Online Appendix to Optimal Income Tax Deductions for Mixed Business and Personal Expenditures

Jacob Goldin Sebastian Koehne Nicholas Lawson

December 11, 2025

A Connection to Prior Literature

To clarify the relationship between our setting and the models that prior work has considered, we can define a virtual skill parameter

$$\widetilde{w} := \phi(w, d, l) := \frac{y(wl, d)}{l}, \tag{1}$$

which yields an income process that appears multiplicative in labor effort: $y(wl, d) = \widetilde{w}l$. In this transformation, the virtual skill level $\widetilde{w}$ depends on the exogenous parameter w, labor effort l and consumption of the hybrid good d. In contrast, prior research in optimal taxation, including the work studying optimal commodity taxation with preference non-separability, generally takes skill to be exogenous (e.g., Atkinson and Stiglitz, 1976; Mirrlees, 1976; Christiansen, 1984).

As highlighted by the transformation in Equation (1), mixed-purpose goods can be reinterpreted as a non-standard model of endogenous skill formation. In this regard, our framework shares some similarities with models of optimal human capital policy. However, those models typically abstract from a consumption value of skills (e.g., Bohacek and Kapička, 2008; Findeisen and Sachs, 2016; Stantcheva, 2017) and often focus on specific skill formation processes, such as the multiplicative case $\widetilde{w} = w \cdot \phi(d)$ explored by Bovenberg and Jacobs (2005).

Alternatively, one could define taxpayers' preferences over the consumption goods (c and d) and pre-tax income (y), rather than over the consumption goods and labor effort (l). Viewed through this lens, we can interpret the presence of a mixed-purpose good as giving rise to a special form of preference non-separability, in that an individual's (dis)utility associated with obtaining a given level of pre-tax income depends on the individual's choice of d. For this interpretation, we invert the income-generating process and express labor effort as $l(d, y, w) = g(y, d)/w$, where $g(y, d)$ is the inverse of $y(x, d)$ with respect to x for any given level of d. Then, we can express utilities as

$$\widetilde{U}(c, d, y, w) = u\left(v(c, d), \frac{g(y, d)}{w}, w\right). \tag{2}$$

This interpretation shifts the effect of the mixed-purpose good d from the budget constraint to preferences. A model that is general enough to accommodate this approach is Koehne and Sachs (2022), if one is willing to impose the additional restriction that utility takes the additive form $v(c, d) + h(y, d, w)$.[1] In addition to the difference in assumptions, that paper focuses on time-saving household services rather than mixed-

[1] Gauthier and Laroque (2009) examine a related framework of individual production sets, where cost functions take the separable form $h(H(y, d), w)$, and derive a generalized Atkinson–Stiglitz result within that setting.

purpose goods, and they do not derive any of our results. More generally, when mixed-purpose goods are incorporated into the model via preferences rather than the income-generating function, preferences obtain a specific form of skill dependence, which tends to limit the transparency and interpretability of the results. In contrast, our representation of mixed-purpose goods in terms of the income-generating function yields a particularly simple and intuitive condition for Pareto optimality.

To illustrate how our approach facilitates the optimal policy analysis relative to the preference-based approach, we apply our Pareto condition (11) to the utility specification in (2). Differentiating the identity $g\left(y\left(x,d\right),d\right)=x$, we obtain $y_d=-g_d/g_y$. Setting $p=1$ as in the model of Koehne and Sachs (2022), our condition becomes

$$1=\frac{v_d}{v_c}-\frac{g_d}{g_y}. \tag{3}$$

By contrast, the Pareto optimality condition derived by Koehne and Sachs (2022) involves cross-partials of the utility function and takes the form

$$\frac{\widetilde{U}_c+\widetilde{U}_y}{\widetilde{U}_{wy}}=\frac{\widetilde{U}_c-\widetilde{U}_d}{-\widetilde{U}_{wd}}. \tag{4}$$

In what follows, we demonstrate that our condition (3) implies the Koehne and Sachs (2022) condition (4).

Under the additive specification of Koehne and Sachs (2022), we have $u_v=1$ and $u_{vl}=u_{vw}=0$. Therefore, the first derivatives of the utility function (2) are given by

$$\begin{aligned}\widetilde{U}_c&=v_c,\\ \widetilde{U}_d&=v_d+\frac{1}{w}u_l g_d,\\ \widetilde{U}_y&=\frac{1}{w}u_l g_y,\end{aligned}$$

and the cross-partials with respect to skill w are

$$\begin{aligned}\widetilde{U}_{wy}&=g_y\cdot\left(-\frac{1}{w^2}u_l-\frac{g}{w^3}u_{ll}+\frac{1}{w}u_{lw}\right),\\ \widetilde{U}_{wd}&=g_d\cdot\left(-\frac{1}{w^2}u_l-\frac{g}{w^3}u_{ll}+\frac{1}{w}u_{lw}\right).\end{aligned}$$

Hence, the left-hand side of (4) can be written as

$$\frac{\widetilde{U}_c+\widetilde{U}_y}{\widetilde{U}_{wy}}=\frac{v_c+\frac{1}{w}u_l g_y}{g_y\cdot\left(-\frac{1}{w^2}u_l-\frac{g}{w^3}u_{ll}+\frac{1}{w}u_{lw}\right)}=-\frac{w^2v_c+wu_l g_y}{g_y\cdot\left(u_l+\frac{g}{w}u_{ll}-wu_{lw}\right)}.$$

The right-hand side of (4) can be written as

$$\frac{\widetilde{U}_c - \widetilde{U}_d}{-\widetilde{U}_{wd}} = -\frac{v_c - v_d - \frac{1}{w}u_l g_d}{g_d \cdot \left(-\frac{1}{w^2}u_l - \frac{g}{w^3}u_{ll} + \frac{1}{w}u_{lw}\right)} = \frac{w^2 v_c - w^2 v_d - w u_l g_d}{g_d \cdot \left(u_l + \frac{g}{w}u_{ll} - w u_{lw}\right)}.$$

After cross multiplication, we note that both sides of Equation (4) are indeed equal under the Pareto condition (3).

In summary, our approach recovers the preference-based optimality condition (4) as a special case. In addition, it provides a more transparent economic interpretation of Pareto optimality in environments with work-related goods, avoiding cross-partial utility terms involving unobservable skill.

B Mechanism Design Approach

This appendix offers an alternative derivation of Equation 7 using a mechanism design framework. The analysis proceeds in two steps. First, we characterize an optimal allocation rule under informational constraints. Second, we examine how this rule restricts the design of possible decentralizations.

B.1 Allocation constraints

An allocation $(c(w), d(w), x(w))_{w\in\mathcal{W}}$ is *feasible* if it satisfies the following resource constraint:

$$\int_{\underline{w}}^{\overline{w}} \left[c\left(w\right) + p \cdot d\left(w\right) - y\left(x\left(w\right), d\left(w\right)\right)\right] f\left(w\right) \mathrm{d}w \leq 0. \tag{5}$$

Assuming that the variables c, d, y and $x = wl$ are observable,[2] while skills w and labor supply l are not, incentive compatibility requires that individuals do not gain from choosing the commodity bundle (c, d) and labor income x of any other type. Thus, allocation $(c(w), d(w), x(w))_{w\in\mathcal{W}}$ is *incentive compatible* if it satisfies

$$u\left(v\left(c\left(w\right), d\left(w\right)\right), \frac{x\left(w\right)}{w}, w\right) \geq u\left(v\left(c\left(w'\right), d\left(w'\right)\right), \frac{x\left(w'\right)}{w}, w\right) \quad \text{for all } w, w' \in \mathcal{W}. \tag{6}$$

Importantly, due to the weak separability of the utility function, the incentive compatibility condition (6) depends on commodity bundles $(c\left(w\right), d\left(w\right))$ only via the subutilities $v\left(c\left(w\right), d\left(w\right)\right)$. Thus, any reform that leaves the subutilities unchanged will not affect the incentives to report truthfully.

[2] If d is observable, then knowledge of the income production function $y(x, d)$ allows one to infer labor income $x = wl$ based on total income y, or vice-versa. Therefore, it is sufficient to assume that at least one of the two income variables is observable.

B.2 Pareto optimal allocations

An allocation is *incentive feasible* if it is feasible and incentive compatible. An allocation is *constrained Pareto optimal* if it is incentive feasible and not Pareto dominated by any other incentive feasible allocation.

To support our resource minimization approach to deriving the optimal allocation rule, we assume throughout the analysis that the resource constraint (5) is binding in every constrained Pareto optimum. Intuitively, we are assuming that if resources are left over, the planner can design a reform that increases consumption without violating the incentive compatibility condition, such as through a lump-sum transfer. This assumption holds in many benchmark cases, including additively and multiplicatively separable utility functions.

B.3 First-Best Rule

Having established that the resource constraint binds, we now demonstrate that constrained Pareto optima satisfy a first-best rule for commodity allocation.

We consider two-dimensional allocation perturbations that change regular consumption c and the hybrid good d, while holding fixed the subutility derived from these commodities. For a given allocation $(c(w), d(w), x(w))_{w\in\mathcal{W}}$, we pick arbitrary types $w_0, w_1 \in \mathcal{W}$ with $w_0 < w_1$ and an arbitrary function $\delta : [w_0, w_1] \to \mathbb{R}$ with $\delta(w_0) = \delta(w_1) = 0$. Using the scaling factor $\varepsilon \in \mathbb{R}$, we perturb the commodity levels as follows:

$$\hat{d}(w) = d(w) + \varepsilon \cdot \delta(w), \tag{7}$$

$$\hat{c}(w) = c(w) - \psi(w, \varepsilon \cdot \delta(w)), \tag{8}$$

where $\psi(w, \varepsilon \cdot \delta(w))$ is defined such that the subutility remains unchanged,[3] i.e.,

$$v(c(w) - \psi(w, \varepsilon \cdot \delta(w)), d(w) + \varepsilon \cdot \delta(w)) = v(c(w), d(w)) =: \bar{v}(w). \tag{9}$$

Equation (9) ensures

$$v\left(\hat{c}(w), \hat{d}(w)\right) = v(c(w), d(w)) \text{ for all } w \in \mathcal{W}. \tag{10}$$

[3]By inverting this condition, we can state the reduction of regular consumption alternatively as

$$\psi(w, \varepsilon \cdot \delta(w)) = c(w) - v^{-1}(\bar{v}(w); d(w) + \varepsilon \cdot \delta(w)),$$

where v^{-1} is the inverse of subutility v with respect to its first argument.

Therefore, if the original allocation $(c(w), d(w), x(w))_{w\in\mathcal{W}}$ is incentive compatible, so is the perturbed allocation $\left(\hat{c}(w), \hat{d}(w), x(w)\right)_{w\in\mathcal{W}}$.

A necessary condition for Pareto optimality is that no perturbation can reduce the resources needed for a given pattern of subutilities. Because the incentive structure is unaffected by such perturbations, this insight holds in the first-best world as well as in the second-best world of constrained optimality. The first-order condition for resource minimization is given in the following result.

Proposition 1 (First-Best Rule)**.** *If an allocation* $(c(w), d(w), x(w))_{w\in\mathcal{W}}$ *is constrained Pareto optimal, commodity choices satisfy the first-best rule*

$$p = \frac{v_d\left(c\left(w\right), d\left(w\right)\right)}{v_c\left(c\left(w\right), d\left(w\right)\right)} + y_d\left(x\left(w\right), d\left(w\right)\right) \tag{11}$$

for almost all skill types w*.*

Proof. First, we show by contradiction that constrained Pareto optimality implies resource minimization for a given pattern of subutilities $v(c, d)$ and labor incomes x. Suppose, to the contrary, that a given allocation is constrained Pareto optimal and does not minimize the resources for the given pattern of subutilities and labor incomes. Then, there exists an alternative allocation with identical subutilities and labor incomes that requires fewer resources. Since the subutilities and labor supplies are the same as in the original allocation, the alternative allocation is also incentive compatible. It also implements the same utility levels, so it is also constrained Pareto optimal. However, since the original allocation was feasible, the resource constraint for the alternative allocation must be slack, which contradicts our assumption that the resource constraint is binding in every constrained Pareto optimum.

Now, we derive the first-order condition for resource minimization. (This step provides a direct proof of a basic version of the Euler-Lagrange equation.) If the given allocation minimizes the resources needed for a given pattern of subutilities, then $\varepsilon = 0$ is a solution of the following resource problem:

$$\min_{\varepsilon} \int_{w_0}^{w_1} \left[p\left(d\left(w\right) + \varepsilon\delta\left(w\right)\right) + c\left(w\right) - \psi\left(w, \varepsilon\delta\left(w\right)\right) - y\left(x\left(w\right), d\left(w\right) + \varepsilon\delta\left(w\right)\right)\right] f\left(w\right) \mathrm{d}w.$$

The FOC evaluated at $\varepsilon = 0$ is

$$\int_{w_0}^{w_1} \left[p\delta\left(w\right) - \delta\left(w\right)\psi_2\left(w, 0\right) - \delta\left(w\right) y_d\left(x\left(w\right), d\left(w\right)\right)\right] f\left(w\right) \mathrm{d}w = 0,$$

where ψ_2 denotes the partial derivative of ψ with respect to its second argument. Equivalently,

$$\int_{w_0}^{w_1} \left[p - \psi_2\left(w, 0\right) - y_d\left(x\left(w\right), d\left(w\right)\right)\right] \delta\left(w\right) f\left(w\right) \mathrm{d}w = 0.$$

Importantly, the perturbation δ was arbitrary and the FOC is valid for every perturbation δ. Therefore, we must have

$$p - \psi_2\left(w, 0\right) - y_d\left(x\left(w\right), d\left(w\right)\right) = 0 \text{ for almost all } w \in \left[w_0, w_1\right].$$

Differentiating Equation (9) with respect to $\varepsilon \cdot \delta\left(w\right)$, we obtain

$$-\psi_2\left(w, \varepsilon \cdot \delta\left(w\right)\right) v_c\left(\hat{c}\left(w\right), \hat{d}\left(w\right)\right) + v_d\left(\hat{c}\left(w\right), \hat{d}\left(w\right)\right) = 0.$$

Hence,

$$\psi_2\left(w, 0\right) = \frac{v_d\left(c\left(w\right), d\left(w\right)\right)}{v_c\left(c\left(w\right), d\left(w\right)\right)}.$$

Hence, we can express the optimality condition as

$$p - \frac{v_d\left(c\left(w\right), d\left(w\right)\right)}{v_c\left(c\left(w\right), d\left(w\right)\right)} - y_d\left(x\left(w\right), d\left(w\right)\right) = 0 \text{ for almost all } w \in \left[w_0, w_1\right].$$

Because w_0 and w_1 were arbitrary, the condition extends to the entire type space $\mathcal{W}$. □

The first-best rule derived above is intuitive. Since the choice of goods is separated from the supply of labor by the weak separability of preferences, the trade-off between the hybrid good and regular consumption should not be distorted. This requires that the relative price of the hybrid good, p, corrected for its marginal income benefit, y_d, equals the marginal rate of substitution between these goods, v_d/v_c. Importantly, the correction of the relative price reflects the *gross* marginal income benefit of the hybrid good, y_d. Thus, to the extent that income is taxed, the tax code must provide compensating measures to support the demand for the hybrid good.

B.4 Consequence for the Optimal Deduction Rate

As discussed above, the first-best rule for commodity choice is compatible with income taxation only if the tax system subsidizes the hybrid good relative to general consumption. We now demonstrate how this requirement constrains the design of a tax system with a simple deduction schedule.

Suppose that a constrained Pareto optimal allocation $(c(w), d(w), x(w))_{w \in \mathcal{W}}$ is implemented by a tax function of the form

$$T\left(y - \alpha\left(x\right) pd + \psi\left(x\right)\right),$$

where $x = wl$ represents labor income. By rearranging Equation 5, the first-order condition for the individual commodity choice problem equals

$$y_d\left(x(w), d(w)\right) + \frac{v_d\left(c(w), d(w)\right)}{v_c\left(c(w), d(w)\right)} = p + \left(y_d\left(x(w), d(w)\right) - \alpha\left(x(w)\right) p\right) T'. \tag{12}$$

Equating this condition with the first-best rule, we obtain

$$p = y_d\left(x(w), d(w)\right) + \frac{v_d\left(c(w), d(w)\right)}{v_c\left(c(w), d(w)\right)} = p + \left(y_d\left(x(w), d(w)\right) - \alpha\left(x(w)\right) p\right) T'. \tag{13}$$

Hence, if $T' \neq 0$, we obtain $y_d - \alpha p = 0$. Solving this condition for α, we obtain Equation 7.

C Supplementary Proofs

C.1 Proof of Proposition 2

This appendix presents the proof of Proposition 2 for the optimal constant deduction rate α. Similar to the proof of Proposition 1, we adapt the tax reform approach of Kaplow (2006, 2012, 2020) to the case of marginal deduction reforms.

Let v^* denote the indirect subutility over c and d that individuals obtain as a function of the tax system and a given choice of labor income $x = wl$:

$$v^*\left(wl; T, \alpha, \gamma\left(wl\right)\right) = \max_{c,d} v\left(c, d\right) \text{ s.t. } c + pd = y\left(wl, d\right) - T\left(y\left(wl, d\right) - \alpha pd + \gamma\left(wl\right)\right).$$

Consider a small increase in the deduction rate α by an amount $\Delta\alpha$. By the envelope theorem, the increased deduction rate changes individual subutilities by

$$v_c \cdot p \cdot d \cdot T' \cdot \Delta\alpha.$$

To compensate for this welfare effect, we adjust the component γ in the tax function. By the envelope theorem, increasing $\gamma\left(wl\right)$ by $\Delta\gamma\left(wl\right)$ reduces individual subutility by $v_c \cdot T' \cdot \Delta\gamma\left(wl\right)$. Thus, to neutralize

the impact of the deduction perturbation on subutility, we set

$$\Delta\gamma\left(wl\right)=p\cdot d\cdot\Delta\alpha.$$

for each level of labor income wl. By construction, this combined reform leaves all individual subutilities unchanged.

The next step is to show that the reform is also neutral with respect to labor effort, and therefore neutral with respect to individual welfare $U^{*}\left(w;T,\alpha,\gamma\right)$. Prior to the reform, an individual of type w chooses the labor effort l to maximize $u\left(v^{*}\left(wl;T,\alpha,\gamma\left(wl\right)\right),l\right)$. Since

$$v^{*}\left(wl;T,\alpha,\gamma\left(wl\right)\right)=v^{*}\left(wl;T,\alpha+\Delta\alpha,\gamma\left(wl\right)+\Delta\gamma\left(wl\right)\right)$$

for every possible value of wl, the utility associated with each labor effort option is the same before and after the reform. Therefore, the individual's labor effort choice remains unchanged. Consequently, each individual's overall utility U^{*} is also unaffected.

Since individual welfare is unchanged by the reform, a necessary condition for the Pareto optimality of the original tax system (T,α,γ) is that it maximizes tax revenue; otherwise welfare could be improved by implementing a welfare-neutral but revenue-increasing reform and distributing the additional revenue to individuals lump-sum. Tax revenue after the reform is given by

$$\int_{\underline{w}}^{\overline{w}}T\left(y\left(x\left(w\right),\hat{d}\left(x(w);\Delta\alpha\right)\right)-\left(\alpha+\Delta\alpha\right)p\hat{d}\left(x(w);\Delta\alpha\right)+\gamma\left(x(w)\right)+\Delta\gamma\left(x(w)\right)\right)\mathrm{d}F\left(w\right),$$

where $\hat{d}\left(x(w);\Delta\alpha\right)$ denotes the demand for the hybrid good under the reform $(\Delta\alpha,\Delta\gamma)=(\Delta\alpha,pd\Delta\alpha)$. Differentiating this expression with respect to $\Delta\alpha$, and evaluating at $\Delta\alpha=0$, yields the following necessary condition:

$$\int_{\underline{w}}^{\overline{w}}\left(y_{d}\frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha}-pd-\alpha p\frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha}+pd\right)T'\,\mathrm{d}F\left(w\right)=0,$$

where $\mathrm{d}\hat{d}/\mathrm{d}\alpha$ denotes the the total derivative of demand for the hybrid good with respect to the combined reform $(\Delta\alpha,\Delta\gamma)=(\Delta\alpha,pd\Delta\alpha)$ and we have used the result that labor effort remains unchanged. Equivalently,

$$\int_{\underline{w}}^{\overline{w}}\left[\left(y_{d}-\alpha p\right)\frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha}T'\right]\mathrm{d}F\left(w\right)=0.$$

Equivalently,

$$\int_{\underline{w}}^{\overline{w}} \frac{y_d}{p} \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} T' \,\mathrm{d}F(w) - \alpha \int_{\underline{w}}^{\overline{w}} \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} T' \,\mathrm{d}F(w) = 0.$$

Equivalently,

$$\alpha = \frac{\int_{\underline{w}}^{\overline{w}} \frac{y_d}{p} \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} T' \,\mathrm{d}F(w)}{\int_{\underline{w}}^{\overline{w}} \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} T' \,\mathrm{d}F(w)} = \int_{\underline{w}}^{\overline{w}} \frac{y_d}{p} \frac{\frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} T'}{\int_{\underline{w}}^{\overline{w}} \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} T' \,\mathrm{d}F(w)} \,\mathrm{d}F(w).$$

Relabelling the terms yields the result in Proposition 2. □

C.2 Proof of Proposition 3

This appendix presents the proof of Proposition 3 for the expenditure-based deduction rate $\alpha(pd)$.

As shown in section 4.2, the allocation $(c(w), d(w), x(w))_{w \in \mathcal{W}}$ induced by the optimal labor-income-based deduction rule satisfies the first-order condition of the commodity choice problem under the expenditure-based deduction rule $\alpha(pd)$ if and only if Equation 28 holds for every $w \in \mathcal{W}$. Setting $e(w) := pd(w)$ and noting $e'(w) = pd'(w)$, Equation 28 can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}w} \left[e(w)\alpha(e(w)) \right] = y_d(x(w), d(w)) d'(w). \tag{14}$$

Integrating both sides over w and solving for α yields

$$\alpha(pd(w)) = \frac{\int y_d(x(w), d(w)) d'(w) \,\mathrm{d}w}{pd(w)}, \tag{15}$$

which establishes Equation 29. □

C.3 Proof of Proposition 4

This appendix presents the proof of Proposition 4 on the concavity of the commodity choice problems under the labor-income-based deduction rule (7) and the expenditure-based rule (29).

Given the concavity and monotonicity of the subutility function, a sufficient condition for the strict concavity of these problems is that disposable income $y - T$ varies strictly concavely with respect to the hybrid good d, holding labor income x fixed.

We begin with the expenditure-based rule (29). For notational convenience, we define the deduction level as

$$\mathcal{D}(pd) := \alpha(pd)\, pd.$$

The individual's disposable income can then be written as:

$$k\left(d\right) := y\left(x,d\right) - T\left(y\left(x,d\right) - \mathcal{D}\left(pd\right) + \gamma\left(x\right)\right).$$

The derivative of $k(d)$ with respect to d is given by:

$$k'\left(d\right) = y_d(x,d) - \left(y_d(x,d) - \mathcal{D}'(pd)\,p\right)T'\left(y(x,d) - \mathcal{D}(pd) + \gamma(x)\right).$$

Differentiating again, we obtain

$$k''\left(d\right) = y_{d,d}(x,d) - \left(y_{d,d}(x,d) - \mathcal{D}''(pd)\,p^2\right)T' - \left(y_d(x,d) - \mathcal{D}'(pd)\,p\right)^2 T'',$$

which can be rearranged as

$$k''\left(d\right) = \left(1 - T'\right)y_{d,d}(x,d) + \mathcal{D}''(pd)\,p^2T' - \left(y_d(x,d) - \mathcal{D}'(pd)\,p\right)^2 T''.$$

By assumptions (i) and (ii), we have $1 - T' > 0$ and $T'' \geq 0$. Moreover, we have $y_{d,d} < 0$ by assumption (iii). Hence,

$$k''\left(d\right) < \mathcal{D}''(pd)\,p^2T'.$$

Using $p^2 > 0$ and $T' > 0$, a sufficient condition for the strict concavity of k is that $\mathcal{D}$ is concave in pd, i.e., $\mathcal{D}''(pd) \leq 0$ for every d.

Differentiating the definition of the deduction rule (29) with respect to type w yields

$$\mathcal{D}'(pd(w)) = \frac{1}{p}y_d(x(w),d(w)),$$

and

$$\mathcal{D}''(pd(w)) = \frac{y_{d,d}(x(w),d(w)) + y_{x,d}(x(w),d(w))\frac{x'(w)}{d'(w)}}{p^2}.$$

Since $y_{d,d} < 0$ by assumption (iii), we conclude that $\mathcal{D}'' \leq 0$ holds if the cross-partial term $y_{x,d}\frac{x'}{d'}$ is not too large. In particular, $\mathcal{D}'' \leq 0$ holds if $y_{x,d} = 0$, or if $y_{x,d} \leq 0$ and both $x', d' \geq 0$.

Finally, consider the labor-income-based deduction rule (7). Holding x fixed, the deduction level takes the form $\mathcal{D}(pd) := \alpha(x)pd$, which is linear in d. Consequently, $\mathcal{D}''(pd) = 0$ for all (d,x), and the term involving $\mathcal{D}''$ drops out of the second derivative. Since the remaining terms in $k''(d)$ are negative under assumptions

(i), (ii), (iii), the objective remains strictly concave in d. □

C.4 Proof of Proposition 5

This appendix presents the proof of Proposition 5, which characterizes the optimal uniform deduction rate under a fixed income tax schedule. To keep the reform budget neutral, we pair the change in α with a change in the lump-sum demogrant b, which we denote by $\frac{\partial b}{\partial \alpha}$.

By the envelope theorem, the total derivative of the indirect utility function with respect to α equals:

$$u_v \left(\frac{\partial v^*}{\partial \alpha} + \frac{\partial v^*}{\partial b} \frac{\partial b}{\partial \alpha} \right) = U_c \cdot \left(pdT' + \frac{\partial b}{\partial \alpha} \right). \tag{16}$$

Hence, the government's first-order condition for welfare maximization with respect to α is:

$$\frac{\mathrm{d}W}{\mathrm{d}\alpha} = \int_{\underline{w}}^{\overline{w}} \pi \cdot U_c \cdot \left(pdT' + \frac{\partial b}{\partial \alpha} \right) \mathrm{d}F(w) = 0. \tag{17}$$

The government budget constraint is given by

$$b = R - G, \tag{18}$$

where R denotes the revenue raised by the income tax (not including the demogrant b),

$$R = \int_{\underline{w}}^{\overline{w}} T\left(y(wl, d) - \alpha pd + \gamma(wl) \right) \mathrm{d}F(w).$$

Below, we will use the facts that

$$R_\alpha := \frac{\partial R}{\partial \alpha} = \mathbb{E}\left[T' \cdot \left(\frac{\partial l}{\partial \alpha} (\gamma' + y_x) w + \frac{\partial d}{\partial \alpha} (y_d - \alpha p) - pd \right) \right], \tag{19}$$

and

$$R_b := \frac{\partial R}{\partial b} = \mathbb{E}\left[T' \cdot \left(\frac{\partial l}{\partial b} (\gamma' + y_x) w + \frac{\partial d}{\partial b} (y_d - \alpha p) \right) \right]. \tag{20}$$

It will also be helpful to define h as the change in the tax revenue collected from an individual in response to the individual having an additional dollar of income:

$$h = T' \cdot \left(\frac{\partial l}{\partial b} (\gamma' + y_x) w + \frac{\partial d}{\partial b} (y_d - \alpha p) \right), \tag{21}$$

so that $R_b = \mathbb{E}[h]$.

Totally differentiating the government budget constraint with respect to α, we obtain

$$\frac{\partial b}{\partial \alpha} = R_\alpha + R_b \frac{\partial b}{\partial \alpha}, \tag{22}$$

This in turn implies the identity

$$\frac{\partial b}{\partial \alpha} = \frac{R_\alpha}{1 - R_b}. \tag{23}$$

Substituting these results into the government's FOC and using expectations notation we have:

$$0 = \mathbb{E}\left[\pi \cdot U_c \cdot \left(pdT' + \frac{\partial b}{\partial \alpha}\right)\right] \tag{24}$$

$$= \mathbb{E}[\pi \cdot U_c \cdot pdT'] + \frac{\mathbb{E}[\pi \cdot U_c]}{1 - R_b}\mathbb{E}\left[T' \cdot \left(\frac{\partial l}{\partial \alpha}(\gamma' + y_x)w + \frac{\partial d}{\partial \alpha}(y_d - \alpha p) - pd\right)\right] \tag{25}$$

$$= \mathbb{E}[\pi \cdot U_c \cdot pdT'] + \frac{\mathbb{E}[\pi \cdot U_c]}{1 - R_b}\mathbb{E}\left[T' \cdot \left(\frac{\partial l}{\partial \alpha}(\gamma' + y_x)w + \frac{\partial d}{\partial \alpha}(y_d - \alpha p)\right)\right] - \frac{\mathbb{E}[\pi \cdot U_c]}{1 - R_b}\mathbb{E}[pdT'] \tag{26}$$

Because a marginal increase in α raises income by pdT', a Slutsky decomposition allows us to write:

$$\frac{\partial l}{\partial \alpha} = \frac{\partial l^c}{\partial \alpha} + pdT'\frac{\partial l}{\partial b} \tag{27}$$

where $\frac{\partial l^c}{\partial \alpha}$ denotes the compensated (Hicksian) derivative, and similarly for d:

$$\frac{\partial d}{\partial \alpha} = \frac{\partial d^c}{\partial \alpha} + pdT'\frac{\partial d}{\partial b}. \tag{28}$$

Substituting these results into the government FOC yields:

$$
\begin{aligned}
0 =& \mathbb{E}\left[\pi \cdot U_c \cdot pdT'\right] + \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[T' \cdot \left(\frac{\partial l}{\partial \alpha}\left(\gamma' + y_x\right)w + \frac{\partial d}{\partial \alpha}\left(y_d - \alpha p\right)\right)\right] - \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[pdT'\right] \\
=& \mathbb{E}\left[\pi \cdot U_c \cdot pdT'\right] + \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[T' \cdot \left(\frac{\partial l^c}{\partial \alpha}\left(\gamma' + y_x\right)w + \frac{\partial d^c}{\partial \alpha}\left(y_d - \alpha p\right)\right)\right] \\
&+ \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[pdT' \cdot T' \cdot \left(\frac{\partial l}{\partial b}\left(\gamma' + y_x\right)w + \frac{\partial d}{\partial b}\left(y_d - \alpha p\right)\right)\right] - \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[pdT'\right] \\
=& \mathbb{E}\left[\pi \cdot U_c \cdot pdT'\right] + \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[T' \cdot \left(\frac{\partial l^c}{\partial \alpha}\left(\gamma' + y_x\right)w + \frac{\partial d^c}{\partial \alpha}\left(y_d - \alpha p\right)\right)\right] \\
&+ \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[pdT' \cdot h\right] - \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[pdT'\right] \\
=& \mathbb{E}\left[\left(\pi\, U_c + h\,\frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\right) \cdot pdT'\right] - \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[pdT'\right] \\
&+ \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[T' \cdot \left(\frac{\partial l^c}{\partial \alpha}\left(\gamma' + y_x\right)w + \frac{\partial d^c}{\partial \alpha}\left(y_d - \alpha p\right)\right)\right].
\end{aligned}
\tag{29}
$$

We use the fact that $\mathbb{E}\left[h\right] = R_b$ implies

$$
\mathbb{E}\left[\pi\, U_c + h\,\frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\right] = \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}, \tag{30}
$$

and rewrite the final condition in (29) as

$$
0 = \operatorname{cov}\left(pdT',\, \pi U_c + h\,\frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\right) + \frac{\mathbb{E}\left[\pi \cdot U_c\right]}{1-R_b}\mathbb{E}\left[T' \cdot \left(\frac{\partial l^c}{\partial \alpha}\left(\gamma' + y_x\right)w + \frac{\partial d^c}{\partial \alpha}\left(y_d - \alpha p\right)\right)\right]. \tag{31}
$$

Dividing this equation by $\frac{\mathbb{E}[\pi \cdot U_c]}{1-R_b}$ and then solving for α yields the result. □

C.5 Proof of Proposition 6

If the tax schedule is optimal, no budget-neutral change to $\gamma(x)$ can be welfare-improving. So our first step will be to repeat the argument used in the proof of Proposition 5 with respect to α, now applied to $\gamma(x)$ for each value of x.

The government budget constraint is given by

$$
b = R - G, \tag{32}
$$

where R denotes the revenue raised by the income tax (not including the demogrant b),

$$R = \mathbb{E}\left[T\left(y\left(x,d\right) - \alpha pd + \gamma\left(x\right)\right)\right]. \tag{33}$$

Totally differentiating the government budget constraint with respect to α, we obtain

$$\frac{\partial b}{\partial \alpha} = R_\alpha + R_b \frac{\partial b}{\partial \alpha}, \tag{34}$$

where $R_\alpha = \frac{\partial R}{\partial \alpha}$ and $R_b = \frac{\partial R}{\partial b}$. This in turn implies the identity

$$\frac{\partial b}{\partial \alpha} = \frac{R_\alpha}{1 - R_b}. \tag{35}$$

By the same logic, we also obtain:

$$\frac{\partial b}{\partial \gamma(x)} = \frac{R_{\gamma(x)}}{1 - R_b}. \tag{36}$$

where $R_{\gamma(x)} = \frac{\partial R}{\partial \gamma(x)}$ denotes the derivative of R with respect to the tax function $\gamma(\cdot)$ at the point x, capturing the marginal change in revenue resulting from an infinitesimal, localized increase of γ at x.

Using this notation, we can express the change in social welfare from a marginal budget-neutral increase in α as:

$$\begin{aligned} \frac{\mathrm{d}W}{\mathrm{d}\alpha} &= \mathbb{E}\left[\pi U_c pdT'\right] + \frac{\partial b}{\partial \alpha}\mathbb{E}\left[\pi U_c\right] \\ &= \frac{1}{1 - R_b}\left(\left(1 - R_b\right)\mathbb{E}\left[\pi U_c pdT'\right] + R_\alpha\,\mathbb{E}\left[\pi U_c\right]\right). \end{aligned} \tag{37}$$

We now establish the FOC with respect to the optimal $\gamma(x)$ for each x. Similar as we did with α, we apply the envelope theorem to get the government's first-order condition for welfare maximization with respect to $\gamma(x)$. That is, for all x we have:

$$0 = -\pi U_c T' g(x) + \frac{\partial b}{\partial \gamma(x)}\mathbb{E}\left[\pi U_c\right], \tag{38}$$

where $g(x)$ denotes the density of x. After substitution of $\partial b/\partial\gamma(x)$ and multiplication, we obtain the equivalent condition:

$$0 = -\pi U_c T'(1 - R_b) pdg(x) + R_{\gamma(x)} pd\,\mathbb{E}\left[\pi U_c\right]. \tag{39}$$

Since the above holds for each x, it also holds when we integrate over x:

$$0 = -(1 - R_b)\mathbb{E}\left[\pi U_c T' p d\right] + \mathbb{E}\left[\pi U_c\right] \cdot \int_x p d R_{\gamma(x)} \mathrm{d}x, \tag{40}$$

where the expectation operator denotes integration with respect to x using the associated probability density $g(x)$, whereas the final integral term represents regular integration with respect to x without density weighting. Substituting this result into (37), we get:

$$\begin{aligned} \frac{\mathrm{d}W}{\mathrm{d}\alpha} &= \frac{1}{1 - R_b}\left(\left(1 - R_b\right)\mathbb{E}\left[\pi U_c p d T'\right] + R_\alpha\, \mathbb{E}\left[\pi U_c\right]\right) \\ &= \frac{1}{1 - R_b}\left(\mathbb{E}\left[\pi U_c\right] \int_x p d R_{\gamma(x)} \mathrm{d}x + R_\alpha\, \mathbb{E}\left[\pi U_c\right]\right) \\ &= \frac{\mathbb{E}\left[\pi U_c\right]}{1 - R_b}\left(\int_x p d R_{\gamma(x)} \mathrm{d}x + R_\alpha\right). \end{aligned} \tag{41}$$

By differentiating the tax revenue definition, we now note

$$R_\alpha = \mathbb{E}\left[T' \cdot \left(\frac{\partial l}{\partial \alpha}\left(\gamma' + y_x\right) w + \frac{\partial d}{\partial \alpha}\left(y_d - \alpha p\right) - p d\right)\right], \tag{42}$$

and

$$R_{\gamma(x)} = T' \cdot \left(\frac{\partial l}{\partial \gamma(x)}\left(\gamma' + y_x\right) w + \frac{\partial d}{\partial \gamma(x)}\left(y_d - \alpha p\right) + 1\right) g(x). \tag{43}$$

Multiplying by $pd(x)$ and integrating with respect to x, the latter result implies:

$$\int_x p d R_{\gamma(x)} \mathrm{d}x = \mathbb{E}\left[p d T' \cdot \left(\frac{\partial l}{\partial \gamma(x)}\left(\gamma' + y_x\right) w + \frac{\partial d}{\partial \gamma(x)}\left(y_d - \alpha p\right) + 1\right)\right]. \tag{44}$$

We combine these results to rewrite the key term in (41) as follows:

$$\begin{aligned} \int_x p d R_{\gamma(x)} \mathrm{d}x + R_\alpha &= \mathbb{E}\left[p d T' \cdot \left(\frac{\partial l}{\partial \gamma(x)}\left(\gamma' + y_x\right) w + \frac{\partial d}{\partial \gamma(x)}\left(y_d - \alpha p\right) + 1\right)\right] \\ &\quad + \mathbb{E}\left[T' \cdot \left(\frac{\partial l}{\partial \alpha}\left(\gamma' + y_x\right) w + \frac{\partial d}{\partial \alpha}\left(y_d - \alpha p\right) - p d\right)\right] \\ &= \mathbb{E}\left[T' \cdot \left(\left(\frac{\partial l}{\partial \alpha} + p d \frac{\partial l}{\partial \gamma(x)}\right)\left(\gamma' + y_x\right) w + \left(\frac{\partial d}{\partial \alpha} + p d \frac{\partial d}{\partial \gamma(x)}\right)\left(y_d - \alpha p\right)\right)\right] \end{aligned} \tag{45}$$

We now want to argue that $\frac{\partial l}{\partial \alpha} + p d \frac{\partial l}{\partial \gamma(x)} = 0$. Because of weak separability, we know that

$$\frac{\partial l}{\partial \gamma(x)} = \frac{\partial l}{\partial v^*}\frac{\partial v^*}{\partial \gamma(x)} = \frac{\partial l}{\partial v^*} \cdot \left(-v_c T'(x)\right), \tag{46}$$

and similarly,

$$\frac{\partial l}{\partial \alpha} = \frac{\partial l}{\partial v^*}\frac{\partial v^*}{\partial \alpha} = \frac{\partial l}{\partial v^*} \cdot \left(v_c p d T'(x)\right). \tag{47}$$

Hence, we have

$$\frac{\partial l}{\partial \alpha} + pd\frac{\partial l}{\partial \gamma(x)} = 0. \tag{48}$$

Substituting this result into (45) implies

$$\int_x pdR_{\gamma(x)}\mathrm{d}x + R_\alpha = \mathbb{E}\left[T' \cdot \left(\left(\frac{\partial d}{\partial \alpha} + pd\frac{\partial d}{\partial \gamma(x)}\right)(y_d - \alpha p)\right)\right]. \tag{49}$$

By construction of the joint reform from our main analysis, we have $\frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} = \frac{\partial d}{\partial \alpha} + pd\frac{\partial d}{\partial \gamma(x)}$. Substituting this result into the above yields:

$$\int_x pdR_{\gamma(x)}\mathrm{d}x + R_\alpha = \mathbb{E}\left[T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} \cdot (y_d - \alpha p)\right]. \tag{50}$$

Substituting this back into (41) yields:

$$\frac{\mathrm{d}W}{\mathrm{d}\alpha} = \frac{\mathbb{E}\left[\pi U_c\right]}{1 - R_b} \cdot \mathbb{E}\left[T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} \cdot (y_d - \alpha p)\right]. \tag{51}$$

Note that $\pi U_c \geq 0$, and it is strictly positive on a set of positive measure, so that $\mathbb{E}[\pi U_c] > 0$. Moreover, if $R_b > 1$, then the tax system could not be optimal because a lump-sum adjustment to T would be Pareto improving (i.e., giving everyone a dollar would raise more than a dollar). Hence, the pre-multiplying factor on the right-hand side of (51) is positive.

Therefore, $\mathrm{d}W/\mathrm{d}\alpha$ is positive if and only if

$$\mathbb{E}\left[T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} \cdot \frac{y_d}{p}\right] > \alpha \cdot \mathbb{E}\left[T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha}\right]. \tag{52}$$

Equivalently,

$$\frac{\mathbb{E}\left[T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha} \cdot \frac{y_d}{p}\right]}{\mathbb{E}\left[T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha}\right]} = \mathbb{E}\left[\frac{T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha}}{\mathbb{E}\left[T' \cdot \frac{\mathrm{d}\hat{d}}{\mathrm{d}\alpha}\right]} \cdot \frac{y_d}{p}\right] > \alpha. \tag{53}$$

Expressing expectations as integrals over the distribution of skills w, we rewrite this condition as

$$\int_{\underline{w}}^{\overline{w}} \frac{T'\left(z(w)\right)\frac{\mathrm{d}\hat{d}(w)}{\mathrm{d}\alpha}}{\int_{\underline{w}}^{\overline{w}} T'\left(z(\omega)\right)\frac{\mathrm{d}\hat{d}(\omega)}{\mathrm{d}\alpha}\mathrm{d}F(\omega)} \cdot \frac{y_d\left(x\left(w\right), d\left(w\right)\right)}{p}\mathrm{d}F(w) > \alpha. \tag{54}$$

Finally, defining the weights

$$\theta_w = \frac{T'\left(z(w)\right)\frac{\mathrm{d}\hat{d}(w)}{\mathrm{d}\alpha}}{\int_{\underline{w}}^{\overline{w}} T'\left(z(\omega)\right)\frac{\mathrm{d}\hat{d}(\omega)}{\mathrm{d}\alpha}\mathrm{d}F(\omega)}, \tag{55}$$

we express the welfare-improvement condition compactly as

$$\int_{\underline{w}}^{\overline{w}} \theta_w \frac{y_d\left(x\left(w\right), d\left(w\right)\right)}{p}\mathrm{d}F(w) > \alpha. \tag{56}$$

This establishes the result. □

References

Atkinson, A. and J. Stiglitz (1976): "The design of tax structure: Direct versus indirect taxation," *Journal of Public Economics*, 6, 55 – 75.

Bohacek, R. and M. Kapička (2008): "Optimal human capital policies," *Journal of Monetary Economics*, 55, 1–16.

Bovenberg, A. L. and B. Jacobs (2005): "Redistribution and education subsidies are Siamese twins," *Journal of Public Economics*, 89, 2005–2035.

Christiansen, V. (1984): "Which commodity taxes should supplement the income tax?" *Journal of Public Economics*, 24, 195–220.

Findeisen, S. and D. Sachs (2016): "Education and optimal dynamic taxation: The role of income-contingent student loans," *Journal of Public Economics*, 138, 1–21.

Gauthier, S. and G. Laroque (2009): "Separability and public finance," *Journal of Public Economics*, 93, 1168–1174.

Kaplow, B. L. (2012): "Optimal control of externalities in the presence of income taxation," *International Economic Review*, 53, 487–509.

Kaplow, L. (2006): "On the undesirability of commodity taxation even when income taxation is not optimal," *Journal of Public Economics*, 90, 1235–1250.

——— (2020): "A unified perspective on efficiency, redistribution, and public policy," *National Tax Journal*, 73, 429–472.

Koehne, S. and D. Sachs (2022): "Pareto-improving reforms of tax deductions," *European Economic Review*, 148, 104214.

Mirrlees, J. A. (1976): "Optimal tax theory: A synthesis," *Journal of Public Economics*, 6, 327–358.

Stantcheva, S. (2017): "Optimal taxation and human capital policies over the life cycle," *Journal of Political Economy*, 125, 1931–1990.