings are monotone in productivities, as assumed in the literature, $H_Y(y_1(\theta_1)) = F_1(\theta_1)$, we have that the formula for the optimal first-period marginal tax rates in Proposition 3 in the paper can be rewritten as

$$\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))} = \frac{1-F_1(\theta_1)}{y_1(\theta_1)\hat{h}_Y(y_1(\theta_1))}\frac{1}{\hat{E}_1(y_1(\theta_1))}\left[\frac{1}{1+\frac{\delta\frac{\partial}{\partial y_1}\int \mathcal{T}_2(y_1(\theta_1),y_2)dH_O(y_2|y_1(\theta_1))}{\tau_1(y_1(\theta_1))}}\right]. \tag{26}$$

Next, use the that fact that $\widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1) = y_1(\theta_1)$ to rewrite $\hat{E}_1(y_1(\theta_1))$ as

$$\hat{E}_1(y_1(\theta_1)) = \frac{1-\tau_1(y_1(\theta_1))}{y_1(\theta_1)}\frac{\partial\widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1)}{\partial(1-\tau_1)}.$$

Then use the fact that $\widehat{y}_1(1-\tau_1,\theta_1)$ is implicitly defined by $1-\tau_1 = \Gamma(y_1,\theta_1)$ to note that

$$\frac{\partial\widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1)}{\partial(1-\tau_1)} = \frac{1}{\Gamma_y(\widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1),\theta_1)} = \frac{1}{\Gamma_y(y_1(\theta_1),\theta_1)}$$

and hence

$$\hat{E}_1(y_1(\theta_1)) = \frac{\Gamma(y_1(\theta_1),\theta_1)}{y_1(\theta_1)\Gamma_y(y_1(\theta_1),\theta_1)}. \tag{27}$$

Now use the fact that, for any (τ_1,θ_1), $\hat{H}_Y(\widehat{y}_1(1-\tau_1,\theta_1)) = F_1(\theta_1)$ to write

$$\hat{h}_Y(\widehat{y}_1(1-\tau_1,\theta_1))\frac{\partial\widehat{y}_1(1-\tau_1,\theta_1)}{\partial\theta_1} = f_1(\theta_1). \tag{28}$$

When evaluated at $\tau_1 = \tau_1(y_1(\theta_1))$, Condition (28) implies that

$$\hat{h}_Y(y_1(\theta_1)) = \frac{f_1(\theta_1)}{\frac{\partial \widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1)}{\partial \theta_1}}.$$

Now use again the fact that $\widehat{y}_1(1-\tau_1,\theta_1)$ is implicitly defined by $1-\tau_1 = \Gamma(y_1,\theta_1)$ to observe that

$$\frac{\partial \widehat{y}_1(1-\tau_1,\theta_1)}{\partial \theta_1} = -\frac{\Gamma_\theta(\widehat{y}_1(1-\tau_1,\theta_1),\theta_1)}{\Gamma_y(\widehat{y}_1(1-\tau_1,\theta_1),\theta_1)}.$$

Because $\widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1) = y_1(\theta_1)$, this means that

$$\hat{h}_Y(y_1(\theta_1)) = \frac{f_1(\theta_1)}{\frac{\partial \widehat{y}_1(1-\tau_1(y_1(\theta_1)),\theta_1)}{\partial \theta_1}} = -\frac{f_1(\theta_1)\Gamma_y(y_1(\theta_1),\theta_1)}{\Gamma_\theta(y_1(\theta_1),\theta_1)}. \tag{29}$$

Combining (27) with (29), we thus have that, under the optimal tax code,

$$\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))} = \frac{1}{\theta_1\gamma_1(\theta_1)}\left[\frac{\left(-\theta_1\frac{\Gamma_\theta(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1)}\right)}{1+\frac{\delta\frac{\partial}{\partial y_1}\int \mathcal{T}_2(y_1(\theta_1),y_2)dH_O(y_2|y_1(\theta_1))}{\tau_1(y_1(\theta_1))}}\right].$$

Next, observe that, in the iso-elastic case,

$$-\frac{\theta_1\psi_{y\theta}(y_1,\theta_1)}{\psi_y(y_1,\theta_1)} = 1+\phi = -\frac{\theta_2\psi_\theta(y_2,\theta_2)}{\psi(y_2,\theta_2)}. \tag{30}$$

Hence, using the definition of the Γ function in the main text, we have that, under this specification,

$$-\theta_1\Gamma_\theta(y_1(\theta_1),\theta_1) =$$

$$(1{+}\phi)\left[\frac{\psi_y(y_1(\theta_1),\theta_1)}{-\frac{\psi_y(y_1(\theta_1),\theta_1)}{\psi_{y\theta}(y_1(\theta_1),\theta_1)}\delta\frac{\partial^2}{\partial\theta_1\partial y_1}\int\left[y_2(\theta)-\mathcal{T}_2(y_1(\theta_1),y_2(\theta))-\psi(y_2(\theta),\theta_2)\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))}\right] =$$

$$(1+\phi)\left[\psi_y(y_1(\theta_1),\theta_1)+\frac{\psi_y(y_1(\theta_1),\theta_1)}{\psi_{y\theta}(y_1(\theta_1),\theta_1)}\delta\rho\frac{\partial}{\partial y_1}\int\left[\psi_\theta(y_2(\theta),\theta_2)\frac{\theta_2}{\theta_1}\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\right]=$$

$$(1+\phi)\left[\psi_y(y_1(\theta_1),\theta_1)+\frac{\delta\rho}{1+\phi}\frac{\partial}{\partial y_1}\int\left[(1+\phi)\psi(y_2(\theta),\theta_2)\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\right]=$$

$$(1+\phi)\psi_y(y_1(\theta_1),\theta_1)\left[1+\frac{\delta\rho}{\psi_y(y_1(\theta_1),\theta_1)}\frac{\partial}{\partial y_1}\int\left[\psi(y_2(\theta),\theta_2)\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\right].$$

Note that, for the second equality, we used (a) the fact that, given $(\theta_1, y_1(\theta_1))$, for any θ_2, $\bar{y}_2(\theta_2; y_1(\theta_1)) = y_2(\theta)$, along with (b) the fact that, for any θ_2,

$$\frac{\partial}{\partial y_2}\left\{\bar{y}_2(\theta_2;y_1(\theta_1))-\mathcal{T}_2(y_1(\theta_1),\bar{y}_2(\theta_2;y_1(\theta_1)))-\psi(\bar{y}_2(\theta_2;y_1(\theta_1)),\theta_2)\right\}=0,$$

along with the fact that, for any Lipschitz continuous function $g(\theta_2)$,

$$\frac{\partial}{\partial\theta_1}\int g(\theta_2)dF_2(\theta_2|\theta_1,y_1)=\int g'(\theta_2)I_1^2(\theta_1,\theta_2,y_1)dF_2(\theta_2|\theta_1,y_1),$$

where, under the assumed functional form relating period-2 productivity to period-1 productivity,

$$I_1^2(\theta_1,\theta_2,y_1)=\rho\frac{\theta_2}{\theta_1}.$$

For the third equality, we used (30).

Accordingly, we have that, under the optimal tax code,

$$\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))}=$$

$$\frac{(1+\phi)}{\theta_1\gamma_1(\theta_1)}\left[1+\frac{\delta\rho}{\psi_y(y_1(\theta_1),\theta_1)}\frac{\partial}{\partial y_1}\int\left[\psi(y_2(\theta),\theta_2)\right]dF_2(\theta_2|\theta_1,y_1(\theta_1))\right]\times$$

$$\left[\frac{\frac{\psi_y(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1;\tau_2)}}{1+\frac{\delta}{\tau_1(y_1(\theta_1))}\frac{\partial}{\partial y_1}\int\mathcal{T}_2(y_1(\theta_1),y_2)dH_O(y_2|y_1(\theta_1))}\right].$$

Using the fact that, when $y_1 = y_1(\theta_1)$, $\overline{y}_2(\theta_2; y_1(\theta_1)) = y_2(\theta)$ and the fact that[1]

$$\frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2) dH_O(y_2|y_1(\theta_1)) = \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2|\theta_1, y_1(\theta_1)),$$

we thus have that

$$\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))} =$$

$$\frac{(1+\phi)}{\theta_1\gamma_1(\theta_1)} \left[1 + \frac{\delta\rho}{\psi_y(y_1(\theta_1), \theta_1)} \frac{\partial}{\partial y_1} \int \left[\psi(y_2(\theta), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1(\theta_1))\right] \times \tag{31}$$

$$\left[\frac{\frac{\psi_y(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1)}}{1 + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2|\theta_1, y_1(\theta_1))}\right].$$

Now use the formula relating the tax code to the wedges

$$W_1(\theta_1) = \frac{\tau_1(y_1(\theta_1)) + \delta \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2|\theta_1, y_1(\theta_1))}{1 + \delta \frac{\partial}{\partial y_1} \int \left[y_2(\theta) - \psi(y_2(\theta), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1(\theta_1))} \tag{32}$$

along with the fact that $1 - \tau_1(y_1(\theta_1)) = \Gamma(y_1(\theta_1), \theta_1)$ to rewrite the relative wedge as

$$\widehat{W}_1(\theta_1) \equiv \frac{W_1(\theta_1)}{1-W_1(\theta_1)}$$

$$= \left[\frac{\tau_1(y_1(\theta_1))}{1-\tau_1(y_1(\theta_1))}\right] \left[\frac{1 + \frac{\delta}{\tau_1(y_1(\theta_1))} \frac{\partial}{\partial y_1} \int \mathcal{T}_2(y_1(\theta_1), y_2(\theta)) dF_2(\theta_2|\theta_1, y_1(\theta_1))}{\frac{\psi_y(y_1(\theta_1),\theta_1)}{\Gamma(y_1(\theta_1),\theta_1)}}\right]. \tag{33}$$

Replacing the formula in (31) into (33), we then have that, under the optimal tax code, the relative wedge at any productivity level θ_1 is equal to

$$\widehat{W}_1(\theta_1) =$$

$$\frac{(1+\phi)}{\theta_1\gamma_1(\theta_1)} \left[1 + \frac{\delta\rho}{\psi_y(y_1(\theta_1), \theta_1)} \frac{\partial}{\partial y_1} \int \left[\psi(y_2(\theta), \theta_2)\right] dF_2(\theta_2|\theta_1, y_1(\theta_1))\right],$$

[1] Recall that, when computing the derivative of the measure $H_O(y_2|y_1(\theta_1))$ with respect to y_1, one holds $\theta_1(y_1)$ constant.

which is the same formula for the optimal period-1 wedge derived through the allocation approach. Q.E.D.

2.2 Period-2 wedges

Observe that

$$\frac{\partial \tilde{H}_O(y_2|y_1)}{\partial y_1} - \frac{\partial H_O(y_2|y_1)}{\partial y_1} = -\frac{\frac{\partial F_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)}{\partial \theta_1}}{\frac{\partial y_1(\theta_1(y_1))}{\partial \theta_1}} = \frac{\rho\theta_2(y_1,y_2) f_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)}{\theta_1(y_1)\frac{\partial y_1(\theta_1(y_1))}{\partial \theta_1}}. \quad (34)$$

Note that the first equality uses the fact that $H_O(y_2|y_1) = F_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)$, along with the fact that, by definition, $\partial \tilde{H}_O(y_2|y_1)/dy_1$ is computed holding θ_1 fixed at $\theta_1 = \theta_1(y_1)$. The second equality uses the fact that

$$\frac{\partial}{\partial \theta_1}[1 - F_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1)] = I_1^2(\theta_1(y_1),\theta_2(y_1,y_2),y_1) f_2(\theta_2(y_1,y_2)|\theta_1(y_1),y_1),$$

along with the fact that, under the assumed specification of the productivity process,

$$I_1^2(\theta_1,\theta_2,y_1) = \rho\frac{\theta_2}{\theta_1}.$$

Replacing (34) into the formula for the wedge in Proposition 4 in the main text, we have that, at the income history $(y_1(\theta_1), y_2(\theta_1,\theta_2))$ induced under the optimal tax code, the period-2 wedge is equal to

$$\frac{\tau_2(y_1(\theta_1),y_2(\theta_1,\theta_2))}{1-\tau_2(y_1(\theta_1),y_2(\theta_1,\theta_2))} = \frac{\left[\rho\frac{\theta_2 f_2(\theta_2|\theta_1,y_1(\theta_1))}{\theta_1\frac{\partial y_1(\theta_1(y_1))}{\partial \theta_1}}\right][1-H_Y(y_1(\theta_1))]}{h_Y(y_1(\theta_1))y_2(\theta_1,\theta_2)\hat{h}_O(y_2(\theta_1,\theta_2)|y_1(\theta_1))\hat{E}_2(y_1(\theta_1),y_2(\theta_1,\theta_2))}, \quad (35)$$

where we used the fact that $y_1 = y_1(\theta_1)$ and $y_2 = y_2(\theta_1,\theta_2)$.

Next, use the fact that, for any θ_1,

$$H_Y(y_1(\theta_1)) = F_1(\theta_1) \tag{36}$$

to note that

$$h_Y(y_1(\theta_1))\frac{\partial y_1(\theta_1(y_1))}{\partial \theta_1} = f_1(\theta_1) \tag{37}$$

and, similarly, the fact that, for any (θ_1, θ_2),

$$\hat{H}_O(y_2(\theta_1,\theta_2)|y_1(\theta_1)) = \hat{H}_O(\hat{y}_2(1-\tau_2(y_1(\theta_1), y_2(\theta_1,\theta_2)), \theta_2)|y_1(\theta_1)) = F_2(\theta_2|\theta_1, y_1(\theta_1))$$

to note that

$$\hat{h}_O(y_2(\theta_1,\theta_2)|y_1(\theta_1)) = \frac{f_2(\theta_2|\theta_1, y_1(\theta_1))}{\frac{\partial \hat{y}_2(1-\tau_2(y_1(\theta_1), y_2(\theta_1,\theta_2)),\theta_2)}{\partial \theta_2}}. \tag{38}$$

Finally, use the fact that

$$\hat{E}_2(y_1, y_2) = \frac{1-\tau_2(y_1,y_2)}{\widehat{y}_2(1-\tau_2(y_1,y_2), \theta_2(y_1,y_2))}\frac{\partial \widehat{y}_2(1-\tau_2(y_1,y_2), \theta_2(y_1,y_2))}{\partial(1-\tau_2)},$$

along with the fact that $\widehat{y}_2(1-\tau_2, \theta_2)$ is implicitly defined by

$$\psi_y(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2) = 1-\tau_2$$

and hence

$$\frac{\partial \widehat{y}_2(1-\tau_2,\theta_2)}{\partial \theta_2} = -\frac{\psi_{y\theta}(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2)}{\psi_{yy}(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2)}$$

and

$$\frac{\partial \widehat{y}_2(1-\tau_2,\theta_2)}{\partial(1-\tau_2)} = \frac{1}{\psi_{yy}(\widehat{y}_2(1-\tau_2,\theta_2),\theta_2)},$$

to write[2]

$$\hat{E}_2(y_1(\theta_1), y_2(\theta_1, \theta_2)) = -\frac{\psi_y(y_2(\theta_1, \theta_2), \theta_2)}{y_2(\theta_1, \theta_2)\psi_{y\theta}(y_2(\theta_1, \theta_2), \theta_2)} \frac{\partial \widehat{y_2}(1 - \tau_2(y_1(\theta_1), y_2(\theta_1, \theta_2)), \theta_2)}{\partial \theta_2}. \quad (39)$$

Also recall that, under the assumed iso-elastic specification

$$-\theta_2 \frac{\psi_{y\theta}(y_2, \theta_2)}{\psi_y(y_2, \theta_2)} = 1 + \phi. \quad (40)$$

Substituting (36)-(39) into (35), and using (40), we have that

$$\hat{W}_2(\theta_1, \theta_2) = \frac{\tau_2(y_1(\theta_1), y_2(\theta_1, \theta_2))}{1 - \tau_2(y_1(\theta_1), y_2(\theta_1, \theta_2))} = \frac{(1 + \phi)\rho}{\gamma_1(\theta_1)\theta_1},$$

which is the same formula obtained through the direct allocation approach. Q.E.D

[2]Note that we also used the fact that $\hat{y}_2(1 - \tau_2(y_1(\theta_1), y_2(\theta_1, \theta_2)), \theta_2) = y_2(\theta_1, \theta_2)$.