Online Appendix for
Optimal income taxation with labor supply responses at two margins: When is an Earned Income Tax Credit optimal?

Emanuel Hansen*

December 31, 2019

B Proofs for Subsection 4.2

Proof of Lemma 3

Proof. Lemma 3 establishes the existence of two functions β_j^U and β_j^D that determine for each couple (α_j, α_{j+1}) which of the IC constraints between skill groups j and $j+1$ are satisfied. The properties of both functions are illustrated in Figure 3. In the following, I first provide a complete and detailed derivation of part (i) on the existence and the properties of function β_j^U. I then sketch the corresponding proof for part (ii) on function β_j^D.

Part (i), first step: *If a is small enough, there is a number $\bar{\beta}_j > 1$ such that the upward IC of ω_j workers is violated if $\alpha_j = \alpha_{j+1} > \bar{\beta}_j$, and satisfied if $\alpha_j = \alpha_{j+1} < \bar{\beta}_j$.*

The claim generalizes the first step of the proof of Proposition 1 for arbitrary skill sets, effort cost functions and type distributions. By (11), allocation $(\tilde{y}, \tilde{\delta})$ satisfies the upward IC of ω_j workers if and only if

$$\tilde{\delta}_j - \tilde{\delta}_{j+1} \geq h(\tilde{y}_{j+1}, \omega_{j+1}) - h(\tilde{y}_{j+1}, \omega_j) \ . \tag{A1}$$

First, I show that (A1) is violated if α' is equal to some fixed number above 1 and a is close enough to 1. For this purpose, fix two numbers $\hat{a} > 1$ and $\gamma \in (1, 2)$ such that $(\tilde{y}, \tilde{\delta})$ is well-behaved and defined by first-order conditions for $\alpha' = \gamma$ and any $a \in [1, \hat{a}]$. In the

*University of Cologne, Center for Macroeconomic Research, Albertus-Magnus-Platz, D-50923 Köln, Germany. Email: hansen@wiso.uni-koeln.de.

limit case $a = 1$, we have $\omega_{j+1} = \omega_j$, $\tilde{y}_{j+1} = \tilde{y}_j$, $\eta_j(\delta) = \eta_{j+1}(\delta)$ and $\tilde{\delta}_{j+1} = \tilde{\delta}_j$. Hence, (A1) is satisfied with a strict equality.

Now consider a small increase in a. For $a = 1$, the derivative of the left-hand side of (A1) with respect to a is given by

$$\left.\frac{d\left[\tilde{\delta}_j - \tilde{\delta}_{j+1}\right]}{da}\right|_{a=1} = -\left.\frac{d\tilde{\delta}_{j+1}}{da}\right|_{a=1} = \frac{\omega_j h_\omega(\tilde{y}_j, \omega_j)\eta_j^2 + \omega_j \partial\eta_j/\partial\omega_j(\gamma - 1)}{\eta_j^2 + \partial\eta_j/\partial\delta_j(\gamma - 1)} \ ,$$

while the derivative of the right-hand side is given by

$$\left.\frac{d\left[h(\tilde{y}_{j+1}, \omega_{j+1}) - h(\tilde{y}_{j+1}, \omega_j)\right]}{da}\right|_{a=1} = \omega_j h_\omega(\tilde{y}_j, \omega_j) \ .$$

Hence, the derivative of the difference between the left-hand side and the right-hand side follows as

$$\frac{\omega_j(\gamma - 1)}{\eta_j^2 + \partial\eta_j/\partial\delta_j(\gamma - 1)} \left[\partial\eta_j/\partial\omega_j - h_\omega(\tilde{y}_j, \omega_j)\partial\eta_j/\partial\delta_j\right] \ .$$

While the fraction before the bracket is strictly positive, the term in the bracket is strictly negative because $h_\omega < 0$ and because η_j is decreasing in both ω_j and δ_j by Assumption 2. Hence, there is a number $z > 1$ such that the left-hand side of (A1) is smaller than the right-hand side for a for all $a \in (1, z)$. This implies that $(\tilde{y}, \tilde{\delta})$ violates the upward IC of ω_1 workers whenever $\alpha' > 1$ and a is close enough to 1.

Second, I show that there is a unique threshold $\bar{\beta}_j \in (1, \gamma)$ such that (A1) is satisfied if and only if $\alpha' < \bar{\beta}_j$. The derivative of $\tilde{\delta}_k$ with respect to α_k can be calculated by implicit differentiation of (28). In particular, $\tilde{\delta}_k$ is strictly increasing in α_k and given by

$$\frac{d\tilde{\delta}_k}{d\alpha_k} = \left[\eta_k - \frac{\partial\eta_k}{\partial\delta_k}\left(\tilde{y}_k - h(\tilde{y}_k, \omega_k) - \tilde{\delta}_k\right)\right]^{-1} = \left[\eta_k - \frac{\partial\eta_k/\partial\delta_k}{\eta_k}(1 - \alpha_k)\right]^{-1} > 0 \ ,$$

for any $k \in J$ whenever $(\tilde{y}, \tilde{\delta})$ satisfies the second-order condition for a welfare optimum. Consequently, the difference $\tilde{\delta}_j - \tilde{\delta}_{j+1}$ is strictly decreasing in α',

$$\begin{aligned} \left.\frac{d[\tilde{\delta}_j - \tilde{\delta}_{j+1}]}{d\alpha'}\right|_{\alpha_j = \alpha_{j+1} = \alpha'} &= \frac{1}{\eta_j - \frac{\partial\eta_j/\partial\delta_j}{\eta_j}(1-\alpha')} - \frac{1}{\eta_{j+1} - \frac{\partial\eta_{j+1}/\partial\delta_{j+1}}{\eta_{j+1}}(1-\alpha')} < 0 \\ &\Leftrightarrow \left(\frac{\partial\eta_j/\partial\delta_j}{\eta_j} - \frac{\partial\eta_{j+1}/\partial\delta_{j+1}}{\eta_{j+1}}\right)(1 - \alpha') < \eta_j - \eta_{j+1} \ . \end{aligned}$$

The first bracket on the left-hand side is equal to the semi-elasticity of the relative participation responses η_j/η_{j+1}. By Assumption 3, the absolute value of this semi-elasticity is smaller than $\eta_j - \eta_{j+1}$. For any $\alpha' \in (0, 2)$, the absolute value of $1 - \alpha'$ is smaller than 1. Consequently, the left-hand side of the previous inequality is smaller than the right-hand side of the last inequality. Hence, the left-hand side of (A1) is strictly decreasing in α'. In contrast, the right-hand side of (A1) does not depend on α'.

Recall from the first step 1 that, as long as a is small enough, the upward IC (A1) is violated for $\alpha' = \gamma$. It is satisfied for $\alpha' = 1$ as shown in the proof of Proposition 1. Hence, the monotonicity in α' ensures that there is a unique threshold $\bar{\beta}_j \in (1, \gamma)$ such that the upward IC is satisfied if $\alpha' \leq \bar{\beta}_j$ and violated if $\alpha' \in (\bar{\beta}_j, 2)$.

Part (i), second step: Existence and properties of function β_{Uj}. The claim generalizes the statement proven in the first part of Proposition 2. The proof is simple and mainly exploits that $\tilde{\delta}_k$ is continuously differentiable and strictly increasing in α_k for any $k \in J$, as shown in the first step. This directly implies that the left-hand side of (A1) is strictly increasing in α_j and strictly decreasing in α_{j+1} (and continuously differentiable in both parameters). Hence, if function β_{Uj} exists, (A1) has to be satisfied with equality for any α_j and $\alpha_{j+1} = \beta_{Uj}(\alpha_j)$ such that there is a well-behaved welfare maximum. Moreover, β_{Uj} must be continuously differentiable and strictly increasing in α_j.

To prove the existence and the properties of function β_{Uj}, I again exploit the monotonicity of (A1) in α_j and α_{j+1}. In the first part of this proof, I have shown that (A1) is satisfied with equality for $\alpha_j = \alpha_{j+1} = \bar{\beta}_j$. First, fix α_j to be equal to some $\gamma > \bar{\beta}_j$. As the left-hand side of (A1) is strictly increasing in α_j, the upward IC is also satisfied if $\alpha_j = \gamma$ and $\alpha_{j+1} = \bar{\beta}_j$. However, I have shown above that it is violated if $\alpha_{j+1} = \alpha_j = \gamma > \bar{\beta}_j$. Due to the monotonicity in α_{j+1}, there is a unique number $z \in (\bar{\beta}_j, \gamma)$ such (A1) is satisfied with equality for $\alpha_{j+1} = z$ and $\alpha_j = \gamma$.

Second, fix α_j so be equal to some number $\gamma < \bar{\beta}_j$. As the left-hand side of (A1) is strictly increasing in α_j, the upward IC is violated if $\alpha_j = \gamma$ and $\alpha_{j+1} = \bar{\beta}_j$. However, I have shown above that it is satisfied if $\alpha_{j+1} = \alpha_j = \gamma < \bar{\beta}_j$. Due to the monotonicity in α_{j+1}, there is a unique number $z \in (\bar{\beta}_j, \gamma)$ such (A1) is satisfied with equality for $\alpha_{j+1} = z$ and $\alpha_j = \gamma$.

Hence, for any α_j such that $(\tilde{y}, \tilde{\delta})$ is well-defined, the function $\beta_{Uj}(\alpha_j)$ is well-defined as the unique number such that (A1) is satisfied with equality for $\alpha_{j+1} = \beta_{Uj}(\alpha_j)$. This completes the proof of part (i) of Lemma 3.

Part (ii): Existence and properties of function β_{Dj}. The proof of part (ii) applies similar arguments as the proof of part (i) to determine conditions on the welfare weights α_j and α_{j+1} such that allocation $(\tilde{y}, \tilde{\delta})$ satisfies the downward IC of ω_{j+1} workers, i.e.,

$$\tilde{\delta}_{j+1} - \tilde{\delta}_j \geq h(\tilde{y}_j, \omega_j) - h(\tilde{y}_j, \omega_{j+1}) \; . \tag{A2}$$

First, it can be shown that (A2) is violated if $\alpha_j = \alpha_{j+1}$ is equal to some number $\gamma < 1$ and a is slightly larger than 1. Again, this is done by comparing the derivatives of the left-hand side and the right-hand side of (A2) in a, evaluated for $a = 1$. Second, for $\alpha_j = \alpha_{j+1}$ equal to some number α', the left-hand side is strictly increasing in α' (this directly follows from the observation that the left-hand side of (A2) is equal to minus one times the left-hand side of (A1)) . For $\alpha' = 1$, (A2) is satisfied. Hence, if a is small enough, there is a unique number $\underline{\beta}_j \in (\gamma, 1)$ such that the downward IC of ω_{j+1} workers is satisfied if $\alpha' \geq \underline{\beta}_j$ and violated if $\alpha' < \underline{\beta}_j$.

Finally, I can again exploit the monotonicity of $\tilde{\delta}_j$ and $\tilde{\delta}_{j+1}$ in the welfare weights α_j and α_{j+1}, respectively, to show that there is a uniquely defined function β_{Dj} such that (34) is satisfied with equality for any combination of α_j and $\alpha_{j+1} = \beta_{Dj}(\alpha_j)$. For $\alpha_j < \underline{\beta}_j$, we have $\beta_{Dj}(\alpha_j) \in (\alpha_j, \underline{\beta}_j)$. For $\alpha_j > \underline{\beta}_j$, we have $\beta_{Dj}(\alpha_j) \in (\underline{\beta}_j, \alpha_j)$. Moreover, the differentiability and monotonicity of $\tilde{\delta}_k$ in α_k ensures that β_{Dj} is continuously differentiable and strictly increasing in α_j. □

Proof of Proposition 3

Proof. The conditions in Proposition 3 imply that the first-and-half-best allocation $(\tilde{y}, \tilde{\delta})$ (i) violates the upward IC of ω_j workers for at least one $j \in \{1, \dots, k-1\}$ and either violates or satisfies with strict equality the upward IC of ω_j workers for each $j \in \{1, \dots, k-1\}$, and (ii) satisfies the downward IC of ω_j workers for each $j \in \{k+1, \dots, n\}$, i.e.,

$$\tilde{\delta}_j - \tilde{\delta}_{j+1} < h(\tilde{y}_{j+1}, \omega_{j+1}) - h(\tilde{y}_{j+1}, \omega_j) \quad \text{for at least one } j \in \{1, \dots, k-1\} \ , \tag{A3}$$

$$\tilde{\delta}_j - \tilde{\delta}_{j+1} \leq h(\tilde{y}_{j+1}, \omega_{j+1}) - h(\tilde{y}_{j+1}, \omega_j) \quad \text{for all } j \in \{1, \dots, k-1\} \ , \tag{A4}$$

$$\tilde{\delta}_j - \tilde{\delta}_{j-1} \geq h(\tilde{y}_{j-1}, \omega_{j-1}) - h(\tilde{y}_{j-1}, \omega_j) \quad \text{for all } j \in \{k, \dots, k-1\} \ . \tag{A5}$$

I now show that this ensures an upward distortion in the second-best output of ω_j workers, $y_j^* > \tilde{y}_j$ for all $j \in \{2, \dots, k\}$. This requires to determine the set of binding IC constraints in the second-best allocation (y^*, δ^*). Assume that the upward IC of ω_j workers is binding for j in the set $J_U \subset J$ only and that the downward IC of ω_j workers is binding for j in the set $J_D \subset J$ only. Then, the Lagrangian of the optimal tax problem is given by

$$\begin{aligned}
\mathcal{L} \ = \ & \sum_{j=1}^{n} f_j \left[G_j(\hat{\delta}_j) \Big(y_j - h\left[y_j, \omega_j\right) \right] + \hat{\delta}_j \left[\alpha_j - 1\right] \Big) - \alpha_j \int_{\underline{\delta}}^{\hat{\delta}_j} \delta dG_j(\delta) \right] \\
& + \sum_{j \in J_U} \mu_j^U \left[\hat{\delta}_j - \hat{\delta}_{j+1} - h(y_{j+1}, \omega_{j+1}) + h(y_{j+1}, \omega_j) \right] \\
& + \sum_{j \in J_D} \mu_j^D \left[\hat{\delta}_j - \hat{\delta}_{j-1} - h(y_{j-1}, \omega_{j-1}) + h(y_{j-1}, \omega_j) \right] \ ,
\end{aligned}$$

where μ_j^U and μ_j^D are the Lagrange multiplier of the upward IC and the downward IC, respectively, of ω_j workers.

In the following, I show that the upward IC of each skill group $j \in \{1, \dots, k-1\}$ is binding and that $y_j^* > \tilde{y}_j$ for each $j \in \{2, \dots, k\}$ in (y^*, δ^*). To provide the relevant arguments in a simple way, I go through three alternative cases for an economy with 4 skill groups.

Case 1: $k = 3$, upward IC of ω_2 workers violated. Assume that $n = 4$, $k = 3$ and that equation (A3) holds for $j = 2$, (A4) holds for $j = 1$ and (A5) holds for $j = 4$. It is immediately clear that (y^*, δ^*) cannot be given by an allocation in which no IC is binding, nor by an allocation in which only some downward ICs are binding. Moreover, I can rule out the possibility that only the upward IC of ω_2 workers is binding, i.e., that $J_U = \{2\}$ and $J_D = \emptyset$. In this case, the FOCs would imply that $\delta_1^* = \tilde{\delta}_1$, $y_2^* = \tilde{y}_2$ and δ_2^* was given by

$$\delta_2^* = y_2^* - h(y_2^*, \omega_2) + \frac{\alpha_2 - 1}{\eta_2^*} + \frac{\mu_2^U}{f_2 g_2(\delta_2^*)} > \tilde{\delta}_2 = \tilde{y}_2 - h(\tilde{y}_2, \omega_2) + \frac{\alpha_2 - 1}{\tilde{\eta}_2}\ , \tag{A6}$$

where η_2^* and $\tilde{\eta}_2$ denote the semi-elasticities of participation of ω_2 workers in (y^*, δ^*) and in $(\tilde{y}, \tilde{\delta})$, respectively. We find that $\delta_2^* > \tilde{\delta}_2$ because $\mu_2^U > 0$ and $\eta_2^* \leq \tilde{\eta}_2$ for $\delta_2^* > \tilde{\delta}_2$ by Assumption 2 (moreover, the second-order condition must be satisfied in a welfare maximum). Combining this inequality with $\delta_1^* = \tilde{\delta}_1$, $y_2^* = \tilde{y}_2$ and (A4) for $j = 1$ implies that $\delta_1^* - \delta_2^* < h(y_2^*, \omega_2) - h(y_2^*, \omega_1)$, i.e., the upward IC of ω_1 workers is violated. Hence, this allocation cannot be second-best. Rather, both upward ICs of ω_1 and ω_2 workers must be binding as long as the upward IC of ω_3 workers and the downward IC of ω_4 workers are not taken into account.

It remains to determine the set of binding IC constraint and the distortions in y_2^* and y_3^* when the full set of IC constraints is taken into account. For this purpose, we can easily generalize the arguments given in the proof of Propositions 1 (ii) and 2 (ii). As a result, the upward ICs of ω_1 and ω_2 workers are binding in (y^*, δ^*), while the downward IC of ω_4 workers might be binding or not. In any case, however, both y_2^* and y_3^* are upwards distorted at the intensive margin.

Case 2: $k = 3$, upward IC of ω_1 workers violated. Assume that $n = 4$, $k = 3$ and that equation (A3) holds for $j = 1$, (A4) holds for $j = 2$ and (A5) holds for $j = 4$. Again, (y^*, δ^*) cannot be given by an allocation in which no IC is binding nor an allocation in which only some downward ICs are binding. Moreover, I can rule out the possibility that only the upward IC of ω_1 workers is binding, i.e., $J_U = \{1\}$ and $J_D = \emptyset$. In this case, the FOCs would imply that $\delta_3^* = \tilde{\delta}_3$, $y_3^* = \tilde{y}_3$, $y_2^* > \tilde{y}_2$ and that δ_2^* was given by

$$\delta_2^* = y_2^* - h(y_2^*, \omega_2) + \frac{\alpha_2 - 1}{\eta_2^*} - \frac{\mu_1^U}{f_2 g_2(\delta_2^*)} < \tilde{\delta}_2 = \tilde{y}_2 - h(\tilde{y}_2, \omega_2) + \frac{\alpha_2 - 1}{\tilde{\eta}_2}\ . \tag{A7}$$

We find that $\delta_2^* < \tilde{\delta}_2$ because $\mu_1^U > 0$, $y_2^* - h(y_2^*, \omega_2) < \tilde{y}_2 - h(\tilde{y}_2, \omega_2)$ and $\eta_2^* \geq \tilde{\eta}_2$ for $\delta_2^* < \tilde{\delta}_2$ by Assumption 2. Combining this inequality with $\delta_3^* = \tilde{\delta}_3$, $y_3^* = \tilde{y}_3$ and (A4) for $j = 2$ implies that $\delta_2^* - \delta_3^* < h(y_3^*, \omega_3) - h(y_3^*, \omega_2)$, i.e., the upward IC of ω_2 workers is violated. Again, this means that both upward ICs of ω_1 and ω_2 workers must be binding as long as the upward IC of ω_3 workers and the downward IC of ω_4 workers are not taken into account. As in case 1, we can apply the arguments given in the proof of Propositions 1 (ii) and 2 (ii) to show that both upwards ICs are binding in (y^*, δ^*) and that both y_2^* and y_3^* are upwards distorted at the intensive margin.

Case 3: $k = 2$**.** Assume that $n = 4$ and $k = 2$. This implies that (A3) holds for $j = 1$ and that (A5) holds for $j \in \{3, 4\}$. For concreteness, assume also that (A4) is not satisfied for $j \in \{2, 3\}$. As before, (y^*, δ^*) cannot be given by an allocation in which no ICs are binding or only some downwards ICs are binding. Consider a relaxed problem that takes into account only the upward IC constraints of ω_1 and ω_2 workers, and only the downward IC constraints of ω_2 and ω_3 workers. The proof of Propositions 1 (ii) and 2 (ii) directly implies that, in the solution (y^R, δ^R) to this relaxed problem, the upward IC of ω_1 workers is binding and y_2^R is upwards distorted. Besides, there are three possible constellations: (1) If no other IC is binding in (y^R, δ^R), then the same is true in the second-best allocation (y^*, δ^*). (2) If the upward IC of ω_2 workers is binding in (y^R, δ^R), then we can apply the arguments in the proof of Propositions 1 (ii) and 2 (ii) to show the same is true in (y^*, δ^*) and that both y_2^* and y_3^* are upwards distorted. (3) If the downward IC of ω_3 workers is binding in (y^R, δ^R), then it is also binding in (y^*, δ^*) but y_2^* is upwards distorted by the arguments given the proof of Propositions 1 (ii) and 2 (ii). Moreover, the downward IC of ω_4 workers may be binding or not and, correspondingly, y_3^* may be undistorted or downwards distorted.

For any economy with any $n > 4$ and $k \in \{2, \ldots, n-1\}$, the complete proof follows from a repeated application of the arguments given in cases 1 to 3. □

Proof of Proposition 4

Proof. In the following, I derive conditions under which some decreasing sequence of welfare weights $(\alpha_1, \ldots, \alpha_n)$ satisfies the conditions in Proposition 3, i.e., upward distortions at both margins are optimal. While this result generalizes Proposition 1, the formal arguments in the following proof differ substantially.

Specifically, the proof shows that a sequence α of welfare weights can at the same time (a) satisfy conditions (i) and (ii) in Proposition 3, (b) be weakly decreasing such that $\alpha_{j+1} \leq \alpha_j$, and (c) imply an average weight of $\sum_{j=1}^{n} f_j \alpha_j = 1$. The strategy of the proof is to construct a candidate sequence $q = (q_1, \ldots, q_{n-1})$ and to derive conditions under which q satisfies (a) to (c). For this purpose, fix $\bar{q} > 1$ to be given by a number such that $\tilde{\delta}_j$ is defined by the first-order condition (28) for any $j \in J$ and $\alpha_j \in [0, \bar{q}]$. Now, consider the sequence q such that $q_j = \bar{q} > 1$ for all $j \in \{1, \ldots, k\}$ and $q_{j+1} = \max\left\{\beta_{Dj}(q_j), \underline{q}\right\}$ for all $j \in \{k, \ldots, n-1\}$, where $\underline{q} := \max_{j \geq k} \underline{\beta}_j < 1$.

Property (a): Conditions in Proposition 3. First, by Lemma 3 (i), there is for each $j \in J$ a number $\bar{\beta}_j > 1$ such that $\beta_{Uj}(q_j) < q_j$ if $q_j > \bar{\beta}_j$. As shown in the proof of Lemma 3 (i), $\bar{\beta}_j < \bar{q}$ if $a = \omega_{j+1}/\omega_j$ is close enough to 1. Then, $q_{j+1} > \beta_{Uj}(q_j)$ for all $j \leq k-1$. Second, by construction, we have $q_{j+1} \geq \beta_{Dj}(q_j)$ for all $j \geq k$ such that $q_{j+1} > \underline{q}$. Finally, by Lemma 3 (ii), $\underline{q} = \max_{j \geq k} \underline{\beta}_j$ ensures that $q_{j+1} \geq \beta_{Dj}(q_j)$ for all $j \geq k$ such that $q_{j+1} = \underline{q}$.

Property (b): Decreasing weights. First, by construction, $q_{j+1} \leq q$ for all $j \leq k-1$. Second, by Lemma 3, $\underline{q} = \max_{j \geq k} \underline{\beta}_j$ implies that $q_{j+1} = \max\left\{\underline{q}, \beta_{Dj}(q_j)\right\} \leq q_j$ for all $j \in J$ (with a strict inequality if $q_j > \underline{\beta}_j$).

Property (c): Average weight of 1. First, with decreasing weights, $\sum_{j=1}^{n} f_j q_j = 1$ and $q_1 > 1$ requires that $q_n < 1$. As $q_j = \bar{q} > 1$ for all $j \leq k$, this obviously requires that $n \geq k+1$. In the following, I show that there is a finite number $m \geq k+1$ such that $q_n < 1$ whenever $n \geq m$. To see why, assume that there is a lower bound $\varepsilon > 0$ such that $q_j - q_{j+1} > \varepsilon$ for each $j \geq k$ such that $\alpha_j \geq 1$. If this is true, then $\alpha_j < \min\{1, \bar{q} - (j-k)\varepsilon\}$ for each $j \geq k+1$. Consequently, there is a unique natural number $m \leq k + (\bar{q}-1)/\varepsilon$ such that $\alpha_j < 1$ is ensured for all $j \in [m, n]$.

I now show that $q_j - q_{j+1}$ is indeed smaller than some number $\varepsilon > 0$ if $q_j \geq 1$, $\omega_{j+1}/\omega_j = a > 1$ and Assumption 4 holds. For this purpose, recall that $\tilde{\delta}_{j+1}$ is an increasing function of α_{j+1} and that, with $q_{j+1} = q_j \geq 1$, $\tilde{\delta}_{j+1} - \tilde{\delta}_j$ would be weakly larger than $\tilde{y}_{j+1} - h(\tilde{y}_{j+1}, \omega_{j+1}) - [\tilde{y}_j - h(\tilde{y}_j, \omega_j)]$. With $q_{j+1} = \beta_{Dj}(q_j)$, instead, the downward IC of ω_{j+1} workers is binding, $\tilde{\delta}_{j+1} - \tilde{\delta}_j = h(\tilde{y}_j, \omega_j) - h(\tilde{y}_j, \omega_{j+1})$. Hence, the welfare weights q_j and q_{j+1} satisfy

$$\int_{q_{j+1}}^{q_j} \frac{d\tilde{\delta}_{j+1}(\alpha)}{d\alpha_{j+1}} d\alpha \geq \tilde{y}_{j+1} - h(\tilde{y}_{j+1}, \omega_{j+1}) - [\tilde{y}_j - h(\tilde{y}_j, \omega_{j+1})] \ . \tag{A8}$$

Define $\zeta_4 := \max_{j \geq k,\ \alpha_j \in [1, \bar{q}]} d\tilde{\delta}_j(\alpha_j)/d\alpha_j > 0$. Then, the left-hand side of (40) is weakly smaller than $(q_j - q_{j+1})\zeta_4$. Due to $h_{y\omega} < 0$ and $h_{yy} > 0$, in contrast, the right-hand side is larger than

$$\begin{aligned} \tilde{y}_{j+1} - h(\tilde{y}_{j+1}, \omega_{j+1}) - [\tilde{y}(z) - h(\tilde{y}(z), z)] &= \int_{\tilde{y}_j}^{\tilde{y}(z)} [1 - h_y(\tilde{y}(\omega), \omega_{j+1})]\, d\omega \\ &> (\tilde{y}(z) - \tilde{y}_j)\ [1 - h_y(\tilde{y}(z), \omega_{j+1})] \end{aligned}$$

with $z \in (\omega_j, \omega_{j+1})$ and $\tilde{y}(z) = \arg\max y - h(y, z)$. Assumption 4 (i) implies that

$$\tilde{y}(z) - \tilde{y}_j > \int_{\omega_j}^{z} \nu_1 \frac{\tilde{y}(\omega)}{d\omega} d\omega > \nu_1 \tilde{y}_j \ln\left(\frac{z}{\omega_j}\right) \ ,$$

while Assumption 4 (ii) implies that

$$\begin{aligned} 1 - h_y(\tilde{y}(z), \omega_{j+1}) &= \int_{z}^{\omega_{j+1}} h_{yy}(\tilde{y}(\omega), \omega_{j+1}) \frac{\tilde{y}(\omega)}{d\omega} d\omega > \int_{z}^{\omega_{j+1}} \frac{\nu_1}{\nu_2} \frac{h_y(\tilde{y}(\omega), \omega_{j+1})}{\omega} d\omega \\ &> \frac{\nu_1}{\nu_2} h_y(\tilde{y}(z), \omega_{j+1}) \ln\left(\frac{\omega_{j+1}}{z}\right) \ . \end{aligned}$$

For $z = \omega_j a^{1/2}$, the right-hand side of (A8) is hence larger than

$$\zeta_5 := \frac{\nu_1^2}{\nu_2} (1/2 \ln a)^2 \min_{j \geq k} \tilde{y}_j h_y(\tilde{y}_j, \omega_{j+1}) > 0 \ .$$

Substituting all terms into (40), it follows that $q_j - q_{j+1} > \zeta_5/\zeta_4$ and, hence, bound away from 0 for all $j \geq 0$. As argued above, this ensures that there is a number $m \leq k + (\bar{q} - 1)\zeta_4/\zeta_5$ such that $q_j < 1$ for all $j \in [m, n]$.

Finally, the average welfare weight is exactly equal to 1 if

$$\begin{aligned} \sum_{j=1}^{n} f_j q_j &= \sum_{j=1}^{m-1} f_j E\left[q_j \mid j < m\right] + \sum_{j=m}^{n} f_j E\left[q_j \mid j \geq m\right] = 1 \\ \Leftrightarrow \sum_{j=m}^{n} f_j &= \bar{z} := \frac{E\left[q_j \mid j < m\right] - 1}{E\left[q_j \mid j < m\right] - E\left[q_j \mid j \geq m\right]} \end{aligned}$$

For $\sum_{j=m}^{n} f_j > \bar{z}$, there exists another (similar) weight sequence q' with average value 1 that satisfies the conditions in Proposition 3 and that is strictly decreasing with $q'_j > q'_{j+1}$ for all $j \in \{1, \ldots, n-1\}$. $\square$

C Supplementary material

Appendix C provides additional results and numerical analyses. Additionally, it provides a review of previews results in the optimal tax literature, a summary of empirical estimates of labor supply elasticities and further supplementary material.

C.1 Limit result

In the previous subsections, I have provided necessary and sufficient conditions for the optimality of an EITC in models with a discrete skill set. A natural question is whether the results also extend to a model with a continuum of skill types. The intuitive arguments provided in Subsection 4.1.3 imply that an EITC with negative marginal taxes at the bottom allows to reduce extensive-margin distortions at the cost of additional intensive-margin distortions. As shown above, this can only be optimal if the desire for redistribution among the poor is sufficiently small. In principle, this intuition does not depend on the properties of the skill set. However, the sufficient conditions in Propositions 1,2 and 4 are derived using formal arguments that exploit the discreteness of the skill set.

I can tackle this question, however, by studying the results of the model for the limit case where a discrete skill $\{\omega_1, \omega_2, \ldots, \omega_n\}$ converges to the set of all rational numbers in the interval $[\omega_1, \omega_n]$. For this purpose, I first impose the same functional-form assumptions on the utility function and the type distribution as in Subsection 3.1. Hence, I assume that the utility function is isoelastic as impied by (15) and that fixed costs are uniformly distributed according to (16). Second, I consider an economy with a large finite number n of skill groups and constant relative distances between adjacent skill types, $\omega_{j+1}/\omega_j = a > 1$ for all $j \in J_{-n}$, as in Subsection 3.2. The highest skill type follows as $\omega_n = \omega_1 a^{n-1}$. I denote by $F(\omega')$ the share of agents with skill types in the interval $[\omega_1, \omega']$. An economy in this class is characterized by a collection $(n, a, F, \sigma, \omega_1, \bar{\delta})$. In the following, I let the relative distance a shrink to 1 and the number n of skill groups go to infinity, while keeping the smallest and the largest skill type fixed. Hence, I consider the limit case where the skill set converges to an infinite number of evenly spaced skill types in the interval $[\omega_1, \omega_n]$. The following proposition demonstrates that the optimal marginal tax can be negative on a substantial income range even in this limit case.

Proposition A1. *Fix $\sigma > 0$, $\bar{\delta}$, ω_1 and ω_n. Consider three skill types ω_a, ω_b and ω_c in (ω_1, ω_n) such that $\tilde{y}(\omega_b) = 2\tilde{y}(\omega_a)$ and $\tilde{y}(\omega_c) = 4\tilde{y}(\omega_a)$. Let a and n converge to 1 and to ∞, respectively, in such a way that $\omega_n = \omega_1 a^{n-1}$ stays constant. In the limit, there are strictly decreasing welfare weights α such that optimal labor supply is upwards distorted at the intensive margin for each skill group $\omega \in (\omega_1, \omega_a)$ if*

$$\frac{1 - F(\omega_c)}{F(\omega_b)} \geq 2 \, . \tag{A9}$$

The formal proof is available on request. Proposition A1 clarifies that an EITC with negative marginal taxes can be optimal even if the limit case where the discrete skill set converges to a continuum. More specifically, it asks whether a tax with strictly negative marginal taxes for all incomes below some phase-in endpoint $y(\omega_a)$ can be optimal. If condition (A9) is satisfied, then the answer is affirmative: the tax schedule is indeed optimal for some strictly decreasing welfare weights α. As in Proposition 2 for the stylized model with three skill types and in

Proposition 4 for the general discrete model, the sufficient condition is expressed in term of the skill distribution; it requires that there exists a sufficiently large share of agents that are sufficiently more productive than those workers with incomes in the phase-in range. In particular, this condition compares, first, the shares of high-skill agents with conditionally optimal incomes above $\tilde{y}(\omega_c) = 4\tilde{y}(\omega_a)$ and, second, the share of low-skill agents with conditionally optimal incomes below $\tilde{y}(\omega_b) = 2\tilde{y}(\omega_a)$ (intuitively, this subset includes both the agents in the phase-in range and in the phase-out range of the EITC). An EITC with phase-in endpoint $\tilde{y}(\omega_a)$ can be rationalized whenever the share of high-skill workers $1 - F(\omega_c)$ is at least twice as large as the second share $F(\omega_b)$ of low-skill workers. Remarkably, this condition does not depend on the level of the intensive-margin elasticity $\sigma > 0$.

Implicitly, Proposition A1 also provides a lower bound on the location of phase-in endpoints that can be rationalized with decreasing welfare weights. For an economy with a given skill distribution F, we can easily determine the highest skill level ω_a that still satisfies condition (A9).[1] As an illustration, consider the subgroup of childless singles in the US in 2015, which is also considered in the numerical simulations in Section 5. As explained in more detail below, I can estimate the skill distribution in this subgroup using data from the March 2016 CPS. Based on this data, Proposition A1 implies that an EITC with a phase-in endpoint up to an income of \$12,600 annual income and a phase-out endpoint up to an income of \$25,200 can be rationalized. Note that these lower bounds are substantially above the actual 2015 phase-in and phase-out endpoints of \$6,580 and \$14,820, respectively.

C.2 Construction of weight sequences for simulations

Figure A1 below plots the exogenous welfare weights α^A and α^B, which are used in the numerical simulation in Section 5. In particular, it plots the welfare weights α_j^A and α_j^B associated to each skill group $j \in J$ (on the vertical axis) against the skill-specific gross income y_j^α under the current US income tax (on the horizontal axis). The figure focuses on skill groups such that the conditionally optimal income under the current US tax system is below \$100.000.

The blue solid line shows sequence α^A, which is constructed by setting $\alpha_j^A = 1.04$ for all skill types with a conditionally optimal income below \$11,000 under the current US tax system. This condition is satisfied for the 49 lowest skill groups, who constitute the lowest 5.6% of the skill distribution according to CPS data. For each skill group $j \in (49, 74)$, I set the weight α_j to the lowest value such that the first-and-half-best allocation still satisfies the downward IC of workers with skill type ω_j. Formally, this implies that $\alpha_j^A = \beta_{j-1}^D(\alpha_{j-1}^A)$. The welfare weight of each skill group $j \geq 75$ is set to a number $\underline{\alpha}$ that implies an average welfare weight of 1. The implied weight sequence α_j is monotonically decreasing over the skill distribution. Obviously,

The green dashed line shows sequence α^B. It associates a welfare weight of 1.04 to all agents with skill type ω_j such the individually optimal annual income under the current US tax is smaller than \$26,000 (conditional on working). This includes approximately the lowest-skilled quartile of the population according the CPS data (26.3%). To ensure an average welfare weight of 1, the welfare weight of all higher-skilled agents is set to a number close to 0.986.

[1] Note that both ω_b and ω_c are increasing with ω_a by construction. Hence, the left-hand side of condition (A9) is strictly decreasing in ω_a. This implies that the maximum of skill levels ω_a that satisfy condition (A9) is well-defined.

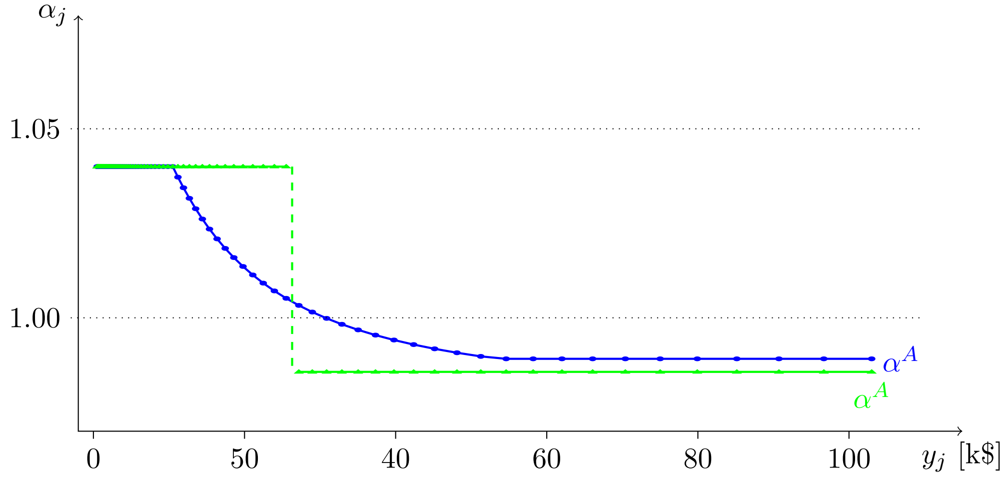
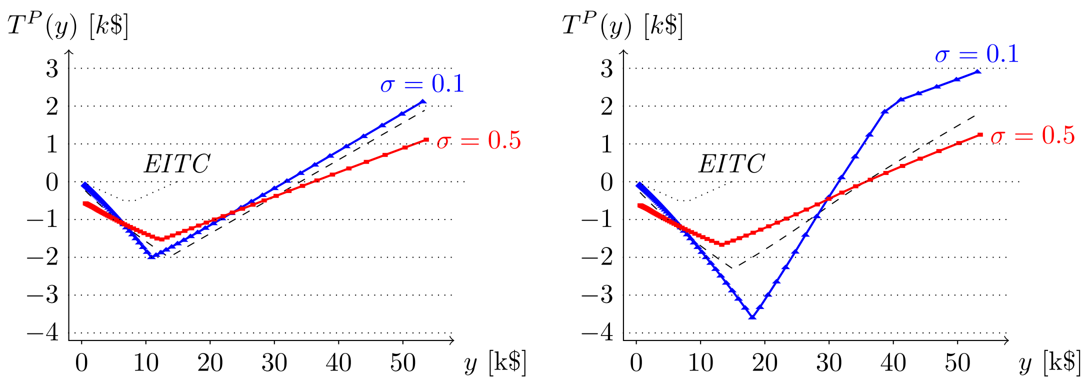


Figure A1: Welfare weight sequence α^A and α^B.

C.3 Sensitivity analysis

The following figures depict the optimal tax schedules for the benchmark calibration and a number of alternative parameter choices. In particular, each figure shows that simulated participation tax $T^P(y) = T(y) - T(0)$ that is optimal given the sequence of welfare weights α^A (left panel) and the sequence of welfare weights α^B (right panel).

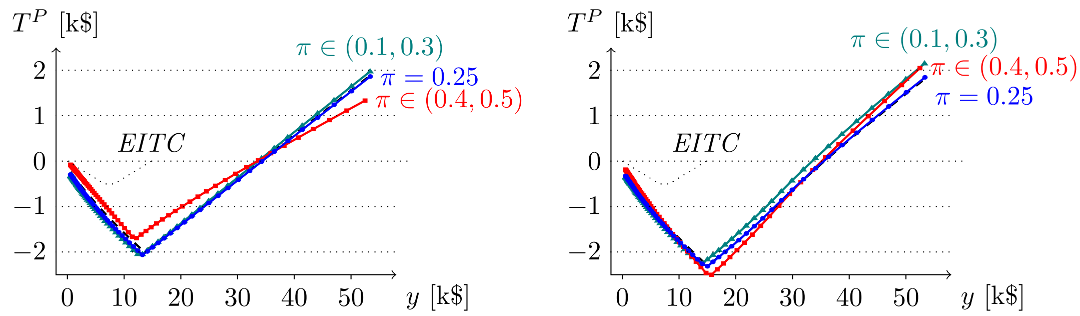
Figure A2 shows how variations in the intensive-margin elasticity affect the optimal tax schedule for both weight sequences. In both panels, the benchmark calibration with parameter value $\sigma = 0.3$ is depicted by the black dashed line. The blue lines with triangles depict the optimal participation taxes that arise for a smaller elasticity of $\sigma = 0.1$. The red lines with squares depict the optimal participation taxes that result for a larger elasticity $\sigma = 0.5$. Both figures show that a higher elasticity leads to a flatter tax schedule, both in the phase-in and in the phase-out income ranges.



(a) Optimal tax for weight sequence α_A (b) Optimal tax for weight sequence α_B.

Figure A2: Optimal taxes for different intensive-margin elasticities.

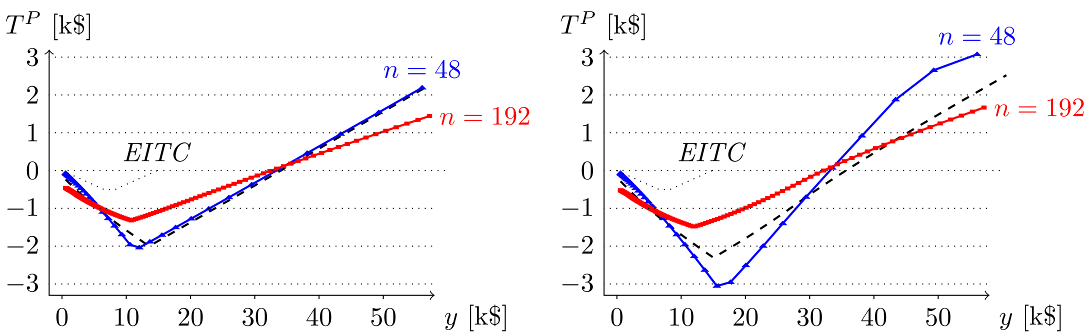
Figure A3 shows how variations in the participation elasticities affect the optimal tax schedule for both weight sequences. In both panels, the benchmark simulations for participation elasticities that are falling from 0.4 in the lowest skill groups to 0.18 in the highest skill groups are depicted by black dashed line. The teal lines with triangles depict the optimal taxes that arise for smaller participation elasticities that are falling from 0.3 to 0.1. The red lines with squares depict the optimal taxes that result for larger elasticities that are falling from 0.5 to 0.4 as in the benchmark calibration by Jacquet et al. (2013) (to be consistent with this calibration, I also set the intensive-margin elasticity to 0.25 instead of 0.3). Finally, the blue lines with circles depict the simulated taxes that arise if participation elasticities are equal to 0.25 in all skill groups. Both figures show that a variations in the average level and in the skill gradient of participation elasticities have only minor effects on the properties of the optimal income tax.



(a) Optimal tax for weight sequence α_A (b) Optimal tax for weight sequence α_b

Figure A3: Optimal taxes for different participation elasticities.

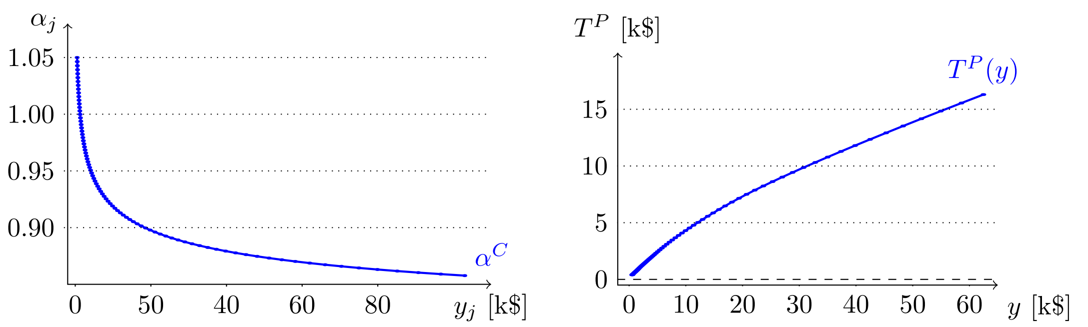
Figure A4 shows how the discretization of the skill set affects the optimal tax schedule for both weight sequences. In both panels, the benchmark simulations for a skill set with 96 are depicted by black dashed line. The relative distance between each pair of adjacent skill types is equal with $\omega_{j+1}/\omega_j = 1.05$. To study the effects of the discretization, I consider two alternative calibrations with skill sets such that the lowest and the highest skill types are identical to the benchmark calibration, but the number of skill types and the relative distance between adjacent types are varied. For each discretization, I have re-estimated the skill distribution based on March 2016 CPS data. In particular, the blue lines with triangles depict the optimal taxes that arise for coarser skill set with 48 skill types. The red lines with squares depict the optimal taxes that result for a finer skill set with 192 skill types. Both figures show that a finer discretization of the skill set has a similar effect as an increase in the intensive-margin elasticity, leading to flatter tax schedules both in the phase-in and the phase-out income ranges.



(a) Optimal tax for weight sequence α_A (b) Optimal tax for weight sequence α_B

Figure A4: Optimal taxes for different discretizations of the skill set.

Finally, I vary the sequence of welfare weights. In particular, I assume that the weight of skill group j is given by $\alpha_j^C = 3/4 + 0.3(\omega_1/\omega_j)^{1/3}$. Figure A5a depicts these welfare weights by plotting them against the conditionally optimal income levels of the different skill groups under the current US tax. As can be seen, the welfare weights are convexly decreasing over the skill distribution as in the calibrations of Saez (2002) and Jacquet et al. (2013). Figure A5b plots the optimal income tax that results from my numerical simulations. In contrast to all previous figures, both participation taxes and marginal taxes are strictly positive at all income levels (below the very top).



(a) Welfare weights α^C

(b) Optimal tax schedule for sequence α^C

Figure A5: An example with strictly positive marginal taxes.

C.4 Optimality of positive marginal taxes

Above, I have shown numerically that the results of Saez (2002) and Jacquet et al. (2013) are recovered in my model if the skill gradient of welfare weights at the bottom is large. To complement this numerical result, I provide an analytical condition for the optimality of strictly positive marginal taxes.

Proposition A2. *Let Assumptions 2, 3 and 4 be satisfied and let the sequence of welfare weights be given by α. Optimal output is downwards distorted at the intensive margin in all skill groups $j \in J_{-n}$ if $\alpha_{j+1} \leq \beta_{Dj}(\alpha_j)$ for all $j \in J_{-n}$ with at least one strict inequality, where function β_{Dj} is defined in Lemma 3.*

Proposition A2 extends the main result of Jacquet et al. (2013) to a model with a discrete instead of a continuous skill set. A formal proof is available on request. Basically, the proposition says that optimal labor supply is downwards distorted at the intensive margin if the first-and-half-best allocation violates the downward IC constraints of the workers in all skill groups $\{2, \ldots, n\}$. The proof that, in this case, every single downward IC constraint is unambiguously binding in the optimal allocation. By Lemma 3, the condition $\alpha_{j+1} \leq \beta_{Dj}(\alpha_j)$ is satisfied if either the difference between the welfare weights α_j and α_{j+1} is large enough, or the difference between both welfare weights and the average weight of 1 is large enough.

C.5 Optimal income taxes for single parents

The following section provides the results of numerical simulations for a second subgroup of the US population, single parents. To calibrate the model for this population group, I consider the same functional forms as for childless singles: (a) utility (1) is quasilinear, (b) the effort cost function (15) gives rise to a constant elasticity at the intensive margin, (c) the distributions of fixed costs (23) are logistic in each skill group, (d) the participation elasticity π_j and the participation share π_j vary along the skill distribution according to (24) and (25), respectively. The last point implies that parameters μ, $\underline{\pi}$ and $\bar{\pi}$ have to be set in order to match empirically plausible labor supply elasticities for single parents.

For this purpose, I again follow the survey by Saez et al. (2012), the meta-study by Chetty, Guren, Manoli & Weber (2013) and the study by Chetty, Friedman & Saez (2013) on EITC recipients. With respect to the intensive margin, there is no convincing evidence for differences between single parents and other population groups. Hence, I again set the intensive-margin parameter to $\sigma = 0.3$.

With respect to the extensive margin, in contrast, the literature suggests substantial differences across population groups. In particular, single mothers are commonly perceived as the group that is most responsive at the participation margin (along with married women). Early studies estimate extremely large elasticities for single mothers receiving the EITC, sometimes even above 1 (see, e.g., Meyer & Rosenbaum 2001). Chetty, Guren, Manoli & Weber (2013) review a number of studies based on quasi-experimental variation and re-calculate the elasticities to be consistent with each other. The resulting elasticities in those studies that focus on single mothers are in the range of 0.43, which is still larger than for other population groups. The two most recent studies exploiting EITC reforms point towards smaller levels: Bastian & Jones (2019) estimate a participation elasticity of 0.33 for single mothers, while Kleven (2019) is a clear outlier in the literature, finding an elasticity close to 0.

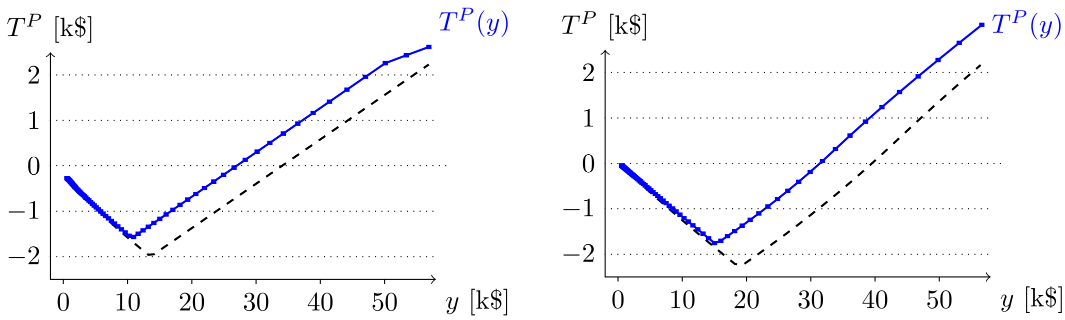
Based on this literature, I set the parameters $\underline{\pi}$ equal to 0.6 and $\bar{\pi}$ equal to 0.3. This gives an average participation elasticity of 0.42, close to the suggestion of Chetty, Guren, Manoli & Weber (2013). In particular, this also implies that elasticities are falling along the skill distribution, in line with the existing evidence for all subgroups of the population. As the participation rate of single parents is 82.1% in the CPS data (similar to the one of childless singles), I set $\underline{L} = 0.7$ and $\bar{L} = 0.85$.

To calibrate the skill distribution of single parents, I again use data from the March 2016 CPS. For this purpose, I consider only single parents at ages 25 to 60 that do neither live with an unmarried spouse nor with any family member except for own children below the age of 18. This ensures that there is no joint labor supply decision problem. The restricted sample contains $7,141$ observations. In the CPS data, single parents are on average less productive than childless singles (average annual income $\$32,958$ compared to $\$42,223$ for childless singles). To back out the skill distribution, I again use the same linear approximation of the US tax system.[2] I can then use the first-order condition for an individual optimum to assign skill levels to all single parents with positive incomes. In particular, I use the same skill set with $n = 96$ skill types and estimate the share of single parents in each skill group based on a kernel density estimation.

To calibrate the planner's redistributive preferences, I cannot simply use the same weight

[2]For single parents, this approximation is less convincing than for childless singles, especially at low income levels where the effective marginal tax depends on the statutory tax schedule, the EITC phase-in and phase-out rates and the marginal reduction rates of welfare transfers. As shown by Maag et al. (2012), the effective marginal tax varies strongly across states and income levels. For example, if a single parents moves from zero income to the federal poverty line, he faces an average marginal tax between a minimum of -13.3% and a maximum of 25.5%. Between the poverty level and twice the poverty level, the average marginal tax is above 40% in most states. Given the lack of superior alternatives, I stick to the same linear approximation as for childless singles.

sequences as for childless singles. As the skill distribution varies across subgroups, the average weight among single parents would differ from 1. Hence, I again construct two exogenous sequences of welfare weights that are monotonically decreasing and have an average value of 1 (within the group of single parents). First, I consider a sequence $\tilde{\alpha}^A$ that assigns a welfare weight of 1.04 to the agents in the lowest 45 skill groups, i.e., to all working agents with incomes below $\$9,050$. The welfare weights of higher-skilled agents are assumed to be gradually decreasing (as with sequence α^A for childless singles). Second, I consider a sequence $\tilde{\alpha}_B$ that assigns a welfare weight of 1.03 to the lowest 66 skill groups. These agents constitute the lower-skilled half of the single parent subpopulation (more precisely: 52.3%) and include all working agents with incomes below $\$31,000$. The average weight of the higher-skilled agents is set to be constant as well and equal to 0.967. As in the case of childless singles, welfare weights $\tilde{\alpha}_A$ satisfy the sufficient conditions for an EITC with negative marginal taxes, while welfare weights $\tilde{\alpha}_B$ only satisfy the necessary conditions.



(a) Optimal tax for weight sequence $\tilde{\alpha}_A$

(b) Optimal tax for weight sequence $\tilde{\alpha}_B$

Figure A6: Optimal income taxes for single parents.

The simulation results are depicted in Figures A6a for sequence $\tilde{\alpha}^A$ and A6b for sequence $\tilde{\alpha}^B$. In particular, the solid blue lines in both figures show the optimal participation taxes for single parents. For sequence α^A, an EITC with negative marginal taxes for incomes up to $\$10,935$, a maximum tax credit of $\$1,566$ and negative participation taxes for incomes up to $\$26,574$ is optimal. For sequence $\tilde{\alpha}^A$, the optimal tax involves an EITC with negative marginal taxes for incomes up to $\$15,017$, a maximum tax credit of $\$1,753$ and negative participation taxes for incomes up to $\$29,912$ is optimal. The transfers to non-working agents only reach levels of $\$1,231$ and $\$1,573$, respectively. However, these transfers have to interpreted as transfers paid by higher-skilled single parents only. If the social planner has a desire for redistribution from other population groups to the group of single parents, then there is an additional lump-sum transfer towards all single parents. With quasi-linear utility, however, these between-subgroup transfers only imply parallel shifts in the budget sets; they do not affect the optimal levels of marginal taxes and participation taxes.

These simulated tax schedules have to be interpreted as the optimal integrated schemes that result from the combination of income taxes and welfare transfers (and, potentially, child care costs). As argued above, the integrated tax-transfer systems effectively imply a larger "net EITC" in some US states and a smaller one in other states (see Maag et al. 2012). Hence, my simulation results can be broadly interpreted as providing support for the current levels of the EITC for single parents. They do not provide support for another substantial extension of the

EITC, however.

For comparison, Figures A6a and A6b also depict by dashed lines the optimal tax schedules that would result for childless singles given similar welfare sequences. According to these simulation results, the optimal EITC should be larger for childless singles than for single parents. This is probably in contrast to the current US policies, although single parents in the US do not only benefit from a more generous EITC but also face large positive reduction rates from a number of welfare transfers such as TANF, SNAP, Public Housing etc. The main reason for the different simulation results is that the share of high-skilled agents is much larger in the group of childless singles than in the group of single parents. As explained in Subsection 3, this implies that a generous EITC can more easily be financed in the group of childless singles without imposing strong distortions at the intensive margin. In contrast, the larger participation elasticities of single parents seem to have only a negligible effect on the optimal size of the subgroup EITC.

C.6 Generalized welfare weights and poverty alleviation

In this section, I discuss whether an EITC is the optimal policy for alleviating poverty. Following Mirrlees (1971), optimal income taxation is commonly studied under the assumption that the social objective is given by a (utilitarian) welfare function (see Weinzierl 2014 for a recent critique). As an alternative, Kanbur et al. (1994) as well as Besley & Coate (1992, 1995) suggest the goal of alleviating poverty as measured by the available income. They advocate this objective as being more consonant with public debates and, consequently, as providing better insights into real-world policy choices. In particular, they argue that neither policy-makers nor taxpayers seem to value the leisure enjoyed by the poor (which is an argument of standard utility functions), but rather seem to focus on income as a more visible sign of poverty (see, e.g., Besley & Coate (1995): 189 and Kanbur et al. (1994): 1615-1616).

Support for this view comes from the recent public debate surrounding a potential expansion of the EITC for childless workers. Most prominently, President Barrack Obama and Paul Ryan, then Republican Chairman of the House of Representative Budget Committee, independently proposed to expand the EITC by doubling the phase-in and phase-out rates, raising the phase-out start and the eligibility threshold, and relaxing age restrictions for childless workers (Executive Office 2014, House Budget Committee 2014). The Obama proposal emphasizes the goal to reduce poverty for childless low-income workers. In particular, the proposal estimates that "the increase in the credit would lift about half a million people above the poverty line and reduce the depth of poverty for 10 million more" (Executive Office 2014: 2). It also criticizes that the current US tax code pushes childless workers with low incomes "into or deeper into poverty" (Executive Office 2014: 3), both directly and indirectly through discouraging work.[3] The Ryan proposal suggests a number of reforms to reduce poverty and increase economic self-sufficiency. It argues that the EITC is the most successful program in fighting poverty among families, and that its expansion would significantly reduce poverty among childless workers. The Ryan proposals also emphasizes that an EITC expansion would provide greater incentives for people to work and "earn enough money to place them above the poverty line" (House Budget Committee 2014:

[3]Besides, the proposal argues that an EITC expansion would increase employment rates, and that this might benefit society through positive external effects such as increasing marriage rates, supporting child outcomes and reducing incarceration rates (Executive Office 2014: 9).

7). A number of further proposals provide similar arguments for an even more generous EITC expansion, emphasizing the goals of lifting people above the poverty line, reducing the depth of poverty for others and, in particular, eliminating the possibility that low-income workers are taxed into poverty. For example, such poverty-related arguments were made to support two recent proposals for EITC expansion, introduced in the House of Representatives in February 2017, and in the Senate in June 2017.[4]

I continue by sketching the most common poverty measures and their formalization in optimal tax problems. Let $\bar{c}$ denote the poverty line, expressed in terms of consumption or available income (after tax and transfers).[5] Foster et al. (1984) introduce a class of poverty measures given by

$$P_a(c, \bar{c}) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\bar{c} - c_i}{\bar{c}} \right)^a \mathbb{1}_{c_i \leq \bar{c}} \, .$$

where n is the number of agents in the population and c_i is agent i's available income. For $a = 0$, this measure is equal to the poverty rate or head count ratio, the share of a population with available incomes below $\bar{c}$. For $a = 1$, it is equal to the poverty gap, the average (percentage) shortfall of available income from the poverty line. In contrast to the poverty rate, the poverty gap also accounts for the intensity (depth) of poverty. For $a > 1$, the measure assigns higher weights to larger shortfalls from the poverty line. Institutions such as the World Bank and the United Nations commonly use the first two measures, the poverty rate and the poverty gap.

Kanbur et al. (1994) were the first to study optimal income taxation under the goal of poverty alleviation instead of welfare maximization. They formalize the objective of poverty alleviation by using what they call a generalized poverty gap measure (with $a > 1$). For an intensive-margin model, they find that the optimal marginal tax might be negative at incomes below the poverty line. Numerical simulations show large positive marginal taxes to be optimal even at the very bottom, however. They also point out that their formalization of poverty alleviation is inconsistent with the Pareto principle, because it fails to account for the agents' disutility of providing output instead of enjoying leisure. To reconcile the poverty gap criterion with the Pareto principle, Saez & Stantcheva (2016) suggest to apply their approach of generalized welfare weights by setting welfare weights equal to $\alpha_p > 1$ for all agents with consumption below the poverty line $\bar{c}$, and equal to $\alpha_{np} \in [0, 1)$ for all agents with consumption above $\bar{c}$. In my numerical simulations, I have considered sequence α_B of welfare weights, which has exactly this shape. As shown in Subsection 5, the optimal tax is given by an EITC with negative marginal taxes and negative participation taxes in my model with labor supply responses at two margins. With intensive-margin responses alone, in contrast, the optimal marginal taxes for this criterion are strictly positive both below and above the poverty line $\bar{c}$ (Saez & Stantcheva 2016).

C.7 Perturbation approach and optimal tax formula

One common method to solve for the optimal income tax is the perturbation approach, introduced by Piketty (1997) and Saez (2001). Jacquet et al. (2013) and Lorenz & Sachs (2016) use this approach in a model with two margins and a continuum of skill (or income) levels. The

[4]For the details of both bills, see *https://www.congress.gov/bill/115th-congress/house-bill/822* and *https://www.congress.gov/bill/115th-congress/senate-bill/1371*.

[5]Sometimes, the poverty line is also defined in terms of pre-tax income.

analysis starts by fixing an initial tax schedule T and the income distribution F_y it implements. It then considers a perturbation of the tax system that increases the marginal tax T' by a small amount τ on a small interval $(y', y' + \ell)$. For the model introduced above, the welfare effect of a small reform with $\tau \to 0$ and $\ell \to 0$ is zero if and only if the marginal tax at income level y' satisfies

$$\frac{T'(y')}{1 - T'(y')} = \frac{1 - F_y(y')}{\tilde{\varepsilon}_T(y')\ y'\ f_y(y')} \left[1 - \bar{\alpha}_T(y') - \frac{1}{1 - F_y(y')} \int_{y'}^{\bar{y}} f_y(y)\ \tilde{\eta}_T(y)\ T^P(y) dy\right], \qquad \text{(A10)}$$

where $\bar{\alpha}_T(y')$ is the average welfare weights associated to agents with incomes above y' under tax schedule T.[6] If tax schedule T is optimal, it satisfies equation (A10) at any income level $y' \in \left[\underline{y}, \bar{y}\right]$. Note that the formula does not account for income effects, which are ruled out in my model by the assumption of quasi-linear preferences. Jacquet et al. (2013) and Lorenz & Sachs (2016) provide generalized versions of this optimal tax formula that allow for income effects. Saez (2002) provides a discretized version of (A10), where the right-hand side contains instead of the integral the sum $\sum_{y'=y}^{\bar{y}} f_y(y)\eta(y)T^P(y')$ over a finite number of income levels (interpreted as occupations).

Equation (A10) has a similar structure as the well-known *ABC* formula for the intensive model (Diamond 1998). The only difference is given by the integral term on the right-hand side, which accounts for the extensive-margin responses at income levels above y'. Due to this integral term, however, the optimal tax formula for the two-margin model is substantially less informative about the qualitative properties of the optimal tax schedule T than the *ABC* formula. To see this, note that the optimal marginal tax at income y' also depends on the values of $\tilde{\eta}$, the income density f_y and the optimal participation tax T^P at all income levels above y'. Most importantly, the optimal participation tax T^P is an endogenous object – more precisely, the object of interest in the optimal tax problem. Hence, we cannot determine T' at any income level without precise knowledge of the entire tax function. Formally, (A10) represents a differential equation with an integral term, which in general cannot be solved analytically. This complication makes it even hard to determine the optimal sign of $T'(y)$. On the right-hand side of (A10), all terms in front of the bracket and the difference $1 - \bar{\alpha}(y)$ are positive for any $y \in \left[\underline{y}, \bar{y}\right]$. However, we have to subtract the extensive-margin terms under the integral, which is also positive in the plausible case that $T^P(y) > 0$ for most high-skill workers. But then, the optimal sign of $T'(y)$ depends on whether the term $1 - \bar{\alpha}(y)$ or the integral term is larger, given the optimal participation tax $T^P(y)$ for all $y > y'$.[7]

Based on Saez (2002), some papers argue that a negative marginal tax is more likely to be optimal if labor supply responds *more strongly* at the extensive margin than at the intensive margin. Indeed, an increase in the semi-elasticity $\eta(y)$ decreases the right-hand side of (A10) if $T^P(y)$ is positive and held constant. However, $T^P(y)$ itself varies with η in a non-trivial way.[8]

[6] Note that $\tilde{\varepsilon}_T$ and $\tilde{\eta}_T$ denote the intensive- and extensive-margin elasticities, respectively, as functions of the income chosen under tax T, rather than functions of the skill type ω (see also the difference between the optimal tax formulas in Jacquet et al. (2013) and Lorenz & Sachs (2016)).

[7] Without the integral term, the optimal marginal tax can be determined point-wise at each income level: It only depends on the values of $\tilde{\varepsilon}$, $\bar{\alpha}$ and the hazard rate at income y'. In particular, the optimal marginal tax is unambiguously positive if $\tilde{\alpha}$ is monotonically decreasing in income.

[8] In the extensive model, $T^P(y)$ is decreasing (increasing) in $\tilde{\eta}(y)$ if $\alpha(y)$ is above (below) 1 (see

Moreover, the intensive-margin elasticity $\varepsilon(y')$ has only a direct effect on the absolute value of the right-hand side, not on its sign; the indirect effects on T^P are not obvious. To conclude, it is hard to sign the optimal marginal tax T' based on the optimal tax formula (A10). This problem cannot even be alleviated by imposing structural assumptions on the labor supply elasticities ε and η.

C.8 The mechanism approach with a continuous skill set

Lemma 1 provides a simplified expression of optimal allocation problem that is valid both with a discrete and a continuous skill set. In the case of a continuous skill set, the dimensionality of the problem can be reduced further by rewriting the incentive compatibility constraints in terms of an envelope condition. Then, the participation threshold in each skill group ω is given by

$$\hat{\delta}(\omega) = \tilde{\delta}(\omega, \delta_0, y) := \delta_0 - \int_{\underline{\omega}}^{\omega} h_\omega(y(\omega'), \omega') d\omega' \ , \tag{A11}$$

where δ_0 is the participation threshold of the lowest skill type $\underline{\omega}$. Plugging (A11) into equation (10) allows to write the optimal tax problem as the unconstrained problem to maximize

$$\begin{aligned} \tilde{W}(\delta_0, y) \quad &= \quad \int_{\omega \in \Omega} f(\omega) G_\delta(\tilde{\delta}(\omega, \hat{\delta}_0, y) \mid \omega) \left[y(\omega) - h(y(\omega), \omega) + (\alpha(\omega) - 1)\tilde{\delta}(\omega, \hat{\delta}_0, y) \right] d\omega \\ &\quad - \int_{\omega \in \Omega} f(\omega) \alpha(\omega) \int_{\underline{\delta}}^{\tilde{\delta}(\omega, \hat{\delta}_0, y)} \delta g_\delta(\delta \mid \omega) d\delta d\omega \end{aligned} \tag{A12}$$

over δ_0 and the output function $y : \Omega \to \mathbb{R}$.

Dealing with an unconstrained maximization problem seems straightforward. Indeed, we can easily use point-wise differentiation with respect to $y(\omega)$ to get the first-order condition

$$\begin{aligned} f(\omega') G(\hat{\delta}(\omega') \mid \omega') \frac{1 - h_y(y(\omega'), \omega')}{h_{y\omega}(y(\omega'), \omega')} &= \\ \int_{\omega'}^{\bar{\omega}} f(\omega) G(\hat{\delta}(\omega) \mid \omega) \left\{ \alpha(\omega) - 1 + \frac{g(\hat{\delta}(\omega) \mid \omega)}{G(\hat{\delta}(\omega) \mid \omega)} \left[y(\omega) - h(y(\omega), \omega) - \hat{\delta}(\omega) \right] \right\} d\omega \ . \end{aligned} \tag{A13}$$

The term $1 - h_y(y(\omega'), \omega')$ on the left-hand side captures the intensive-margin distortion and the *implicit marginal tax* at income level $y(\omega')$ in terms of the model's primitives, while the term $y(\omega) - h(y(\omega), \omega) - \hat{\delta}(\omega)$ at the right-hand side captures the extensive-margin distortions and the *implicit participation tax* at income $y(\omega)$.

As in the case of the optimal tax formula (A10) derived from the perturbation approach, however, the sign of either side cannot be determined for skill type ω' in isolation. By equation (A13), the intensive-margin distortions in skill group ω' rather depends on the extensive-margin distortions in all skill groups above ω' (and vice versa). As a result, the first-order condition (A13) is not directly informative about the optimal signs of marginal taxes and participation taxes. Actually, the first-order condition (A13) only differs from the optimal tax formula (A10) derived in Subsection C.7 in that the former is expressed in terms of the model's primitives, while the latter is expressed in terms of empirically estimable objects.

Diamond (1980), Saez (2002), Choné & Laroque (2011)).

C.9 Empirical evidence on labor supply elasticities

In Section 5, I calibrate my model to match empirical moments for the subgroup of childless singles in the US. This requires to choose calibration targets for the labor supply elasticities at both margins. The current state of the empirical literature is summarized in the survey by Saez et al. (2012) and the meta-study by Chetty, Guren, Manoli & Weber (2013). Unfortunately, there is no clear consensus on the empirical levels of elasticities at both margins. In particular, participation elasticities seem to differ substantially across subgroups of the population. For my calibration, I mainly use the preferred estimates suggested by Chetty, Guren, Manoli & Weber (2013) and Saez et al. (2012). Besides, I consider the studies by Chetty, Friedman & Saez (2013) on the elasticities on EITC recipients with different marital status and by Bargain et al. (2014) on the elasticities of childless singles as particularly informative for my purposes.

With respect to the intensive-margin elasticity, Saez et al. (2012) consider the best available estimates to be in the range between 0.12 and 0.4. Chetty, Guren, Manoli & Weber (2013) suggest 0.33 as their preferred estimate, and Chetty, Friedman & Saez (2013) estimate an average elasticity for EITC recipients of 0.21 (wage earnings) and 0.36 (total earnings). Bargain et al. (2014) estimate an elasticity of 0.13 for childless singles, with somewhat higher levels among low-skill workers.

With respect to the extensive-margin elasticity, Chetty, Guren, Manoli & Weber (2013) suggest 0.25 as their preferred estimate for the population average based on a meta-analysis. They also emphasize that elasticities are probably higher in groups such as single mothers who have a lower attachment to the labor force. For childless singles in the US, Bargain et al. (2014) estimate an average elasticity around 0.24. Besides, most studies find that participation elasticities are strictly decreasing along the income distribution (see Juhn et al. 1991, 2002 for the US and Meghir & Phillips 2010 for the UK). Miller et al. (2018) and Bastani et al. (2019) provide evidence that participation responses are even decreasing within the group of low-income workers. In particular, Miller et al. (2018) document the results from a recent randomized control trial that studied the effects of a substantial increase in the EITC for childless workers, finding much larger responses among the most vulnerable subgroups. Bastani et al. (2019) exploit a tax reform in Sweden to estimate the extensive-margin responses of married females from low-income families. In contrast, Bargain et al. (2014) find participation elasticities to be by and large constant along the income distribution.

Summing up, there remains a considerable uncertainty about the levels of labor supply responses at both margins. In my benchmark simulations for childless singles, I mainly follow the suggestions of Chetty, Guren, Manoli & Weber (2013). In particular, I consider an intensive-margin elasticity of 0.3 in all skill groups. With respect to the extensive margin, I assume that participation elasticities are decreasing from 0.4 in the lowest skill groups to 0.18 in the highest skill groups. This gives an average elasticity of 0.25, the preferred value of Chetty, Guren, Manoli & Weber (2013).

References

Bargain, O., Dolls, M., Neumann, D., Peichl, A. & Siegloch, S. (2014), 'Comparing inequality aversion across countries when labor supply responses differ', *International Tax and Public Finance* **21**(5), 845–873.

Bastani, S., Moberg, Y. & Selin, H. (2019), 'The Anatomy of the Extensive Margin Labor Supply Response', *Scandinavian Journal of Economics* (forthcoming).

Bastian, J. & Jones, M. R. (2019), Do EITC Expansions Pay for Themselves? Effects on Tax Revenue and Public Assistance Spending, Working paper. URL: *http://jacobbastian.squarespace.com/research*.

Besley, T. & Coate, S. (1992), 'Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs', *American Economic Review* **82**(1), 249–261.

Besley, T. & Coate, S. (1995), 'The design of income maintenance programmes', *Review of Economic Studies* **62**(2), 187–221.

Chetty, R., Friedman, J. N. & Saez, E. (2013), 'Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings', *American Economic Review* **103**(7), 2683–2721.

Chetty, R., Guren, A., Manoli, D. & Weber, A. (2013), 'Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities', *NBER Macroeconomics Annual* **27**(1), 1 – 56.

Choné, P. & Laroque, G. (2011), 'Optimal taxation in the extensive model', *Journal of Economic Theory* **146**(2), 425–453.

Diamond, P. (1980), 'Income taxation with fixed hours of work', *Journal of Public Economics* **13**(1), 101–110.

Diamond, P. A. (1998), 'Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates', *The American Economic Review* **88**(1), 83–95.

Executive Office of the President and U.S. Treasury Department (2014), The President's proposal to expand the Earned Income Tax Credit, March 3, 2014. URL: *https://obamawhitehouse.archives.gov/sites/default/files/docs/eitc_report_0.pdf*.

Foster, J., Greer, J. & Thorbecke, E. (1984), 'A class of decomposable poverty measures', *Econometrica* **52**(3), 761–766.

House Budget Committee Majority Staff, Chairman Paul Ryan (2014), Expanding opportunity in America. A discussion draft from the House Budget Committee, July 24, 2014. URL: *https://republicans-budget.house.gov/uploadedfiles/expanding_opportunity_in_america.pdf*.

Jacquet, L., Lehmann, E. & Van der Linden, B. (2013), 'Optimal redistributive taxation with both extensive and intensive responses', *Journal of Economic Theory* **148**(5), 1770–1805.

Juhn, C., Murphy, K. M. & Topel, R. H. (1991), 'Why has the natural rate of unemployment increased over time?', *Brookings Papers on Economic Activity* **1991**(2), 75–142.

Juhn, C., Murphy, K. M. & Topel, R. H. (2002), 'Current unemployment, historically contemplated', *Brookings Papers on Economic Activity* **2002**(1), 79–136.

Kanbur, R., Keen, M. & Tuomala, M. (1994), 'Optimal non-linear income taxation for the alleviation of income-poverty', *European Economic Review* **38**, 1613–1632.

Kleven, H. (2019), The EITC and the Extensive Margin: A Reappraisal, Working Paper 26405, National Bureau of Economic Research. URL: *http://www.nber.org/papers/w26405*.

Lorenz, N. & Sachs, D. (2016), 'Identifying laffer bounds: A sufficient-statistics approach with an application to germany', *Scandinavian Journal of Economics* **118**(4), 646–665.

Maag, E., Steuerle, C. E., Chakravarti, R. & Quakenbush, C. (2012), 'How marginal tax rates affect families at various levels of poverty', *National Tax Journal* **65**(4), 759–782.

Meghir, C. & Phillips, D. (2010), Labour supply and taxes, *in* J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles & J. Poterba, eds, 'Dimensions of Tax Design: The Mirrlees Review', Oxford University Press, chapter 3, pp. 202–274.

Meyer, B. D. & Rosenbaum, D. T. (2001), 'Welfare, the Earned Income Tax Credit and the labor supply of single mothers', *Quarterly Journal of Economics* **116**(3), 1063–1114.

Miller, C., Katz, L. F., Azurdia, G., Isen, A., Schultz, C. & Aloisi, K. (2018), 'Boosting the Earned Income Tax Credit for Singles: Final Impact Findings from the Paycheck Plus Demonstration in New York City'. URL: *https://www.mdrc.org/publication/boosting-earned-income-tax-credit-singles*.

Mirrlees, J. A. (1971), 'An exploration in the theory of optimum income taxation', *Review of Economic Studies* **38**(114), 175–208.

Piketty, T. (1997), 'La redistribution fiscal face au chômage', *Revue française d'économie* **12**(1), 157–201.

Saez, E. (2001), 'Using elasticities to derive optimal income tax rates', *Review of Economic Studies* **68**, 205–229.

Saez, E. (2002), 'Optimal income transfer programs: Intensive versus extensive labor supply responses', *The Quarterly Journal of Economics* **117**(3), 1039–1073.

Saez, E., Slemrod, J. & Giertz, S. H. (2012), 'The elasticity of taxable income with respect to marginal tax rates: A critical review', *Journal of Economic Literature* **50**(1), 3–50.

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

Weinzierl, M. (2014), 'The promise of positive optimal taxation: normative diversity and a role for equal sacrifice', *Journal of Public Economics* **118**, 128–142.