Efficient Child Care Subsidies
Online Appendix

Christine Ho Nicola Pavoni

A Proofs

For ease of exposition, we report the statement in each Lemma and Proposition from the main text (*in italics*) before each proof. We start with a couple of preparatory results.

Claim 1 *In an optimal contract, we have $y(z_1) = 0$; and if $h(z_1) > 0$, then it solves*

$$1 - \frac{1}{\omega} v'(h(z_1)) \geq 0, \tag{A1}$$

with equality whenever $v'(1) \geq \omega$. If $v'(1) \leq \omega$, then $h(z_1) = 1$. In addition, if for some z we have $y(z) = 0$, then agent z gets the same allocation as type $z_1 = 0$.

Proof. Since $z_1 = 0$, we must have $y(z_1) = 0$. From the first order conditions of agent $z_1 = 0$ with respect to c and h, we have $v'(h(z_1)) \leq \omega$. Since $v'(0) = 0$ and $\omega > 0$, we have $h(z_1) > 0$. If $v'(1) \leq \omega$, utility can be increased strictly by increasing h whenever $h(z_1) < 1$; Hence, it must be that $h(z_1) = 1$. Consider now type $z > 0$ declaring $\sigma = z_1$. When $y(\sigma) = 0$, all agents have the same preferences over c and h and get the same utility when declaring $\sigma = z_1$. Thus, by DIC, agent z must receive at least the same utility as agent z_1. If $y(z) = 0$, UIC implies the reverse inequality, so that the utility between z and z_1 must be the same. Since agent z_1's problem is strictly concave, the allocation designed for z_1 minimizes the budget cost. Hence, we should use for all z with $y(z) = 0$ the allocation designed for z_1. ■

Claim 2 *Let λ be the multiplier associated to the budget constraint (1). We have $\lambda = \sum_{i=1}^{N} \pi(z_i)\phi(z_i) = \mathbf{E}[\phi]$.*

Proof. This result is shown by a simple variational exercise. Since we can increase $c(z_i)$ by the same amount for all i without violating the incentive constraints, it must be that $\sum_{i=1}^{N} \pi(z_i)\phi(z_i) \leq \lambda$. Since we can also decrease all $c(z_i)$ uniformly in an incentive compatible way, it must be that $\sum_{i=1}^{N} \pi(z_i)\phi(z_i) \geq \lambda$. Combining the two we get the desired equality. ■

A.1 Proof of Lemma 1

Under Assumptions 1, 2 and 3, any solution to the second best problem where only downward incentive constraints are imposed - that is, when the set of conditions (2) is relaxed to be $\sigma \leq z$ - delivers an optimal allocation. In addition, the 'local' downward incentive constraints can be imposed as equalities. Finally, if the upward incentive constraint is binding for two types $z_j < z_k$, then it is optimal for all agents with type $z_i : z_j \leq z_i \leq z_k$ to receive the same allocation (i.e, bunching).

Proof. We want to show that the solution from the relaxed second best problem, where the government maximizes the objective (3) subject to the budget constraint (1) and only the DIC in (2), is the solution to the original problem. In particular, we delete the UIC when the allocation is of employment whereas the unemployed allocation is one of pooling from Claim 1. The problem is a relaxed one, although the set of constraints that we neglect is endogenous to the chosen allocation. The relaxed second best problem is:

$$\max_{c(\cdot),y(\cdot),h(\cdot)} \quad \sum_{i=1}^{N} \pi(z_i)\,\phi(z_i)\left[c(z_i) - v\left(\frac{y(z_i)}{z_i} + h(z_i)\right)\right] \tag{R}$$

s.t.

$$\sum_{i=1}^{N} \pi(z_i)\,c(z_i) + \omega \leq \sum_i \pi(z_i)\left[y(z_i) + \omega h(z_i)\right] + M,$$

and for all i with $y(z_i) > 0$:

$$c(z_i) - v\left(\frac{y(z_i)}{z_i} + h(z_i)\right) \geq c(z_j) - v\left(\frac{y(z_j)}{z_i} + h(z_j)\right)\ \forall j < i; \qquad \text{DIC(i)}$$

Finally: $y(z_1) = 0$, and $\forall i$ such that $y(z_i) = 0$, we impose $c(z_i) - v(h(z_i)) = c(z_1) - v(h(z_1))$.

Step 1: We start with a lemma that shows that the double crossing condition described by Matthews and Moore (1987) holds for our framework.

Lemma 2 *Assume that Assumptions 1 and 2 hold. Let $z_- < z_0 < z_+$ be three ordered values of productivity. Let $\bar{w} := (\bar{c}, \bar{y}, \bar{h})$ and $\hat{w} := (\hat{c}, \hat{y}, \hat{h})$ be two allocations, and for all $z > 0$ define:*

$$u(w; z) := c - v\left(\frac{y}{z} + h\right).$$

Suppose that we have

$$u(\bar{w}; z_-) \geq u(\hat{w}; z_-),$$

$$u(\bar{w}; z_+) \geq u(\hat{w}; z_+),$$

but

$$u(\hat{w}; z_0) \geq u(\bar{w}; z_0),$$

with at least one inequality holding as strict. Then **(a)** *$\bar{h} > \hat{h}$, $0 < \bar{y} < \hat{y}$, and $\frac{\bar{y}}{z_*} + \bar{h} > \frac{\hat{y}}{z_*} + \hat{h}$, where $z_* \in (z_-, z_+)$ is the value for which the function $f(\cdot) := u(\hat{w}; \cdot) - u(\bar{w}; \cdot)$ takes the max with respect to z;* **(b)** *if $u(\bar{w}; z_+) > u(\hat{w}; z_+)$, then we have $u(\bar{w}; z_{++}) > u(\hat{w}; z_{++})$ for all $z_{++} > z_+$; and* **(c)** *if $u(\bar{w}; z_-) > u(\hat{w}; z_-)$, then we have $u(\bar{w}; z_{--}) > u(\hat{w}; z_{--})$ for all $z_{--} < z_-$.*

Double Crossing Property

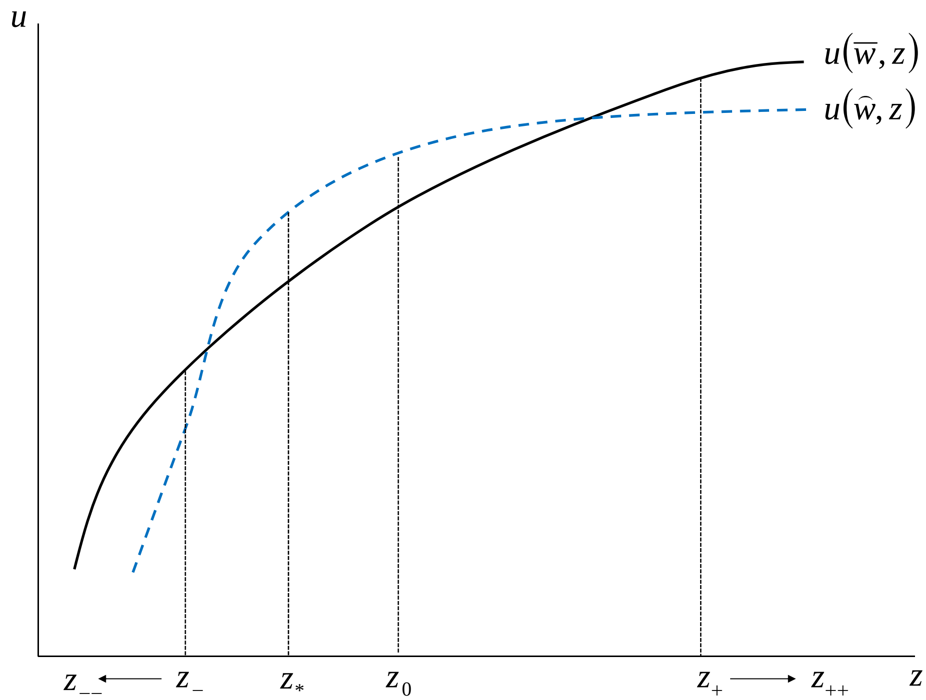

Figure A1: The above figure illustrates the double crossing property of indifference curves.

Proof. A graphical representation of the result is reported in Figure A1 for ease of exposition. Let $z_* \in (z_-, z_+)$ be the value for which the function $f(z) := u(\hat{w}; z) - u(\bar{w}; z)$ takes the max. The necessary first order condition (FOC) and second order condition (SOC) are respectively:

$$f'(z) = \frac{\hat{y}}{z_*^2} v'\left(\frac{\hat{y}}{z_*} + \hat{h}\right) - \frac{\bar{y}}{z_*^2} v'\left(\frac{\bar{y}}{z_*} + \bar{h}\right) = 0,$$

$$f''(z) = -\frac{\hat{y}^2}{z_*^4} v''\left(\frac{\hat{y}}{z_*} + \hat{h}\right) + \frac{\bar{y}^2}{z_*^4} v''\left(\frac{\bar{y}}{z_*} + \bar{h}\right) \leq 0 \iff \frac{\hat{y}}{z_*^4} \frac{v''\left(\frac{\hat{y}}{z_*} + \hat{h}\right)}{v'\left(\frac{\hat{y}}{z_*} + \hat{h}\right)} - \frac{\bar{y}}{z_*^4} \frac{v''\left(\frac{\bar{y}}{z_*} + \bar{h}\right)}{v'\left(\frac{\bar{y}}{z_*} + \bar{h}\right)} \geq 0,$$

where we used the FOC to simplify and rearrange the expression for the SOC. From the FOC, if either $\bar{y} = 0$ or $\hat{y} = 0$, then it must be that both $\bar{y}$ and $\hat{y}$ are equal to 0. But then, both $u(\hat{w}; z)$ and $u(\bar{w}; z)$ are independent on z and the only way to satisfy the three inequalities mentioned in the Lemma is to set $\hat{h} = \bar{h}$. So none of the three inequalities would be be strict and hence the assumptions of the lemma would not be satisfied. So we can safely presume that both $\bar{y}$ and $\hat{y}$ are positive so that the expression for the SOC is well defined.

(a) Suppose that $\bar{h} < \hat{h}$ and recall that v is strictly convex. Then from the FOC, it must be that $\bar{y} > \hat{y}$. Since by Assumption 1, the ratio $\frac{v''}{v'}$ decreases in the argument, the SOC can be satisfied only if $\frac{\bar{y}}{z_*} + \bar{h} > \frac{\hat{y}}{z_*} + \hat{h}$. But then, the FOC would not be satisfied. So it must be that $\bar{h} \geq \hat{h}$ and as a consequence of the FOC, $\bar{y} \leq \hat{y}$ and $\frac{\bar{y}}{z_*} + \bar{h} \geq \frac{\hat{y}}{z_*} + \hat{h}$. We can exclude the equality since we assumed that at least one inequality is strict, while if $\bar{h} = \hat{h}$ we must have $\bar{y} = \hat{y}$ and all types will be indifferent between the two allocations. We hence obtain the inequalities as stated in (a).

(b) Assume that $u(\bar{w}; z_+) > u(\hat{w}; z_+)$ and that for a $z_{++} > z_+ > z_*$ we have instead $u(\bar{w}; z_{++}) \leq$

$u(\hat{w}; z_{++})$. Then, there must be a minimum of f in the interval $(z_*, z_{++}]$. It is easy to see, by reverting the SOC inequality in the previous set of necessary conditions, that FOC and SOC would imply that $\hat{h} \geq \bar{h}$ and $\hat{y} \leq \bar{y}$, thereby contradicting the result in (a). Result **(c)** is shown symmetrically. ∎

Step 2: We can now start the core proof of Lemma 1. We use an induction argument. Denote the set of upward incentive constraints associated with mimicking agent i:

$$c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_j} + h\left(z_j\right)\right) \geq c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_j} + h\left(z_i\right)\right) \ \forall j < i. \qquad \text{UIC(i)}$$

First, note that UIC(1) is empty. Note also that since z_1=0, if $y(z_2) > 0$, then the UIC associated with z_1 mimicking z_2 would be satisfied as type $z_1 = 0$ agent would have an infinite cost of effort (since v is convex, we can bound the derivative downwards: $\lim_{e \to \infty} v(e) = \infty$). If $y(z_2) = 0$, then Claim 1 implies that z_2 must receive the same allocation as that of agent z_1 so that both DIC and UIC between z_1 and z_2 are satisfied. In addition, if $y(z_2) > 0$, then the DIC between z_2 and z_1 must also be binding. Otherwise, it would be possible to find a small enough $\epsilon > 0$ such that decreasing $c(z_i)$ for all $i > 1$ by ϵ and increasing $c(z_1)$ by $\frac{\sum_{i=2}^{N} \pi(z_i)}{\pi(z_1)} \epsilon$, would leave the budget constraint unchanged. This consumption perturbation, however, would be incentive compatible as long as the DIC between z_2 and z_1 is slack and ϵ is small enough. Note indeed that each agent $z_i > z_2$ receives the same utility by mimicking z_1. Recall that we impose DIC for all agents. A slack DIC between z_2 and z_1 implies that for all $z_i > z_2$ the DIC of agent i mimicking agent z_1 would also be slack. Finally, since consumption for all $z_i > z_1$ change by the same amount, incentive constraints are not affected among agents $i > 1$. This perturbation would weakly increases welfare since $\phi(z_1) \geq \mathbf{E}[\phi]$, by Assumption 3, hence generating a contradiction to optimality. In summary, we have shown that for $i = 1, 2$ the LDIC is binding and all UIC(i) constraints are satisfied for all $i \leq 2$. This is our starting point for the induction argument.

Now let $1 < k < N$ and assume that all UIC(i) constraints are satisfied for all $i \leq k$. We will show that in the relaxed problem (R) the LDIC must be satisfied with equality. In particular, we will show that, *if the local LDIC is slack and the induction hypothesis is true, then none of the non-local downward constraints can be binding*. Suppose that we have $z_j < z_k < z_{k+1}$ such that the LDIC between z_k and z_{k+1} is slack while the DIC between z_j and z_{k+1} is binding. Since we have

$$c\left(z_{k+1}\right) - v\left(\frac{y\left(z_{k+1}\right)}{z_{k+1}} + h\left(z_{k+1}\right)\right) > c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_{k+1}} + h\left(z_k\right)\right),$$

it must be that:

$$c\left(z_{k+1}\right) - v\left(\frac{y\left(z_{k+1}\right)}{z_{k+1}} + h\left(z_{k+1}\right)\right) = c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_{k+1}} + h\left(z_j\right)\right) > c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_{k+1}} + h\left(z_k\right)\right). \tag{A2}$$

On the other hand, from the DIC we have:

$$c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_k} + h\left(z_k\right)\right) \geq c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_k} + h\left(z_j\right)\right),$$

while from the UIC (which, by the inductive hypothesis, are assumed to be satisfied for $j \leq k$) we have:

$$c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_j} + h\left(z_j\right)\right) \geq c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_j} + h\left(z_k\right)\right).$$

We thus have the conditions to apply Lemma 2, where the three ranked types are $z_j < z_k < z_{k+1}$ and the supposedly optimal bundle for type z_k takes the role of bundle $(\hat{c}, \hat{y}, \hat{h})$ while the bundle for type z_j plays the role of the $(\bar{c}, \bar{y}, \bar{h})$ bundle in the Lemma. Lemma 2(b) implies that for all $z_{++} > z_{k+1}$,

$$c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_{++}} + h\left(z_j\right)\right) > c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_{++}} + h\left(z_k\right)\right).$$

But then from DIC, we have that all $z_{++} > z_{k+1}$ prefer their own bundle to that of agent z_j. Hence, we have:

$$c\left(z_{++}\right) - v\left(\frac{y\left(z_{++}\right)}{z_{++}} + h\left(z_{++}\right)\right) \geq c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_{++}} + h\left(z_j\right)\right) > c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_{++}} + h\left(z_k\right)\right),$$

which is impossible if $y_k > y_j$ as indicated in point (a) of the Lemma. So, no DIC is binding in terms of mimicking z_k. The first order conditions in the relaxed problem will therefore be those of full information. In particular,

$$z_k = v'\left(\frac{y\left(z_k\right)}{z_k} + h\left(z_k\right)\right).$$

At the same time, since z_{k+1} has a binding constraint with z_j, we have:

$$z_j \geq v'\left(\frac{y\left(z_j\right)}{z_j} + h\left(z_j\right)\right).$$

Since $z_k > z_j$, these conditions imply that:

$$v'\left(\frac{y\left(z_k\right)}{z_k} + h\left(z_k\right)\right) > v'\left(\frac{y\left(z_j\right)}{z_j} + h\left(z_j\right)\right). \tag{A3}$$

Now, consider the DIC between z_k and z_j:

$$c(z_k) - v\left(\frac{y(z_k)}{z_k} + h(z_k)\right) \geq c(z_j) - v\left(\frac{y(z_j)}{z_k} + h(z_j)\right).$$

Since v is convex and $z_k > z_j$, we also have $v'\left(\frac{y(z_k)}{z_k} + h\left(z_k\right)\right) > v'\left(\frac{y(z_j)}{z_k} + h\left(z_j\right)\right)$ from inequality (A3), which together with the DIC implies that $c(z_k) > c(z_j)$.

From (A2), since $c\left(z_j\right) - v\left(\frac{y(z_j)}{z_{k+1}} + h\left(z_j\right)\right) > c\left(z_k\right) - v\left(\frac{y(z_k)}{z_{k+1}} + h\left(z_k\right)\right)$, it must therefore be that:

$$v\left(\frac{y(z_k)}{z_{k+1}} + h(z_k)\right) > v\left(\frac{y(z_j)}{z_{k+1}} + h(z_j)\right) \quad \Longleftrightarrow \quad \frac{y(z_k)}{z_{k+1}} + h(z_k) > \frac{y(z_j)}{z_{k+1}} + h(z_j). \tag{A4}$$

Since Lemma 2(a) implies that $y(z_k) > y(z_j)$ and $h(z_k) < h(z_j)$, inequality (A4) implies that $\frac{y(z_k)}{z} + h(z_k) > \frac{y(z_j)}{z} + h(z_j)$ for all $z \leq z_{k+1}$. On the other hand, Lemma 2(a) also implies that $\frac{y(z_k)}{z_*} + h(z_k) < \frac{y(z_j)}{z_*} + h(z_j)$ for some $z_j < z_* < z_{k+1}$. This is hence a contradiction. It must therefore be that if LDIC for z_{k+1} is slack, then all DIC for z_{k+1} must also be slack.

But then, if the DIC for agent z_{k+1} mimicking a lower type are slack, we can find an incentive compatible $\epsilon > 0$, such that we can increase $c(z_i)$ for all $i \neq k+1$ by ϵ and decrease $c(z_{k+1})$ by $\frac{\sum_{i \neq k+1} \pi(z_i)}{\pi(z_{k+1})}\epsilon$, such that the budget constraint remains the same. This will be incentive compatible since the LDIC between z_{k+1} and z_k is slack by assumption and consumption for all $i \neq k+1$ increase by the same amount. Welfare changes by the amount

$$\left[\sum_{i \neq k+1} \pi(z_i)\phi(z_i) - \sum_{i \neq k+1} \pi(z_i)\phi(z_{k+1})\right]\epsilon \geq 0,$$

where the inequality is implied by Assumption 3. Thus, it must be that the LDIC are binding.

We now show that binding LDIC implies that the UIC are satisfied. Note that the binding LDIC for z_{k+1} mimicking z_k implies:

$$y(z_{k+1}) + \omega h(z_{k+1}) - c(z_{k+1}) \geq y(z_k) + \omega h(z_k) - c(z_k).$$

The statement must be true, otherwise, the budget constraint could be relaxed (strictly) by replacing allocation $(y(z_{k+1}), h(z_{k+1}), c(z_{k+1}))$ with allocation $(y(z_k), h(z_k), c(z_k))$. Namely, we can give to agent z_{k+1} the contract designed for agent z_k. By eliminating one contract, all incentive constraints will remain satisfied and agent z_{k+1}'s utility will be the same as the agent is indifferent between the two allocations. By the induction assumption, since all LDIC are satisfied with equality, we have that $y(z_i) + \omega h(z_i) - c(z_i)$ *weakly increases with* z_i for $i = 1, \ldots, k+1$.

Now, assume that some of the UIC(k+1) are not satisfied. Namely, for some $1 < j \leq k$ (recall that the UIC(1) is empty), we have:

$$c(z_{k+1}) - v\left(\frac{y(z_{k+1})}{z_j} + h(z_{k+1})\right) > c(z_j) - v\left(\frac{y(z_j)}{z_j} + h(z_j)\right).$$

Then, it must be that for such z_j we have:

$$y(z_{k+1}) + \omega h(z_{k+1}) - c(z_{k+1}) < y(z_j) + \omega h(z_j) - c(z_j).$$

Otherwise, we could replace allocation $(y(z_j), h(z_j), c(z_j))$ with $(y(z_{k+1}), h(z_{k+1}), c(z_{k+1}))$. The budget constraint would be weakly relaxed, the incentive constraints remain satisfied, and agent z_j would have higher utility. Hence welfare would increase (strictly). But this provides a contradiction to the monotonicity obtained from the binding LDIC. Hence, it must be that all UIC(k+1) are satisfied.

Finally, *we show that bunching may occur and characterize when this happens.* Claim 1 shows that if $y(z_i) = y(z_j) = 0$, then UIC are trivially satisfied and bunching arises. So, assume that $z_i > z_j$ and $y(z_i) > 0$. If the UIC between z_j vs z_i is binding, it must be that $y(z_i) + \omega h(z_i) - c(z_i) \leq y(z_j) + \omega h(z_j) - c(z_j)$. Otherwise, we can eliminate the allocation for agent z_j and give agent z_j the allocation now in place for agent

z_i. This would keep welfare the same as UIC is binding and also relax the budget constraint. But since the DIC are binding for all $k \leq i$ and $i > j$, the argument we made above implies that the reverse inequality must also be true. Thus, it much be that $y(z_i) + \omega h(z_i) - c(z_i) \leq y(z_j) + \omega h(z_j) - c(z_j)$. Let's now look at type z_{i-1}. The previous argument implies that $y(z_i) + \omega h(z_i) - c(z_i) = y(z_{i-1}) + \omega h(z_{i-1}) - c(z_{i-1}) = y(z_j) + \omega h(z_j) - c(z_j)$. Recall that the binding LDIC between z_i and z_{i-1} implies that agent z_i is indifferent between the allocation designed for him and the allocation designed for z_{i-1}. So, we can eliminate the allocation designed for him and use the allocation designed for z_{i-1} instead. As usual, this will keep the budget and welfare the same, and possibly relax the incentive constraints. This same argument can be done till agent z_j. Hence a bunching allocation among these agents would be optimal when UIC is binding. ■

A.2 Proof of Proposition 1

Under Assumptions 1, 2 and 3, we have:

- **(a)** *The 'net surplus' $y^*(z) + \omega h^*(z) - c^*(z)$ is non-decreasing in z;*
- **(b)** *Utility of agents in equilibrium $V^*(z|z)$ is non-decreasing in z, and strictly increasing between any two levels $z_{i+1} > z_i$ when $y^*(z_i) > 0$.*
- **(c)** *For all z, $h^*(z) \leq 1$.*

Proof. We omit the superscript * on the optimal allocation for notational simplicity.

(a) The monotonicity property of the net surplus has been shown in the proof of Lemma 1, using the fact that LDIC are satisfied with equality.

(b) We know from Claim 1 that if $y(z_i) = y(z_{i-1}) = 0$, then we have pooling so that z_i and z_{i-1} will get the same utility. Now, suppose that $y(z_i) > y(z_{i-1}) = 0$. Recall that $z_1 = 0$ and $y(z_1) = 0$. Since LDIC implies

$$c\left(z_1\right) - v\left(h\left(z_1\right)\right) \leq c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_i} + h\left(z_i\right)\right),$$

utility must be weakly increasing. Now, assume that the lower type has $y(z_{i-1}) > 0$. Then DIC implies that

$$c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_i} + h\left(z_i\right)\right) \geq c\left(z_{i-1}\right) - v\left(\frac{y\left(z_{i-1}\right)}{z_i} + h\left(z_{i-1}\right)\right) > c\left(z_{i-1}\right) - v\left(\frac{y\left(z_{i-1}\right)}{z_{i-1}} + h\left(z_{i-1}\right)\right),$$

where the first inequality uses the incentive compatibility constraint while the second inequality uses the fact that $z_i > z_{i-1}$ and $y(z_{i-1}) > 0$.

(c) Suppose that we have $h(z_i) > 1$ for some i. Since the marginal return to providing household child care beyond child care needs is zero while the marginal cost is positive, we can reduce both $h(z_i)$ and $c(z_i)$ so that the utility of agent z_i is unchanged (if we denote $\hat{c}$ and $\hat{h}$ as the new values, we have: $c(z_i) - \hat{c} = v\left(\frac{y(z_i)}{z_i} + h(z_i)\right) - v\left(\frac{y(z_i)}{z_i} + \hat{h}\right)$), and relax the budget constraint (which can subsequently be translated to an increase in welfare by a uniform increase in consumption). The incentives of agent z_i to mimic lower type agents are unchanged since utility of agent z_i is unchanged. We also need to show that such a change weakly relaxes the DIC of higher types. When $y(z_i) = 0$, incentives are unchanged since higher type agents get the same utility as agent z_i when mimicking type z_i. When $y(z_i) > 0$, the convexity of v implies that higher types will now get a lower utility when pretending to be type z_i. This is so since for $z > z_i$, we have $v\left(\frac{y(z_i)}{z} + h(z_i)\right) - v\left(\frac{y(z_i)}{z} + \hat{h}\right) < v\left(\frac{y(z_i)}{z_i} + h(z_i)\right) - v\left(\frac{y(z_i)}{z_i} + \hat{h}\right)$ by the convexity of v. ■

A.3 Proof of Proposition 2

Under Assumptions 1, 2, and 3, we have:

(a) *Unemployment: Recall that* $z_1 = 0$*. We have* $y^*(z_1) = 0$ *and* $h^*(z_1) > 0$*, where*

$$1 - \frac{1}{\omega} v'\left(h^*(z_1)\right) \geq 0, \tag{A5}$$

with equality whenever $v'(1) \geq \omega$*. If* $v'(1) \leq \omega$*, then* $h^*(z_1) = 1$*. In addition, for all* z *such that* $y^*(z) = 0$*, type* z *gets the same allocation as type* z_1*.*

(b) *Low productivity: Let* $z \leq \omega$*. We have* $h^*(z) > 0$*, and if* $y^*(z) > 0$*, then* $h^*(z) = 1$*.*

(c) *Segmentation: If* $y^*(z) > 0$*, then* $y^*(z') > 0 \; \forall \; z' > z$*; hence if* $y^*(z) = 0$*, then* $y^*(z') = 0 \; \forall \; z' < z$*;*

(d) *Monotonicity: Let* $z' > z$ *for which we have no bunching. If* $h^*(z') \leq h^*(z)$*, then* $y^*(z') > y^*(z)$*; and if* $y^*(z') \leq y^*(z)$*, then* $h^*(z') > h^*(z)$*.*

(e) *Wedges for the employed: Let* z_i *be such that* $y^*(z_i) > 0$*. Then labor wedges are non-negative:*

$$1 - \frac{1}{z_i} v'\left(e^*(z_i)\right) \geq 0; \tag{A6}$$

If, in addition, $h^*(z_i) > 0$*, then the child care wedges are also non-negative:*

$$1 - \frac{1}{\omega} v'\left(e^*(z_i)\right) \geq 0. \tag{A7}$$

Both wedges are strictly positive whenever $\phi(z_{i+1}) < \mathbf{E}[\phi]$*.*
For $i = N$*, the labor wedge is zero and* $h^*(z_N) = 0$*.*

Proof. We omit the superscript * on the optimal allocation for notational simplicity.
(a) This result has been shown in Claim 1.
(b) If $y(z) = 0$, then agent z gets the unemployed allocation and provides $h(z) = h(0) > 0$, as shown in Claim 1. Now, suppose that for $z \leq \omega$, $y(z) > 0$ but $h(z) < 1$. We show that we can reduce y and increase h so as to keep agent z's utility constant without violating any DIC nor the budget constraint. The budget constraint can only improve since keeping agent's $z \leq \omega$ utility constant implies a change $\Delta y + z\Delta h = 0 \leq \Delta y + \omega \Delta h$, where the last quantity is the change in the budget constraint. The DIC are not affected since all $z' > z$ will now find the new allocation less attractive than before (note that $\Delta h(z) = \frac{\Delta y(z)}{z} \geq \frac{\Delta y(z)}{z'}$ so that type z' mimicking z will now face higher effort cost). This change would strictly improve welfare since we can uniformly redistribute the increase in the budget, $\Delta y + \omega \Delta h$, among all types without altering incentives.
(c) We want to show that if for z_i we have $y(z_i) > 0$, then it must be that $y(z_n) > 0$ for all $n > i$ (we know that $y(z_1) = 0$, so $i > 1$). Suppose that for $k > i > j$ we have both $y(z_j) = y(z_k) = 0$ and $y(z_i) > 0$. From Proposition 1(b), we know that utility between $i+1$ and i must be strictly increasing since $y(z_i) > 0$. We also know that utility is weakly increasing in type from the DIC. So, the utility of agent $z_k \geq z_{i+1}$ must be strictly larger than the utility of agent $z_j \leq z_i$. But we know that we have pooling among the unemployed, and they all receive the same utility, so that z_k and z_j must have the same utility if $y(z_j) = y(z_k) = 0$. Hence, we get a contradiction.
(d) Since the case where $y(z_{i-1}) = 0$ is trivial to show, we show it for the case where $y(z_{i-1}) > 0$ (hence $i > 1$). From (c) above, we have $y(z_i) > 0$ as well. Using the DIC and UIC between z_i and z_{i-1} (recall that

the local DIC binds):

$$
\begin{aligned}
c\left(z_i\right)-v\left(\tfrac{y(z_i)}{z_i}+h\left(z_i\right)\right) &= c\left(z_{i-1}\right)-v\left(\tfrac{y(z_{i-1})}{z_i}+h\left(z_{i-1}\right)\right)\\
c\left(z_{i-1}\right)-v\left(\tfrac{y(z_{i-1})}{z_{i-1}}+h\left(z_{i-1}\right)\right) &\geq c\left(z_i\right)-v\left(\tfrac{y(z_i)}{z_{i-1}}+h\left(z_i\right)\right).
\end{aligned}
$$

Adding the two inequalities together and rearranging, we get:

$$
v\left(\frac{y\left(z_i\right)}{z_{i-1}}+h\left(z_i\right)\right)-v\left(\frac{y\left(z_i\right)}{z_i}+h\left(z_i\right)\right)\geq v\left(\frac{y\left(z_{i-1}\right)}{z_{i-1}}+h\left(z_{i-1}\right)\right)-v\left(\frac{y\left(z_{i-1}\right)}{z_i}+h\left(z_{i-1}\right)\right).
$$

By the first fundamental theorem of calculus, this implies that:

$$
\int_{z_{i-1}}^{z_i} v'\left(\frac{y\left(z_i\right)}{s}+h\left(z_i\right)\right)\frac{y\left(z_i\right)}{s^2}ds\geq\int_{z_{i-1}}^{z_i} v'\left(\frac{y\left(z_{i-1}\right)}{s}+h\left(z_{i-1}\right)\right)\frac{y\left(z_{i-1}\right)}{s^2}ds.
$$

If $h\left(z_i\right)\leq h\left(z_{i-1}\right)$, then convexity of $v\left(.\right)$ and $0<y\left(z_i\right)<y\left(z_{i-1}\right)$ would imply that the integrand on the left hand side is everywhere smaller than the right hand side, a contradiction. Thus, it must be that $y\left(z_i\right)\geq y\left(z_{i-1}\right)$. A similar argument implies that if $y\left(z_i\right)\leq y\left(z_{i-1}\right)$ then $h\left(z_i\right)\geq h\left(z_{i-1}\right)$. The equivalent statement, in both cases, says that a strict inequality in the first statement implies a strict inequality in the second. Clearly, if there is no bunching, a slightly stronger result holds: If $h\left(z_i\right)\leq h\left(z_{i-1}\right)$, then $y\left(z_i\right)>y\left(z_{i-1}\right)$ and: If $y\left(z_i\right)\leq y\left(z_{i-1}\right)$, then $h\left(z_i\right)>h\left(z_{i-1}\right)$. Since the index i was generic, we have shown the monotonicity result reported in the text. **(e)** Recall the relaxed problem (R). From Lemma 1, we can charaterize the problem by focusing on problem (R). For z_i such that $y(z_i)>0$, the first order conditions with respect to consumption, earnings and household child care, are given by:

$$
\begin{aligned}
c\left(z_i\right) &: \left[\pi\left(z_i\right)\phi(z_i)+\delta\left(z_i\right)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right) = \left[\textstyle\sum_{k>i}\mu_i(z_k)+\lambda\pi\left(z_i\right)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right),\\
y\left(z_i\right) &: \left[\pi\left(z_i\right)\phi(z_i)+\delta(z_i)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right) = \textstyle\sum_{k>i}\mu_i(z_k)\tfrac{z_i}{z_k}v'\left(\tfrac{y(z_i)}{z_k}+h(z_i)\right)+\lambda\pi\left(z_i\right)z_i,\\
h\left(z_i\right) &: \left[\pi\left(z_i\right)\phi(z_i)+\delta(z_i)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right) \geq \textstyle\sum_{k>i}\mu_i(z_k)v'\left(\tfrac{y(z_i)}{z_k}+h(z_i)\right)+\lambda\pi\left(z_i\right)\omega,
\end{aligned}
$$

where $\mu_j(z_i)\geq 0$ is the Kuhn-Tucker multiplier associated with the DIC guaranteeing that agent z_i does not mimic type $z_j<z_i$. For all i, we defined $\delta\left(z_i\right):=\sum_{j<i}\mu_j(z_i)$ and multiplied the first condition by $v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)$ and the second one by $z_i>0$. Clearly, the first order conditions for y and h are satisfied with equality when we have an interior solution for them.

Substituting the first order condition with respect to $c(z_i)$ into those with respect to $y(z_i)$ and $h(z_i)$, and rearranging, we get:

$$
\lambda\pi(z_i)z_i=\lambda\pi(z_i)v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)+\sum_{k>i}\mu_i(z_k)\left[v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)-\frac{z_i}{z_k}v'\left(\frac{y(z_i)}{z_k}+h(z_i)\right)\right],
$$

and

$$
\lambda\pi(z_i)w\leq\lambda\pi(z_i)v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)+\sum_{k>i}\mu_i(z_k)\left[v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)-v'\left(\frac{y(z_i)}{z_k}+h(z_i)\right)\right],
$$

with equality if $h(z_i)>0$. The term in square brackets are strictly positive since $z_k>z_i$ and v is convex. It is now easy to see how we can get the wedges.

We now want to show that when $\phi(z_{i+1}) < \lambda = \mathbf{E}[\phi]$ *(where the last equality is from Claim 2), then some DIC must be binding with* $\mu_i(z_k) > 0$ *for some* $k > i$.

Suppose that all $\mu_i(z_k)$ are nil. Then, from the first order conditions with respect to $y(z_i)$ and $h(z_i)$, we have $\lambda\pi(z_i)z_i \geq \lambda\pi(z_i)\omega$, which implies $z_i \geq \omega$. Excluding the knife edge case of $z_i = \omega$,[1] the first order conditions imply $h(z_i) = 0$ and:

$$z_i = v'\left(\frac{y(z_i)}{z_i}\right).$$

In addition, for all $z_j < z_i$ for which $y(z_j) > 0$, we have:

$$z_j \geq v'\left(\frac{y(z_j)}{z_j} + h(z_j)\right);$$

and if $y(z_j) = 0$, we have $h(z_j) > 0$ and:

$$\omega \geq v'\left(h(z_j)\right).$$

Since both $z_i > z_j$ and $z_i \geq \omega$, the two inequalities imply:

$$v'\left(\frac{y(z_i)}{z_i}\right) > v'\left(\frac{y(z_j)}{z_j} + h(z_j)\right) \quad \text{or} \quad v'\left(\frac{y(z_i)}{z_i}\right) \geq v'\left(h(z_j)\right).$$

Convexity of v and $h(z_j) \geq 0$ imply $y(z_j) < y(z_i)$ as a consequence of the first inequality, and $y(z_j) = 0 < y(z_i)$ as a consequence of the second. Allowing for zeros, we can summarize the above conditions by:

$$\frac{y(z_i)}{z_i} > \frac{y(z_j)}{z_j} + h(z_j) \quad \text{and} \quad y(z_j) < y(z_i),\ h(z_j) \geq h(z_i) = 0.$$

In addition, since $z_{i+1} > z_i > z_j$ we have $\frac{y(z_i)}{z_i} > \frac{y(z_j)}{z_{i+1}} + h(z_j)$. From Proposition 1(ii) since utility is increasing, we have:

$$c_i - v\left(\frac{y(z_i)}{z_i}\right) > c_j - v\left(\frac{y(z_j)}{z_j} + h(z_j)\right).$$

This, together with $\frac{y(z_i)}{z_i} > \frac{y(z_j)}{z_{i+1}} + h(z_j)$ implies $c(z_i) > c(z_j)$.

Now, if we look at the first order condition for $c(z_{i+1})$, since by assumption, $\phi(z_{i+1}) < \lambda$, we must have $\delta(z_{i+1}) > 0$. As a consequence, by the definition of $\delta(z_{i+1})$, some DIC is binding for agent $i+1$, with positive multiplier: $\mu_j(z_{i+1}) > 0$ for some $j < i$ (by assumption, $\mu_i(z_{i+1}) = 0$). Consider such a j. Since the

[1] If $z_i = \omega$, point (b) above implies that in this case, $y(z_i) > 0$ implies $h(z_i) = 1$. It would also imply either $y(z_j) = 0$ or $h(z_j) = 1$ for all $z_j < z_i$. Hence $y(z_i) > y(z_j)$ and the proof would follow the same line as we do here assuming that $z_i > \omega$.

LDIC for z_{i+1} is binding (and by assumption for such j the non-local DIC z_{i+1} vs z_j binds), we have:

$$\begin{aligned} c_{i+1} - v\left(\frac{y(z_{i+1})}{z_{i+1}}\right) &= c_i - v\left(\frac{y(z_i)}{z_{i+1}}\right), \\ &= c_j - v\left(\frac{y(z_j)}{z_{i+1}} + h(z_j)\right). \end{aligned}$$

Since we saw above that $c(z_i) > c(z_j)$, it must be that:

$$v\left(\frac{y(z_i)}{z_{i+1}}\right) > v\left(\frac{y(z_j)}{z_{i+1}} + h(z_j)\right) \quad \Longleftrightarrow \quad \frac{y(z_i)}{z_{i+1}} > \frac{y(z_j)}{z_{i+1}} + h(z_j). \tag{A8}$$

Now, define:

$$\Delta(z) := c(z_i) - v\left(\frac{y(z_i)}{z}\right) - \left[c(z_j) - v\left(\frac{y(z_j)}{z} + h(z_j)\right)\right].$$

Using again the fact that both the LDIC for z_{i+1} and the non-local DIC z_{i+1} vs z_j bind, we have that $\Delta(z_{i+1}) = 0$. Moreover, DIC (recall $j < i$) implies $\Delta(z_i) \geq 0$. At the same time, deriving with respect to z, we have:

$$\Delta'(z) = \frac{y(z_i)}{z^2} v'\left(\frac{y(z_i)}{z}\right) - \frac{y(z_j)}{z^2} v'\left(\frac{y(z_j)}{z} + h(z_j)\right).$$

Since $y(z_i) > y(z_j)$, *if we show that for all* $z \in [z_i, z_{i+1}]$ *we have* $v'\left(\frac{y(z_i)}{z}\right) > v'\left(\frac{y(z_j)}{z} + h(z_j)\right)$, *then we would be done*. This is so since the above inequality implies that $\Delta'(z) > 0$ for $z \in [z_i, z_{i+1}]$, contradicting the fact that $\Delta(z_i) \geq \Delta(z_{i+1}) = 0$. This would hence mean that the initial assumption was false, namely we must have some $\mu_k(z_i) > 0$. Recall that from (A8) we have $\frac{y(z_i)}{z_{i+1}} > \frac{y(z_j)}{z_{i+1}} + h(z_j)$. Since $y(z_i) > y(z_j)$, then $\frac{y(z_i)}{z} > \frac{y(z_j)}{z} + h(z_j)\ \forall z \leq z_{i+1}$ as desired.

We have hence shown that for z_i with $y(z_i) > 0$ and $\phi(z_{i+1}) < \lambda = \mathbf{E}[\phi]$ (thus, $z_i < z_N$):

$$z_i > v'(e(z_i)).$$

Now, suppose that for z_i we have both $y(z_i) > 0$ and $h(z_i) > 0$. If $z_i > \omega$, then the first order conditions exclude the possibility that all multipliers $\mu_i(z_k)$ are zero. On the other hand, we saw above that if all multipliers $\mu_i(z_k)$ are zero and $y(z_i) > 0$, then it must be that $z_i \geq \omega$. If we exclude the case $z_i = \omega$, we hence have that:[2]

$$\omega > v'(e(z_i)).$$

The standard no distortion at the top result is easily obtained from the first order conditions as no DIC exists for this agent. Since $z_N > \omega$, we must have $h(z_N) = 0$ and:

$$1 - \frac{1}{z_N} v'\left(\frac{y(z_N)}{z_N}\right) = 0 \Rightarrow 1 - \frac{1}{\omega} v'\left(\frac{y(z_N)}{z_N}\right) < 0.$$

$\blacksquare$

[2]The case were all multipliers are zero and $z_i = \omega$ also has $h(z_i) = 1$ from point (b) above. If $v'(1) > \omega$, then it cannot be. This case hence can only happen when $v'(1) \leq \omega$. For this case, we can follow the same line of proof as before to derive the wedge for $y(z_i)$ to obtain a contradiction (recall that if $z_j < z_i$, then $z_j < \omega$).

A.4 Proof of Proposition 3

Let $f^*(\sigma)$ be the optimal formal child care cost associated with the constrained optimal $h^*(\sigma)$. The following subsidy rates and transfers implement the constrained optimum.
(a) For employed agents, we have:

$$\text{If} \quad \sigma \notin Z_0^*, \quad \text{then} \quad s(\sigma, f) = \begin{cases} \left(1 - \frac{1}{\omega} v'\left(\frac{y^*(\sigma)}{z_N} + h^*(\sigma)\right)\right)^+ & if \quad f \leq f^*(\sigma); \\ \left(1 - \frac{1}{\omega} v'\left(\frac{y^*(\sigma)}{\bar{z}_0} + h^*(\sigma)\right)\right)^- & if \quad f > f^*(\sigma). \end{cases}$$

(b) For unemployed agents, the subsidy rate is zero: If $\sigma \in Z_0^*$, then $s(\sigma, f) = 0 \; \forall f$.
(c) For all $\sigma \in Z$, the optimal transfer scheme is set as follows:

$$T(\sigma) = y^*(\sigma) - c^*(\sigma) - f^*(\sigma) + s(\sigma, f^*(\sigma)) f^*(\sigma);$$

where $c^*(\cdot)$ and $y^*(\cdot)$ are the consumption and income functions of the second best allocation.

Proof. Consider the following maximisation problem for all $z, \sigma \in Z^2$:

$$\hat{V}(\sigma|z) := \max_f \; y^*(\sigma) - T(\sigma) - (1 - s(\sigma, f)) f - v\left(\frac{y^*(\sigma)}{z} - \frac{\omega - f}{\omega}\right).$$

From the arguments we made in the main text, the piecewise linear function $s(\sigma, f)$ implies that the solution to the above maximization problem is $f^*(\sigma)$ for each $z \notin Z_0$ and for $z = \bar{z}_0$. From the expression of T, when $f = f^*(\sigma)$, we have:

$$y^*(\sigma) - T(\sigma) - (1 - s(\sigma, f^*(\sigma))) f^*(\sigma) = c^*(\sigma).$$

Since declaring σ forces the agent to choose $y^*(\sigma)$, for all such z we have $\hat{V}(\sigma|z) = V^*(\sigma|z)$, the second best value for each declaration σ. That is, each type z, such that either $z \notin Z_0$ or $z = \bar{z}_0$ declaring σ and possibly deviating in f, gets at most $V^*(\sigma|z)$. Since the second best allocation is incentive compatible, we have shown that the proposed scheme is robust to joint deviations in σ and f for all such types. Consider now unemployed agents: $z \in Z_0$ and $z < \bar{z}_0$. Since all unemployed receive the same utility in equilibrium, we have for all these agents $V^*(z|z) = V^*(\bar{z}_0|\bar{z}_0)$. On the other hand, it is immediate to see that for all $z < \bar{z}_0$, $\hat{V}(\sigma|z) \leq \hat{V}(\sigma|\bar{z}_0)$. The fact that agent $\bar{z}_0$ does not want to deviate hence implies that none of these agents want to deviate either. In summary, we have shown that each type $z \in Z$ chooses to tell the truth, produces $y^*(z)$ and spends $f^*(z)$ in formal child care. Since transfers $T(\cdot)$ are adjusted by the child care subsidy to satisfy the government budget constraint, the proof is complete. ∎

A.5 Proof of Proposition 4

Under Assumption 4, there is a $\overline{T} \in \mathbb{R}$ such that the following subsidy rates and transfers implement the constrained optimum.

(a) For employed agents (who earn $y > 0$), we have:

$$s(y,f) = \begin{cases} \left(1 - \frac{1}{\omega} v'\left(\frac{y}{z_N} + 1 - \frac{f(y)}{\omega}\right)\right)^{+} & if \quad f \leq f(y); \\ \left(1 - \frac{1}{\omega} v'\left(\frac{y}{\bar{\bar{z}}_0} + 1 - \frac{f(y)}{\omega}\right)\right)^{-} & if \quad f > f(y); \end{cases}$$

$$\textit{if } y \in \mathcal{Y} \textit{ then } T(y) = y - c(y) - f(y) + s(y, f(y)) f(y); \textit{ otherwise } T(y) = \overline{T}.$$

(b) For unemployed agents (with $y = 0$), the second best allocation is implemented by having:

$$s(0,f) \equiv 0, \qquad \textit{and} \qquad T(0) = -c(0) - f(0).$$

Proof. Let $\overline{y}$ be such that $v'\left(\frac{\overline{y}}{z_N}\right) = z_N$ and $\overline{T} := \overline{y} + \max_{(\sigma,z)\in Z^2} V^*(\sigma|z)$. Agent z solves:

$$\max_{y\geq 0, f\leq\omega} \quad y - T(y) - (1 - s(y,f))f^{+} - v\left(\frac{y}{z} - \frac{\omega - f}{\omega}\right).$$

Under Assumption 4, for each $y \in \mathbb{R}_+$, there is only one value for $f(y)$ and hence, a well defined subsidy rate schedule $s(y,f)$. Since for each σ there is only one value of $y \in \mathcal{Y}$, to each σ in the direct mechanism there is only one pair of values y and f. Moreover, it is immediate to see that the punishment induced by $\overline{T}$ implies that no agent will ever choose $y \notin \mathcal{Y}$. For $y \in \mathcal{Y}$, however, the agent has weakly less joint deviations available compared to those considered in the implementation of Proposition 3. Moreover, for $y \in \mathcal{Y}$, the welfare and net revenues for the government are as in the direct mechanism. The result hence follows from Proposition 3. ■

B Numerical Appendix

This section describes the calibration process and presents further results and sensitivity analysis. We first present and solve the agent's private problem given the actual tax and benefit system, impute wages, and describe the existing US tax and benefit scheme. Those ingredients contribute to the calibration exercise. We then present sensitivity analyses based on specifications that vary the baseline parameters and the social welfare criteria, and that allow for income effects.

B.1 Agent's Private Problem

The private problem of an agent is given by

$$\max_{\{c,l,h\}} \quad \frac{c^{1-\alpha} - \kappa}{1-\alpha} - \frac{1}{\theta}\frac{e^{1+\gamma}}{1+\gamma} \tag{B1}$$

s.t.

$$c = y - T^E(y,f) - f,$$

where effort $e = l + h$, earnings $y = zl$, and formal child care cost $f = \omega(1-h)$. $T^E(y,f)$ are net taxes faced by the agent under the existing US tax and benefit system, E.

Let $l(z; T, \omega)$ be the labor supply choice for agent z under the transfer program T and child care cost ω. In our framework, the *reservation wage* $R(T, \omega)$ can be identified with the hypothetical type for which $l(z; T, \omega) > 0$ if $z > R(T, \omega)$ and $l(z; T, \omega) = 0$ if $z < R(T, \omega)$. Obviously, we do not observe the distribution of types for wages below $R(T, \omega)$. In order to identify the distribution of z for $z < R(T, \omega)$, in the next section, we assume that (the types of) mothers with kids *above* six years of age are distributed as our group of reference. When kids are grown up, mothers plausibly face a lower amount of child care needs. To ease the exposition, suppose that mothers with grown up kids face no child care needs at all. This would correspond to a $\omega' = 0$. It is immediate to see that, since in the US tax scheme $T_y^E(0, 0) < 1$ (i.e, income tax is less than 100% at zero income) and when there are no child care needs $f(z; T, 0) = 0$ for all z, the reservation wage for these agents is zero, i.e., $R(T, 0) = 0$. In general, we will assume that a reduction in child care needs reduces the reservation wage.

B.2 Wage Imputation

Wages for working mothers in the CPS are computed as yearly gross earnings divided by total hours of work in one year.[3] On the other hand, non working mothers have no earnings. In our model, a mother may not be working either because (i) she has no employment opportunities ($z_1 = 0$) or (ii) her wage is below her reservation wage. As described in the text, we consider the involuntarily unemployed as those with no employment opportunities. We now focus our analysis on the remaining mothers, that is, those who are either working or voluntarily unemployed.

Our log wage function is given by:

$$\ln wage_i = X_i \beta + \epsilon_i,$$

where X_i is a vector of demographic characteristics such as age, education, health, ethnicity and number of children, that are correlated with wages, and ϵ_i is an unobserved component. β is a vector of coefficients that we aim to estimate so that we can impute wages for the unemployed. Note, however, that wages are observed only for the employed. If there is a correlation between wages and the decision to work, the distribution of ϵ_i will be truncated. We therefore would not be able to rely on Ordinary Least Squares (OLS) regressions to estimate β and would need to account for the selection of agents into work.

From the agent's private problem in Section B.1, she works if $z > R(T, \omega)$, where R is the reservation wage. As we just saw, mothers with child care needs have a positive reservation wage. We therefore model the work decision of the agent as:

$$work_i = 1 \left[\delta lnwage_i - \gamma K_i - \eta_i > 0 \right],$$

where K_i is a dummy variable that captures the child care needs of agents (presence of children aged below 6) and therefore $K_i = 1$ reflects a positive reservation wage ($R > 0$) of mothers. The random variable η_i is an unobserved determinant of work participation that may be present in the real world, and δ and γ are coefficients to be estimated.

[3] We drop mothers with wages above \$40 which consist of 39 observations (approximately 1% of the sample). Those mothers are very sparsely distributed between a wage range of \$40 to \$276.

Using the equation for log wage, we can rewrite the work decision as:

$$work_i = 1\left[X_i\psi - \gamma K_i - u_i > 0\right],$$

where $u_i = \delta\epsilon_i - \eta_i$. The unobserved terms are assumed to follow a bivariate normal distribution:

$$\begin{bmatrix} \epsilon_i \\ u_i \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_\varepsilon & \rho \\ \rho & 1 \end{bmatrix}\right),$$

where σ_ε is the variance of ϵ_i and ρ is the correlation between ϵ_i and u_i.

Thus, the conditional mean of log wages is given by:

$$E\left[\ln wage_i | work_i = 1\right] = X_i\beta + \rho\sigma_\epsilon\lambda_i\left(X_i\psi - \gamma K_i\right),$$

where $\lambda_i\left(X_i\psi - \gamma K_i\right) = \frac{\phi(X_i\psi - \gamma K_i)}{\Phi(X_i\psi - \gamma K_i)}$ and ϕ and Φ are the normal pdf and cdf respectively. λ_i is the inverse mills ratio that takes into account the fact that the distribution of ϵ_i is truncated. Note that even in the absence of the unobserved determinant of work, η_i, we would still have a selection issue. This is because in this case, $u_i = \delta\epsilon_i$, so that there would still be a correlation between wages and the work decision.

We use the whole sample of single mothers aged between 18 and 50 and with children under 18 from March 2010 CPS data for our estimation purposes. Table B1 reports summary statistics for this group.

Table B1: **Summary Statistics for Single Mothers with Children under 18**

Variable	Mean	s.d.	Variable	Mean	s.d.
Age	30.7	9.64	Black	0.21	0.41
High school graduate	0.29	0.46	Proportion working	0.62	0.49
College or university	0.51	0.50	Yearly hours of work (if > 0)	1,558	768
No. of children under 6	0.45	0.69	Wage per hour (if > 0)	13.63	8.31
No. of children under 18	1.67	0.91	Has a child under age 6	0.35	0.48
White	0.70	0.46	No. of observations	7,060	

Source: March 2010 CPS data on single women with at least one child aged below 18. We limit the sample to women who are not involuntarily unemployed but who are either working or voluntarily unemployed (out of the labor force).

The imputation is done using the Heckman two step estimation procedure (Wooldridge, 2002). First run a probit using work status as the dependent variable and construct the inverse mills ratio. In the second stage, run an OLS regression using log of wages as the dependent variable and controlling for demographics X and the inverse mills ratio.

In order to identify our selection correction term, we rely on the non-linearity of the inverse Mills ratio and on the use of an exclusion restriction in the work equation. The exclusion restriction needs to be a variable which may affect mother's work decision but not her wages. We use a dummy variable indicating whether a mother has a child under 6 for this purpose. While the total number of children may be correlated with a women's past work decisions and therefore work experience and wages, once we control for the total number of children, we do not expect the presence of a child under 6 to immediately affect her wages although it may affect her current work decision. This corresponds to our variable K_i which captures child care needs of mothers.

Table B2 reports regression results for our selection and wage equations. As can be seen from our selection equation in column (i), having a child aged below 6 has a negative and statistically significant impact on the work decision of mothers. Moreover, from our wage equation in column (ii), the coefficient of the mills ratio is positive and significant suggesting that individuals who work tend on average to have higher wages.

Table B2: **Selected Coefficients from Work and Wage Regressions**

Dependent variable	(i) work		(ii) lnwage	
	coef	s.e.	coef	s.e.
Age	0.194	(0.015)	0.148	(0.021)
High school graduate	1.111	(0.346)	0.697	(0.294)
Undergraduate degree	1.731	(0.352)	1.225	(0.314)
No. of children under 18	-0.038	(0.056)	-0.022	(0.033)
Fair health	0.788	(0.118)	0.249	(0.129)
Good health	1.300	(0.110)	0.464	(0.167)
Very good health	1.300	(0.110)	0.606	(0.182)
Excellent health	1.490	(0.111)	0.621	(0.182)
White	0.251	(0.065)	0.065	(0.044)
Black	0.173	(0.075)	0.047	(0.046)
Any child under 6	-0.073	(0.037)		
Mills			0.529	(0.177)
No. of observations	7,060			

Standard errors reported in brackets. Controls also include age squared, number of children squared, average unemployment rate in state of residence and state dummies.

As discussed in the text, we are interested in the potential wage distribution of voluntarily unemployed mothers who would have been working if they did not have child care needs. We therefore impute their potential log wage as:

$$E\left[\ln wage_i | X_i\psi - \gamma K_i \leq u_i \leq X_i\psi\right] = X_i\beta - \rho\sigma_\epsilon \left[\frac{\phi\left(X_i\psi - \gamma K_i\right) - \phi\left(X_i\psi\right)}{\Phi\left(X_i\psi - \gamma K_i\right) - \Phi\left(X_i\psi\right)}\right].$$

From there, we can infer the potential wage distribution of voluntarily unemployed mothers by finding out the proportion of mothers with a given potential wage.

B.3 2010 US Tax and Benefits System

Unemployment benefits Unemployment benefits are set at \$5,500 such that, given the US tax and benefit system, the proportion of working mothers predicted by our model fit the proportion of single working mothers with a child below 6 (56%) in the CPS. Since families with two children receive on average \$412 TANF benefits per month (US Department of Health and Human Services, 2011), we interpret unemployment benefits as the sum of yearly TANF benefits of \$4,944 and of additional benefits of \$556 which may constitute of unemployment insurance benefits or food stamps that an unemployed individual may be eligible for. We do not explicitly set unemployment insurance benefits as young mothers may not be eligible for them if they have no work experience.

Federal taxes Taxable income is based on earnings minus standard deductions of $5,700 for a single childless person and of $8,400 for the head of household. Each taxpayer and dependent also get personal exemptions of $3,650. The tax rates are as follows (Taxes About):

Tax rate	Taxable income
10%	Less than $8,375
15%	$8,375 - $34,000
25%	$34,000 - $82,400
28%	$82,400 - $171,850
33%	$171,850 - $373,650
35%	$373,650 and above

Social Security taxes The Social Security base wage was $106,800 in 2010 and the employee rate 7.65% (Payroll Experts).

Earned Income Tax Credit The EITC is a refundable tax credit payable to working families. Earned income must be below $40,363 for a single parent with two children aged below 18 and the maximum credit was $5,036 in 2010 (Tax Policy Center, 2010). The phase-in rate was 40% and the phase-out rate 21.06% while the phase-out income range starts at $16,420. In the phase-in income range, EITC benefits are computed as 40% of earned income up to the maximum credit of $5,036. In the phase-out income range, EITC benefits are the difference between the maximum credit and 21.06% of income earned above $16,420. For a single childless individual, EITC benefis are phased-in at a rate of 7.65% up to a maximum of $457. Benefits are subsequently phased-out at a rate of 7.65% until earnings of $13,460 beyond which EITC benefits are zero.

Dependent Care Tax Credit The dependent care tax credit is a non-refundable tax credit as described in Section 2. It covers 35% of cost of formal child care up to a cap of $6k for two children in families earning less than $15,000. The tax credit rate declines by 1% for each $2,000 additional income until it reaches a constant rate of 20% for families with annual gross income above $43,000.

Child Care and Development Fund We set the CCDF rate to 90% which is the recommended subsidy rate under Federal guidelines. We take into account the fact that only a certain proportion of eligible households received the CCDF subsidy: 39% of potentially eligible children living in households below the poverty threshold, 24% of potentially eligible children living in households with income between 101 and 150% of the poverty threshold, and 5% of potentially eligible children living in household with income above 150% of the poverty threshold but below the CCDF eligibility threshold of 85% of state median income (US Department of Health and Human Services, 2009). We therefore compute the average CCDF subsidy rate as 35.1%, 21.6%, and 4.5% for households with income below, between 101% and 150%, and above 150% of the poverty threshold, respectively. The poverty threshold for a single parent with two children was $17,568 and US median earnings was $32,349 in 2010.

B.4 Calibration

Adverse labor market conditions We specify the probability, $\pi(0)$, of people suffering from adverse labor market conditions ($z_1 = 0$) as the proportion of involuntarily unemployed mothers in our CPS sample. The involuntarily unemployed include those who lost their jobs and those who entered or re-entered the labor force but could not get a job. This definition excludes those who voluntarily left their jobs or are out of the labor force. Around 11% of mothers with children under 6 were involuntarily unemployed and represent our mass of people suffering from adverse labor market conditions.

Empirical distribution of productivities In our model, individuals have heterogeneous market productivities when working outside of the house. We interpret market productivity types $z > 0$ as hourly wages when agents are not involuntarily unemployed. A standard approach to calibrating the skills distribution in the literature has been to fit a smooth function (typically log normal with upper Pareto tail) on empirical wages of the employed (Mankiw, Weinzierl and Yagan, 2009). This approach, however, does not consider the skills distribution among the voluntarily unemployed, which we interpret as those who voluntarily left their jobs or are out of the labor force.

In our framework, agents can be voluntarily unemployed if their wage is lower than their reservation wage which, in our model, depends on formal child care cost as well as the actual US tax and benefit system. In particular, we care about the *potential* wage distribution of mothers who would have worked if they did not have child care needs or faced a more generous child care subsidy scheme. This is because while it may be privately optimal for them not to work given the current real world situation, it may be efficient for some of them to work in the optimal program. We, therefore, impute the potential wage distribution of the voluntarily unemployed using two step selection correction methods à la Heckman, as described in Section B.2.

Figure B1 illustrates the wage distribution of mothers conditional on not being unlucky, $\pi(z)$ for $z > 0$. The wage distribution of the employed is based on their actual hourly wages, which are computed as yearly gross earnings divided by total hours of work in one year. The potential wage distribution of voluntarily unemployed mothers has been imputed. In the numerical exercise, we discretize the wage space into 50 wage centiles ranging between \$2.40 to \$32.21 such that we have 2% of mothers within each centile.

Child care needs Data from the Survey of Income and Program Participation indicates that on average, preschool age children spent 16 hours per week in the care of grandparents and other relatives (Rosenbaum and Ruhm, 2007; Laughlin, 2010). We interpret child care needs as the amount of child care time that can be substituted for paid care. Given a normal working week of 40 hours and family provided care of 16 hours per week, mothers need to make alternative child care arrangements for the remaining 24 hours per week. We, therefore, calibrate our model such that one unit of effort is equal to 24 hours per week.

Child care cost To calibrate average hourly cost of formal child care ω, we use the 2010 US average cost which ranged between \$6,380 for a four year old in family care homes and \$9,520 for an infant in child care centres (Child Care Aware of America, 2012).[4] Assuming that full time child care corresponds to 50 weeks

[4] State wide average annual costs for a four year old (infant) in full time centre based care ranged between \$3,900 (\$4,600) in Mississippi and \$11,700 (\$15,000) in Massachusetts.

Figure B1: March 2010 Current Population Survey data on women with at least one child aged below 6. Wages for non working mothers are imputed using Heckman selection correction methods.

of 40 hours each, average hourly cost is \$3.98 per child. Since single mothers have on average 1.28 children under age 6, we set $\omega = \$5.10$.

Calibration of θ The calibration of θ is done as follows: define a grid over $\theta \in [0.5, 2.5]$ with equally spaced intervals of 0.1. For each θ and z, we find the optimal labor supply predicted from the agent's private problem (B1), $l(\theta, z)$, and given the actual US tax and benefit system. Given the selection corrected empirical distribution of wages $\pi(z)$, we then compute the average labor supply predicted by our model for each θ, that is, we compute $\bar{l}(\theta) = \sum_z \pi(z)\, l(\theta, z)$. We find θ by minimizing the square of the distance between average labor supply predicted in our model and average labor supply in the data $\bar{l}_{data} = \sum_z \pi(z)\, l_{data}(z)$:

$$\hat{\theta} = argmin \left[\bar{l}(\theta) - \bar{l}_{data}\right]^2$$

After obtaining $\hat{\theta}$, we define a finer grid over θ (within a smaller interval that is inclusive of $\hat{\theta}$) with equally spaced intervals of 0.01 and repeat the procedure in order to get a more precise estimate of θ. The calibrated θ for different child care needs, child care cost and labor supply elasticity are reported in Table B3.

Calibration of M Net transfers that the US government already allocates to mothers are computed as follows. Given our calibrated values of θ, we simulate the chosen allocations $(c(z), y(z), f(z))$ for each type z, given the actual tax and benefit system. Based on the computed earnings $y(z)$, we then compute the Federal and SS taxes, EITC, DCTC and CCDF benefits as described above. Unemployed mothers receive unemployment benefits inclusive of TANF. Given the different net transfers received by each z, we take the average:

$$M = -\sum_{z} \pi(z)\, T^{E}(y(z), f(z))$$

where $T^{E}(y(z), f(z))$ are the net taxes computed based on the actual US tax and benefit system. The calibrated values of M for different child care needs, child care cost and labor supply elasticity are reported in Table B3.

Table B3: **Calibrated Effort Cost Parameter and Net Transfers**

	Baseline	$1e = 34$	$\omega = 6.4$	$\gamma = 2$	*log(c)*
θ	1.27	0.88	1.27	1.71	1.39
M	\$3,896	\$4,306	\$4,141	\$4,275	\$5,256

Note: θ is calibrated such that, given the 2010 US tax and benefit system and the distribution of wages, the average hours of work predicted from our model match the average hours of work observed for single mothers with at least one child under 6 in 2010 CPS. Baseline preferences with $1e = 24$, $\omega = \$5.1$, $\gamma = 1$, and $\alpha = 0$. In sensitivity analysis, we recalibrate θ and M by varying the parameter of reference while keeping the other ones at the baseline level.

B.5 Numerical Algorithm

We numerically solve for the constrained optimal allocations using Matlab. First, we impose non-negativity constraints on h, y. We then solve for the government problem using the following steps:

1. *Relaxed problem.* (a) Make an initial guess of values for the optimal allocations (c, y, h) and maximize welfare by imposing the government budget constraint with equality and the LDIC with inequality. (b) Use the solution in (a) as the new initial guess of values for (c, y, h) and maximize welfare by imposing the government budget constraint with equality and all the DIC with inequality. We have 1,275 DIC in total.

2. *Ex-post verification.* After having obtained the solution for the relaxed problem in point 1, we check whether all the UIC are satisfied. If all LDIC are binding we do not need to check the UIC as they are automatically satisfied from the proof of Lemma 1. Similarly, in the Ralwsian case, Lemma 1 guarantees this ex-post check step is not needed.

3. *Full problem.* If the conditions in Step 2 are not satisfied, then use the solution in Step 1(b) as the new initial guess of values for (c, y, h) and maximize welfare by imposing the government budget constraint with equality and all the DIC and UIC with inequality. That is, solve the full-blown problem. Excluding the 50 UIC for z_1, we have $2,500$ incentive constraints in total.

B.6 Further Results and Sensitivity Analyses

We now present results from the different specifications displayed in Table B4.

Table B4: **Sensitivity Analysis with Different Specifications**

	Specification	*Preferences*	*Social Welfare*	*Pareto*	*Description*
(1)	*Baseline*	Quasi-linear	Logarithmic	No	Baseline
(2)	$1e = 34$	Quasi-linear	Logarithmic	No	High child care needs
(3)	$\omega = 6.4$	Quasi-linear	Logarithmic	No	High cost of child care
(4)	$\gamma = 2$	Quasi-linear	Logarithmic	No	Low labor supply elasticity
(5)	*Rawls*	Quasi-linear	Rawlsian	No	Rawlsian
(6)	$log(c)$	Logarithmic	Utilitarian	No	Income effects
(7)	$\rho = 0$	Quasi-linear	Logarithmic	Yes	Baseline Pareto improving
(8)	$\rho = 0.99$	Quasi-linear	Almost Utilitarian	Yes	Low redistributive taste
(9)	$\rho = 1$	Logarithmic	Utilitarian	Yes	log(c) Pareto improving

Note: As reported in Table B3, we recalibrate θ and M for specifications (2), (3), (4), (6), and (9) by varying the parameter of reference while keeping the other ones at the baseline level.

Table B5: **Average Welfare Gains relative to the US Status Quo**

Unrestricted Second Best						Pareto Improving		
Baseline	$1e = 34$	$\omega = 6.4$	$\gamma = 2$	*Rawls*	$log(c)$	$\rho = 0$	$\rho = 0.99$	$\rho = 1$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
11.9%	13.3%	9.7%	7.7%	71.1%	9.1%	8.0%	3.6%	3.8%

Note: The table reports the consumption-equivalent welfare gains relative to the US status quo. The welfare gains are measured as the common relative increase in consumption across all z_i, that generates the same welfare level as under target optimum: $\sum_{i=1}^{N} \pi(z_i)\phi^*(z_i)U(c(z_i)(1+k), e(z_i)) = \sum_{i=1}^{N} \pi(z_i)\phi^*(z_i)U(c^*(z_i), e^*(z_i))$. The baseline specification in column (1) sets $1e = 24$, $\omega = \$5.1$ and $\gamma = 1$. In sensitivity analysis, we recalibrate θ and M by varying the parameter of reference while keeping the other ones at the baseline level. Columns (5) to (8) report, respectively, the welfare gains for the specifications with quasi-linear preferences and Rawlsian social welfare, log-consumption preferences and utilitarian social welfare, and Pareto improving optimal allocations with quasi-linear preferences and social welfare with $\rho = 0$ and $\rho = 0.99$, and log-consumption preferences and Utilitarian social welfare ($\rho = 1$).

Welfare gains relative to US Table B5 reports the average consumption-equivalent welfare gains relative to the US status quo. The welfare gains are measured as the common relative increase in consumption across all z_i, k, that generates the same welfare level as under target optimum:

$$\sum_{i=1}^{N} \pi(z_i)\phi^*(z_i)U(c(z_i)(1+k), e(z_i)) = \sum_{i=1}^{N} \pi(z_i)\phi^*(z_i)U(c^*(z_i), e^*(z_i)).$$

Following the literature, we interpret the increase in consumption as a windfall in the status quo while holding labor effort fixed.

Figure B2 reports the individual consumption-equivalent welfare gains relative to the US status quo. The welfare gains for each z_i are measured as the relative increase in consumption, $k(z_i)$, that generates the same welfare level as under target optimum:

$$U(c(z_i)(1+k(z_i)), e(z_i)) = U(c^*(z_i), e^*(z_i)).$$

Welfare Gains by Market Productivity Type Relative to the US Status Quo

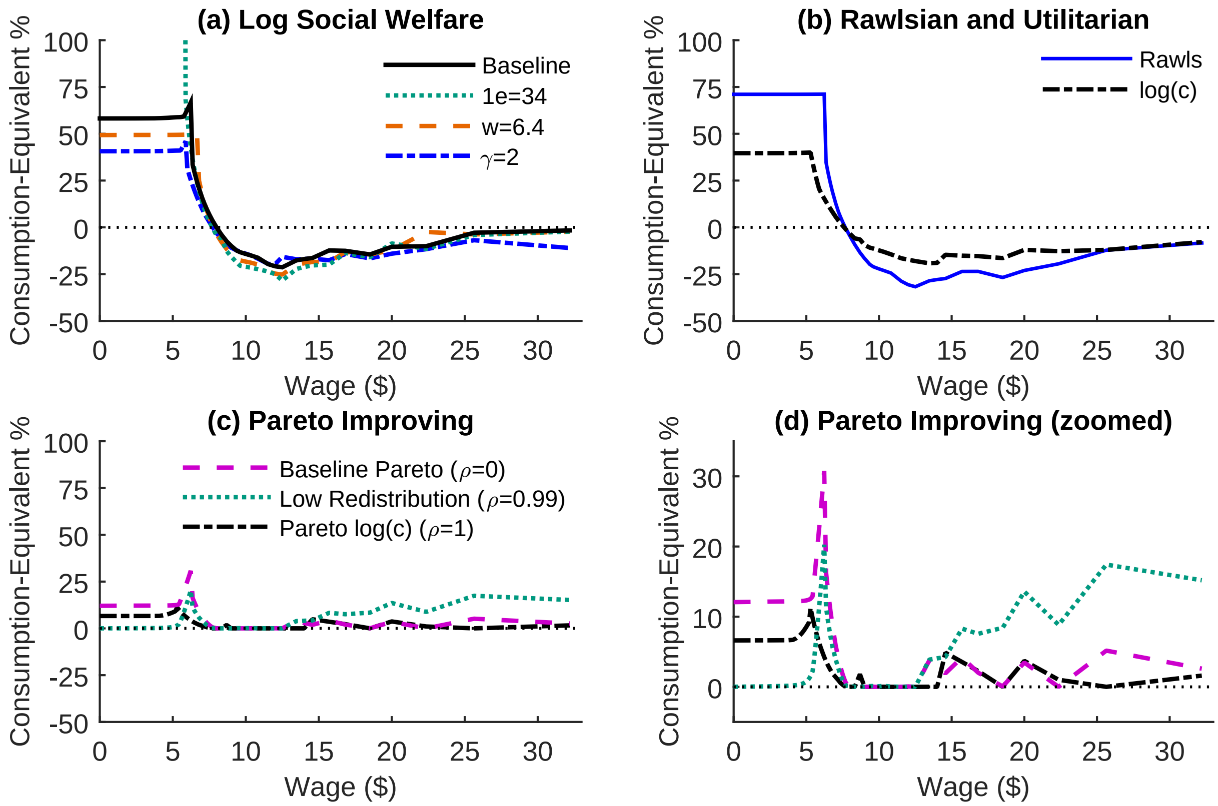

Figure B2: Panel (a) reports the welfare gains under quasi-linear preferences and logarithmic social welfare. Panel (b) reports the welfare gains under quasi-linear preferences and Rawlsian social welfare, and under log-consumption preferences and utilitarian social welfare. Panel (c) reports the welfare gains for the Pareto improving optimal allocations with quasi-linear preferences and social welfare function with $\rho = 0$ and $\rho = 0.99$, and log-consumption preferences and Utilitarian social welfare ($\rho = 1$). Panel (d) reports the same welfare gains as in Panel (c) but zoomed on the vertical axis. The welfare gains for each z_i are measured as the relative increase in consumption, $k(z_i)$, that generates the same welfare level as under target optimum: $U(c(z_i)(1+k(z_i)), e(z_i)) = U(c^*(z_i), e^*(z_i))$. For the specification with $1e = 34$, the welfare gains are truncated at 100% in Panel (a): The welfare gains are up to 351% for the lowest types because they have extremely low consumption under the US Status Quo.

The welfare gains originate from incentivizing mothers to work and earn more, which facilitiates redistribution to lower productivity mothers. The scope for redistribution from top to bottom is lower in the Pareto-improving specifications. Nevertheless, the Pareto-improving specifications still generate potentially sizeable welfare gains averaging 3.6% to 8%, as can be seen from Table B5.

Child care subsidies and marginal taxes Table B6 reports the optimal child care subsidy rates and Table B7 reports the marginal tax rates for the different specifications of the model. We note that employed agents tend to face higher marginal income tax rates under the Rawlsian criterion. This is consistent with income taxes also playing a redistributive role in our framework. Under the Rawlsian criterion, the welfare of the worse-off individual (i.e., the unemployed) is what matters. The employed are therefore taxed (and child care subsidized) to ensure that they are efficiently incentivized for output production.

Table B6: **Child Care Subsidy Rates**

Earnings	*USA*	*Baseline*	$1e = 34$	$\omega = 6.4$	$\gamma = 2$	*Rawls*	*log(c)*	$\rho = 0$	$\rho = 0.99$	$\rho = 1$
		Unrestricted Second Best						Pareto Improving		
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
\$5,000	0.70	0.80	0.79	0.65	1.00	0.80	1.00	0.80	0.79	0.52
\$10,000	0.70	0.60	0.59	0.68	0.92	0.60	0.71	0.60	0.60	0.71
\$15,000	0.70	0.40	0.38	0.52	0.83	0.40	0.43	0.40	0.40	0.32
\$20,000	0.53	0.20	0.18	0.36	0.69	0.20	0.08	0.20	0.20	0
\$25,000	0.51	0.02	0	0.20	0.52	0.06	0	0.02	0.04	0
\$30,000	0.31	0	0	0.10	0.30	0.02	0	0	0	0
\$35,000	0.25	0	0	0.02	0.06	0	0	0	0	0
\$40,000	0.22	0	0	0	0.02	0	0	0	0	0

Note: US subsidy rates in column (0) take into account the DCTC and CCDF. The baseline specification in column (1) sets $1e = 24$, $\omega = \$5.1$ and $\gamma = 1$. In sensitivity analysis, we recalibrate θ and M by varying the parameter of reference while keeping the other ones at the baseline level. Columns (5) to (9) report, respectively, the welfare gains for the specifications with quasi-linear preferences and Rawlsian social welfare, log-consumption preferences and utilitarian social welfare, and Pareto improving optimal allocations with quasi-linear preferences and social welfare with with $\rho = 0$ and $\rho = 0.99$, and log-consumption preferences and Utilitarian social welfare $\rho = 1$. Because z is discrete, we do not always observe a z with earnings level exactly equal to say $5k, 10k, 20k, 25k, 30k, 35k, 40k$. We use linear interpolation to approximate the subsidy rates in between discretized earnings levels where necessary.

Sensitivity analysis on wage bins We conduct further sensitivity analysis with respect to the efficient child allowances and child care subsidy rates. The implementation is as in Section 6.2, where we keep US net taxes (federal and SS taxes net of EITC if employed and TANF if unemployed) for the childless fixed. Figure B3 illustrates the baseline and US optimal child allowances and child care subsidy rates as in Figure 5. Recall that the number of wage bins as per our calibration in Section 6 is 50, which implies a maximum wage of \$32.21. As the optimal child care subsidy rate relies on z_N, we vary the number of wage bins to see how sensitive the rate is to the maximum wage. Wage bin grids of 25, 100, and 500 points yield z_N equal to \$28.80, \$35.60, and \$38.80, respectively. As can be seen from Figure B3, child allowances and subsidy rates for the different wage bins follow closely the baseline optimal ones.

Table B7: **Marginal Income Tax Rates for Employed Mothers**

			Unrestricted Second Best					Pareto Improving		
Earnings	*USA*	*Baseline*	$1e = 34$	$\omega = 6.4$	$\gamma = 2$	*Rawls*	*log(c)*	$\rho = 0$	$\rho = 0.99$	$\rho = 1$
	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$5,000	-0.32	0.28	0.29	0.41	0.19	0.55	0.19	0.04	0.03	0.03
$10,000	-0.32	0.42	0.42	0.42	0.42	0.48	0.39	-0.11	-0.16	-0.10
$15,000	0.08	0.38	0.37	0.38	0.39	0.42	0.39	0.02	0.03	0.02
$20,000	0.39	0.30	0.36	0.31	0.34	0.42	0.35	0.38	0.22	0.37
$25,000	0.39	0.35	0.38	0.36	0.41	0.43	0.33	0.37	0.24	0.37
$50,000	0.23	0.30	0.36	0.30	0.29	0.33	0.08	0.30	0.19	0.05
$75,000	0.33	0.24	0.24	0.24	0.12	0.26	0	0.26	0.15	0
$100,000	0.33	0.17	0.19	0.17	0	0.18	0	0.18	0.11	0

Note: US taxes in column (0) take into account Federal and SS tax rates as well as the EITC rates. The remaining columns reports the adjusted optimal marginal income tax rates (MTR) for different income levels, computed as the sum of the labor wedges and marginal child care subsidies. The baseline specification in column (1) sets $1e = 24$, $\omega = \$5.1$ and $\gamma = 1$. In sensitivity analysis, we recalibrate θ and M by varying the parameter of reference while keeping the other ones at the baseline level. Columns (5) to (9) report, respectively, the welfare gains for the specifications with quasi-linear preferences and Rawlsian social welfare, log-consumption preferences and utilitarian social welfare, and Pareto improving optimal allocations with quasi-linear preferences and social welfare with $\rho = 0$ and $\rho = 0.99$, and log-consumption preferences and Utilitarian social welfare $\rho = 1$. Because z is discrete, we do not always observe a z with earnings level exactly equal to say $5k, 10k, 20k, 25k, 30k, 35k, 40k$. We use linear interpolation to approximate the MTR in between discretized earnings levels where necessary.

Sensitivity Analysis on Wage Bins

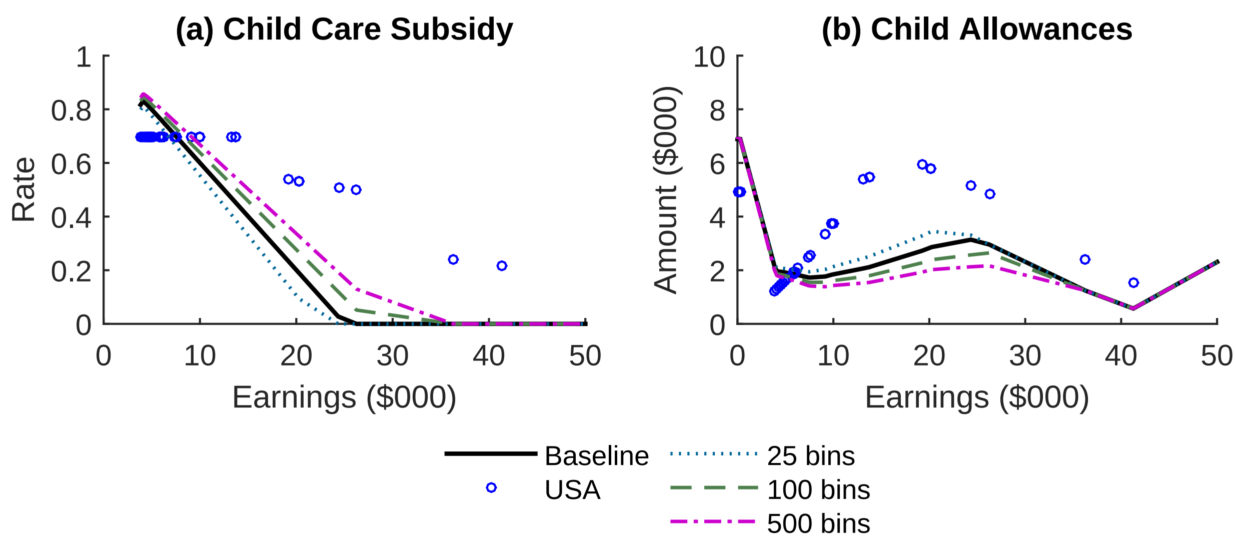

Figure B3: Panel (a) illustrates child care subsidy rates and Panel (b) illustrates child allowances. The US net income taxes are kept fixed (federal and SS taxes net of EITC if employed and TANF if unemployed) as in Figure 5. The rates and allowances are illustrated for the baseline optimal scheme, for the US, and for the baseline optimal scheme when using 25, 100 and 500 wage bins.

References

Child Care Aware of America. 2012. "Parents and the High Cost of Child Care." *Report*.

Laughlin, Lynda. 2010. "Who's Minding the Kids? Child Care Arrangements."

Mankiw, Gregory N., Matthew Weinzierl, and Danny Yagan. 2009. "Optimal Taxation in Theory and Practice." *Journal of Economic Perspectives*, 23(4): 147–174.

Matthews, Steven, and John Moore. 1987. "Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening." *Econometrica*, 55(2): 441–467.

Rosenbaum, Dan T., and Christopher Ruhm. 2007. "Family Expenditures on Child Care." *The BE Journal of Economic Analysis & Policy*, 7(1).

Tax Policy Center. 2010. "Earned Income Tax Credit Parameters." *Report*.

US Department of Health and Human Services. 2009. "Child Care Eligibility and Enrolment Estimates for Fiscal Year 2009." *Report*.

US Department of Health and Human Services. 2011. "Characteristics and Financial Circumstances of TANF Recipients: FY2010." *Report*.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge:MIT Press.