Econometrica Supplementary Material

SUPPLEMENT TO "THE OPTIMAL INCOME TAXATION OF COUPLES"

(*Econometrica*, Vol. 77, No. 2, March, 2009, 537–560)

BY HENRIK JACOBSEN KLEVEN, CLAUS THUSTRUP KREINER, AND EMMANUEL SAEZ

Section S1 shows that a given path of earnings $(z_0(n), z_1(n))$ is implementable. Section S2 provides conditions for the existence of a solution to the maximization problem. Section S3 discusses conditions ensuring no bunching in the optimum. Section S4 discusses the outcome of a more general model with heterogeneity in both work costs and home production abilities. Section S5 provides technical details of the simulations.

S1. IMPLEMENTATION

AS IN THE ONE-DIMENSIONAL MECHANISM DESIGN theory, we define implementability as follows: An action profile $(z_0(n), z_1(n))_{n\in(\underline{n},\bar{n})}$ is implementable if and only if there exist transfer functions $(c_0(n), c_1(n))_{n\in(\underline{n},\bar{n})}$ such that $(z_l(n), c_l(n))_{l\in\{0,1\},n\in(\underline{n},\bar{n})}$ is a simple truthful mechanism.[1] The central implementability theorem of the one-dimensional case carries over to our model.

LEMMA S.1: *An action profile $(z_0(n), z_1(n))_{n\in(\underline{n},\bar{n})}$ is implementable if and only if $z_0(n)$ and $z_1(n)$ are both nondecreasing in n.*

PROOF: The utility function $c - nh(z/n)$ satisfies the classic single crossing (Spence–Mirrlees) condition (here equal to $x \cdot h''(x) > 0$ for all $x > 0$). Hence, from the one-dimensional case, we know that $z(n)$ is implementable, that is, there is some $c(n)$ such that $c(n) - nh(z(n)/n) \geq c(n') - nh(z(n')/n)$ for all n, n' if and only if $z(n)$ is nondecreasing.[2]

Suppose $(z_0(n), z_1(n))$ is implementable, implying that there exists $(c_0(n), c_1(n))$ such that $(z_l(n), c_l(n))_{l\in\{0,1\},n\in(\underline{n},\bar{n})}$ is a simple truthful mechanism. That implies in particular that $c_l(n) - nh(z_l(n)/n) \geq c_l(n') - nh(z_l(n')/n)$ for all n, n' and for $l = 0, 1$. Hence, the one-dimensional result implies that $z_0(n)$ and $z_1(n)$ are nondecreasing.

Conversely, suppose that $z_0(n)$ and $z_1(n)$ are nondecreasing. Because $z_0(n)$ is nondecreasing, the one-dimensional result implies there is $c_0(n)$ such that

[1]A mechanism is defined as truthful if there is a $\bar{q}(n)$ so that (i) when $q < \bar{q}(n)$, the set $(l' = 1, n' = n)$ maximizes $u(z_{l'}(n'), l', c_{l'}(n'), (n, q))$ over all (l', n'); (ii) when $q \geq \bar{q}(n)$, the set $(l' = 0, n' = n)$ maximizes $u(z_{l'}(n'), l', c_{l'}(n'), (n, q))$ over all (l', n').

[2]As an informal reminder, recall that if $z(n)$ is implementable, then the first-order condition is $\dot{c}(n) - h'(z(n)/n)\dot{z}(n) = 0$ and the second-order condition is $\ddot{c} - \ddot{z}h'(z(n)/n) - (\dot{z}^2/n)h''(z(n)/n) \leq 0$. Differentiating the first-order condition leads to $\ddot{c} - \ddot{z}h'(z(n)/n) - (\dot{z}^2/n)h''(z(n)/n) + (z(n)/n)h''(z(n)/n)(\dot{z}/n) = 0$. Combining with the second-order condition implies $(z(n)/n)h''(z(n)/n)\dot{z} \geq 0$, which implies $\dot{z} \geq 0$ using the Spence–Mirrlees condition.

 DOI: 10.3982/ECTA7343

$c_0(n) - nh(z_0(n)/n) \geq c_0(n') - nh(z_0(n')/n)$. Similarly, there is $c_1(n)$ such that $c_1(n) - nh(z_1(n)/n) \geq c_1(n') - nh(z_1(n')/n)$.

It is easy to show that the mechanism $(z_l(n), c_l(n))_{l\in\{0,1\}, n\in(\underline{n},\bar{n})}$ is actually truthful. Define $V_l(n) = c_l(n) - nh(z_l(n)/n)$ for $l = 0, 1$ and $\bar{q}(n) = V_1(n) - V_0(n)$. We only need to prove the cross-inequalities. For all $n, n', q \geq \bar{q}(n)$,

$$\begin{aligned} &u(z_0(n), 0, c_0(n), (n, q)) \\ &\quad = V_0(n) \geq V_1(n) - q \geq u(z_1(n'), 1, c_1(n'), (n, q)); \end{aligned}$$

for all $n, n', q < \bar{q}(n)$,

$$\begin{aligned} &u(z_1(n), 1, c_1(n), (n, q)) \\ &\quad = V_1(n) - q \geq V_0(n) \geq u(z_0(n'), 0, c_0(n'), (n, q)). \end{aligned}$$

The key assumption that allows us to obtain those simple results is the fact that q is separable in our utility specification. *Q.E.D.*

S2. EXISTENCE OF A SOLUTION TO THE MAXIMIZATION PROBLEM

Formally, our maximization problem is the optimal control problem $\dot{V} = b(n, V, z)$ with maximization objective $B^0 = \int_{\underline{n}}^{\bar{n}} b^0(n, V(n))\,dn$ and constraint $\int_{\underline{n}}^{\bar{n}} b^1(n, z(n), V(n))\,dn \geq 0$, where

$$\begin{aligned} b(n, V, z) &= \left(-h\left(\frac{z_0}{n}\right) + \left(\frac{z_0}{n}\right)h'\left(\frac{z_0}{n}\right), -h\left(\frac{z_1}{n}\right) + \left(\frac{z_1}{n}\right)h'\left(\frac{z_1}{n}\right)\right), \\ b^0(n, V) &= \left[\int_0^{V_1-V_0} \Psi(V_1 - q^w)p(q|n)\,dq \right. \\ &\qquad \left. + \int_{V_1-V_0}^{\infty} \Psi(V_0 + q^h)p(q|n)\,dq\right] f(n), \\ b^1(n, V, z) &= \left\{\left[z_1 + w - nh\left(\frac{z_1}{n}\right) - V_1\right]P(V_1 - V_0|n) \right. \\ &\qquad \left. + \left[z_0 - nh\left(\frac{z_1}{n}\right) - V_0\right](1 - P(V_1 - V_0|n))\right\} f(n). \end{aligned}$$

The functions b, b^0, and b^1 are continuous in n and class C^1 in (V, z) by assumption. Some convexity assumptions are required to demonstrate the existence of a solution (V, z). Strict concavity of the functions b^0 and b^1, and strict convexity of b in (V, z) are sufficient to obtain existence (and uniqueness); see, for example, Mangasarian (1966, Theorem 1, p. 141). However, in our application, concavity of b^0 and b^1 would be an unduly strong assumption.

It is possible to obtain existence without such strong assumptions using our Assumption 1 and the regularity assumptions on functions f, Ψ, P, and h. More precisely, according to Macki (1982, Theorem 3, p. 96), if we assume (i) an a priori bound on the path of admissible z,[3] (ii) b, b^0, and b^1 are continuous, and (iii) the sets $B(n,V,\lambda) = \{(y, b(n,V,z)) | z_0 \geq 0, z_1 \geq 0, y \geq -b^0(n,V) - \lambda \cdot b^1(n,V,z)\}$ are convex for all n, V and $\lambda \geq 0$, then there exists an optimal control z measurable on $(\underline{n}, \bar{n})$.[4]

Assumption (iii) is the only one that requires checking. In our problem, we have:

$$B(n,V,\lambda) = \Bigg\{\bigg(y, -h\Big(\frac{z_0}{n}\Big) + \Big(\frac{z_0}{n}\Big)h'\Big(\frac{z_0}{n}\Big),\\ -h\Big(\frac{z_1}{n}\Big) + \Big(\frac{z_1}{n}\Big)h'\Big(\frac{z_1}{n}\Big)\bigg) \Big| z_0 \geq 0, z_1 \geq 0,\\ y \geq -b^0(n,V)\\ - \lambda f(n) \cdot \bigg[(1-P) \cdot \Big(z_0 - nh\Big(\frac{z_0}{n}\Big) - V_0\Big)\\ + P \cdot \Big(w + z_1 - nh\Big(\frac{z_0}{n}\Big) - V_1\Big)\bigg]\Bigg\}.$$

Let us denote by $K(\cdot)$ the inverse of the strictly increasing function $x \to -h(x) + xh'(x)$. Note that $K(0) = 0$. Hence, we have

$$B(n,V,\lambda)\\ = \big\{(y, x_0, x_1) | x_0 \geq 0, x_1 \geq 0,\\ y + b^0(n,V) \geq nf(n)\lambda\big[(1-P) \cdot \big(h(K(x_0)) - K(x_0) + V_0\big)\\ + P \cdot \big(h(K(x_1)) - K(x_1) - w + V_1\big)\big]\big\}.$$

Therefore, $B(n,V,\lambda)$ is convex if $x \to h(K(x)) - K(x) \equiv \phi(x)$ is convex. By definition of $K(x)$, we have $-h(K(x)) + K(x)h'(K(x)) = x$, hence $K(x) \cdot h''((K(x)) \cdot K'(x)) = 1$. Therefore, we have $\phi'(x) = (h'(K(x)) - 1)K'(x) = -(1 - h'(K(x)))/[K(x)h''(K(x))]$. As $x \to K(x)$ is strictly increasing, Assumption 1 implies that $\phi'(x)$ is increasing.

[3]That means that we know a priori that there is some $Z > 0$ possibly large such that $0 \leq z_l(n) \leq Z$ for all $n \in (\underline{n}, \bar{n})$ and $l = 0, 1$. This assumption is weak when $\bar{n} < \infty$ as we do not expect the optimal tax system to generate infinitely large subsidies that drive up earnings z without bound.

[4]Macki (1982) presented optimal control as a minimization problem. Our maximization problem can be seen as minimizing $-\int b^0\, dn$. Macki (1982) did not include constraints such as $\int b^1\, dn \geq 0$, but such a constraint can be added by using a standard Lagrange multiplier λ and considering the objective $b^0 + \lambda \cdot b^1$.

S3. NO BUNCHING WITH LOW REDISTRIBUTIVE TASTES

As discussed in the main text, when redistributive tastes are low, the optimal solution is close to the laissez-faire no tax solution (where $z_0 = z_1 = n$), and, therefore, will have the property that z_l is strictly increasing in n and hence display no bunching.

A formal proof of this statement requires using advanced functional analysis (see Kleven, Kreiner, and Saez (2007)), but the argument is easy to understand. Let us parametrize redistributive tastes with γ and assume that social welfare is CRRA so that $\Psi(V) = V^{1-\gamma}/(1-\gamma)$. The no redistributive case is $\gamma = 0$. When $\gamma = 0$, the unique solution is $z_0 = z_1 = n$.[5] Let us denote by z^γ the solution for $\gamma \geq 0$ and assume that the strong convexity assumptions hold so that the solution is unique for $\gamma > 0$. It is possible to show that the solution is smooth in γ and can be written as $z^\gamma = z^0 + \gamma \cdot Z + o(\gamma)$, where $n \to Z(n)$ is the first-order deviation from z^0 for small γ and $o(\gamma)$ is small relative to γ (in a C^1 sense). Z actually satisfies a *linear* second-order differential equation with a unique smooth solution. As a result, $\dot{z}_l^\gamma(n) = 1 + \gamma \cdot \dot{Z}_l(n) + o(\gamma) > 0$ for γ small so that z^γ does not display bunching.

This result is of course true as well in the one-dimensional case and can be demonstrated without using advanced functional analysis. To our knowledge, this result has not been presented in the literature before[6] and is formally proven below.

PROPOSITION S.1: *Consider the one-dimensional Mirrlees* (1971) *optimal income tax problem with* $\Psi(V) = V^{1-\gamma}/(1-\gamma)$. *Assume that Assumption* 1 *in the main text is satisfied*, $n \to f(n)$ *is of class* C^1 *and bounded away from* 0, $x \to h(x)$ *is of class* C^3, $\underline{n} > 0$, *and* $\bar{n} < \infty$. *Then the solution does not display bunching for* $\gamma > 0$ *small enough*.

PROOF: In the one-dimensional case, under the assumptions of the proposition, the Hamiltonian is strictly concave in (z, V) for $\gamma > 0$ so that the solution is unique and given by the maximum principle first-order condition:

$$\varphi\left(\frac{z}{n}\right) \cdot nf(n) = \int_n^{\bar{n}} \left(1 - \frac{V(m)^{-\gamma}}{\lambda}\right) f(m)\, dm \tag{S.1}$$

with $\varphi(x) = (1 - h'(x))/(xh''(x))$, $\lambda = \int_{\underline{n}}^{\bar{n}} V(n)^{-\gamma} f(n)\, dn$, and $\dot{V}(n) = -h'(z/n) + (z/n)h'(z/n) \geq 0$. Transversality conditions imply that $z(\underline{n}) = \underline{n}$ and $z(\bar{n}) = \bar{n}$.

[5]In that case, it is actually possible to prove by contradiction directly that only $z_0 = z_1 = n$ can satisfy the first-order conditions spelled out in Proposition 1.

[6]Except in the monopoly problem (where social marginal welfare weights are constant), the literature does not seem to have presented any conditions that rule out bunching.

Obviously, if $\gamma = 0$, then $\lambda = 1$, and (S.1) implies $z = n$. With $\gamma > 0$, for all n, $0 < \underline{n}(1 - h(1)) \leq V(\underline{n}) \leq V(n) \leq V(\bar{n}) \leq \bar{n}(1 - h(1)) < \infty$ (as redistribution will increase the utility of the lowest skilled relative to laissez-faire and decrease utility of the highest skilled). Hence, $V(n)^{-\gamma} \to 1$ uniformly in n when $\gamma \to 0$. Hence $\lambda \to 1$ when $\gamma \to 0$. Assumption 1 (φ strictly decreasing and smooth) along with (S.1) and the normalization assumption $h'(1) = 1$ then implies that $z/n \to 1$ uniformly in n when $\gamma \to 0$. Differentiating (S.1) implies

$$\varphi'\left(\frac{z}{n}\right) \cdot \left[\dot{z} - \frac{z}{n}\right] f(n) + \varphi\left(\frac{z}{n}\right) \cdot (n + nf'(n)) = \left(\frac{V(n)^{-\gamma}}{\lambda} - 1\right) f(n).$$

As $\varphi(1) = 0$ and $\varphi'(1) < 0$, $z/n \to 1$ and $V(n)^{-\gamma}/\lambda \to 1$ uniformly in n when $\gamma \to 0$, we have $\dot{z} \to 1$ uniformly in n when $\gamma \to 0$. Hence, for γ small enough, $\dot{z} > 0$ for all n, implying that there is no bunching for γ small enough. *Q.E.D.*

S4. MODEL WITH BOTH WORK COSTS AND HOME PRODUCTION: OPTIMAL ZERO TAX CONDITION

In the main text, we are considering the polar models with either only work costs ($q^w = q, q^h = 0$) or only home production ($q^h = q, q^w = 0$). We consider here the more general model with both work costs and home production. We assume that (q^w, q^h) are distributed with density $k(q^w, q^h|n)$ conditional on primary earnings ability n. We characterize conditions on $k(\cdot, \cdot|n)$ so that there should be no tax on secondary earnings so that $T_1 \equiv T_0$.

PROPOSITION S.2: *If, for each n, (q^w, q^h) is distributed symmetrically around the diagonal $q^h + q^w = w$, that is, $k(q^h, q^w|n) = k(w - q^w, w - q^h|n)$ for all $q^h + q^w \leq w$, and the first-order conditions described in Proposition* 1 *are* sufficient *for a solution, then $T_0 \equiv T_1$, that is, there should be no tax on secondary earnings.*

PROOF: In the general model (q^w, q^h), equation (1) implies that secondary earners work if and only if $q^w + q^h \leq V_1 - V_0$. Let us denote (as in the polar cases) by $P(V_1 - V_0|n)$ the probability that $q^w + q^h \leq V_1 - V_0$. The symmetry property implies that $P(w|n) = 1/2$. Suppose that $V_1 - V_0 = w$. Then

$$g_0 = \frac{\int_{q^h + q^w > w} \Psi'(V_0 + q^h) k(q^h, q^w|n)\, dq^h\, dq^w}{(1 - P(w|n)) \cdot \lambda}$$
$$= \frac{\int_{w - q^h + w - q^w < w} \Psi'(V_1 - (w - q^h)) k(q^h, q^w|n)\, dq^h\, dq^w}{P(w|n) \cdot \lambda}.$$

Changing variables to $r^h = w - q^w$ and $r^w = w - q^h$, we have

$$g_0 = \frac{\int_{r^h + r^w < w} \Psi'(V_1 - r^w) k(w - r^w, w - r^h | n) \, dr^h \, dr^w}{P(w|n) \cdot \lambda} = g_1,$$

where the last equality is obtained using the symmetry property. This implies that if the tax system is such that $T_0 \equiv T_1$, then $V_1 - V_0 = w$, $\Delta T = 0$, $z_0 = z_1$, $\varepsilon_0 = \varepsilon_1 \equiv \varepsilon$, $g_0 = g_1 \equiv g$, $P = 1/2$, and $T_0' = T_1' \equiv T'$ with $T'/(1 - T') = (1/(\varepsilon n f(n))) \int_n^{\bar{n}} (1 - g) f(m) \, dm$ (the standard Mirrlees (1971) formula) satisfy both first-order conditions (7) and (8). If those conditions are sufficient for an optimum, that means that the standard Mirrlees (1971) tax system with no tax on secondary earnings is the full optimum. The sufficiency condition would be satisfied under concavity assumptions (as we discussed in Section S2). *Q.E.D.*

Intuitively, if q^h and q^w have the symmetry property, then under no tax on secondary earnings, $(V_0 + q^h)/(q^h + q^w) > w$ and $V_1 - q^w = V_0 + (w - q^w)/(q^h + q^w) < w$ have the same distribution and hence one- and two-earner couples have the same marginal welfare weights ($g_0 = g_1$). As a result, there is no point in that case for the government to tax (or subsidize) secondary earnings. The symmetry property holds in the particular case where q^h and q^w are identically and independently distributed with density $p(q)$ symmetric around $w/2$ ($p(w - q) = p(q)$). The property can also hold when q^h and q^w are positively (or negatively) correlated. For example, when $q^h = q^w$ (perfect correlation), the property holds if again the density is symmetric around $w/2$.

When the symmetry property fails, under no tax on secondary earnings, $(V_0 + q^h)/(q^h + q^w) > w$ will have a less favorable distribution than $V_1 - q^w = V_0 + (w - q^w)/(q^h + q^w) < w$ if there is more "heterogeneity" in q^w than in q^h. In that case, $g_1 < g_0$ under no tax on secondary earnings. Hence, imposing a positive tax on secondary earners is desirable. As Assumption 2 in the main text, strict convexity of Ψ' will tend to make the difference between g_0 and g_1 shrink with n so that we would expect the optimal system to display negative jointness. Symmetrically, if there is more "heterogeneity" in q^h than q^w, $g_1 > g_0$ and secondary earnings should be subsidized, and we should expect the size of subsidy to shrink with n if Ψ' is convex.

S5. NUMERICAL SIMULATIONS

Simulations are performed with Matlab software and our programs are available upon request. We select a grid for n, from $\underline{n}$ to $\bar{n}$ with 1000 elements: $(n_k)_k$. Integration along the n variable is carried out using the trapezoidal approximation. All integration along the q variable is carried out using explicit closed form solutions using the incomplete β function:

$$\begin{aligned}\int_0^{V_1-V_0} \Psi'(V_1-q)p(q)\,dq &= \int_0^{V_1-V_0} \frac{1}{(V_1-q)^\gamma}\frac{\eta\cdot q^{\eta-1}}{q_{\max}^\eta}\,dq \\ &= \frac{\eta}{q_{\max}^\eta}\int_0^{V_1-V_0}(V_1-q)^{-\gamma}q^{\eta-1}\,dq \\ &= \frac{\eta\cdot V_1^{\eta-\gamma}}{q_{\max}^\eta}\int_0^{1-V_0/V_1} t^{\eta-1}(1-t)^{-\gamma}\,dt \\ &= \frac{\eta\cdot V_1^{\eta-\gamma}}{q_{\max}^\eta}\cdot\beta\left(1-\frac{V_0}{V_1},\eta,1-\gamma\right),\end{aligned}$$

where the incomplete beta function β is defined as (for $0\le x\le 1$)

$$\beta(x,a,b)=\int_0^x t^{a-1}(1-t)^{b-1}\,dt.$$

Matlab does not compute it directly for $\gamma\ge 1$ ($b\le 0$), but we have used the development in series to compute it very accurately and quickly with a subroutine:

$$\beta(x,a,b)=1+\sum_{n=1}^{\infty}\frac{(1-b)(2-b)\cdots(n-b)}{n!}\cdot\frac{x^{n+a}}{n+a}.$$

Simulations proceed by iteration:

We start with given T_0' and T_1' vectors, derive all the vector variables z_0, z_1, V_0, V_1, $\bar{q}$, T_0, T_1, λ, and so forth which satisfy the government budget constraint and the transversality conditions.[7] This is done with a subiterative routine that adapts T_0 and T_1 as the bottom $\underline{n}$ until those conditions are satisfied. We then use the first-order conditions (7) and (8) from Proposition 1 to compute new vectors T_0' and T_1'. To allow convergence, we use adaptive iterations where we take as the new vectors T_0' and T_1', a weighted average of the old vectors and newly computed vectors. The weights are adaptively adjusted downward when the iteration explodes. We then repeat the algorithm.

This procedure converges to a fixed point in most circumstances. The fixed point satisfies all the constraints and the first-order conditions. We check that the resulting z_0 and z_1 are nondecreasing so that the fixed point solution is implementable. Hence, the fixed point is expected to be the optimum.[8]

[7]Then adjust the constants for $T_l(\underline{n})$ until all those constraints are satisfied. This is done using a secondary iterative procedure.

[8]We also compute total social welfare and verify on examples that it is higher than social welfare generated by other tax rates T_1' and T_0' that satisfy the government budget constraint.

The central advantage of our method is that the optimal solution can be approximated very closely and quickly. In contrast, direct maximization where we search the optimum over a large set of parametric tax systems by computing directly social welfare would be much slower and less precise.

REFERENCES

KLEVEN, H. J., C. T. KREINER, AND E. SAEZ (2007): "The Optimal Income Taxation of Couples as a Multi-Dimensional Screening Problem," Working Paper 2092, CESifo.

MACKI, J. (1982): *Introduction to Optimal Control Theory*. New York: Springer-Verlag.

MANGASARIAN, O. L. (1966): "Sufficient Conditions for the Optimal Control of Nonlinear Systems," *Journal of SIAM Control*, 4, 139–152.

MIRRLEES (1971): "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38, 175–208.

Dept. of Economics, London School of Economics, Houghton Street, London WC2A 2AE, U.K. and Economic Policy Research Unit, Dept. of Economics, University of Copenhagen, Copenhagen, Denmark and Centre for Economic Policy Research, London, U.K.; h.j.kleven@lse.ac.uk,

Dept. of Economics, University of Copenhagen, Studiestraede 6, 1455 Copenhagen, Denmark and Economic Policy Research Unit, Dept. of Economics, University of Copenhagen, Copenhagen, Denmark and CESifo, Munich, Germany,

and

Dept. of Economics, University of California–Berkeley, 549 Evans Hall 3880, Berkeley, CA 94720, U.S.A. and NBER; saez@econ.berkeley.edu.

Manuscript received August, 2007; final revision received August, 2008.