Online Appendix

Optimal Income Taxation with Unemployment and Wage Responses: A Sufficient Statistics Approach

Kory KROFT*
University of Toronto, NBER

Kavan KUCKO†
Cornerstone Research

Etienne LEHMANN‡
CRED(TEPP), Paris II

Johannes SCHMIEDER§
Boston University, NBER and IZA

*Email: kory.kroft@utoronto.ca. Address: 150 St. George Street, 354 Toronto, ON, Canada.

†Email: kavan.kucko@gmail.com. Address: Boston University, Department of Economics, 270 Bay State Road, Boston, MA 02215, USA.

‡Email: etienne.lehmann@gmail.com. Address: CRED Université Panthéon-Assas Paris II, 12 Place du Panthéon, 75 231 Paris Cedex, France. Etienne Lehmann is also research fellow at TEPP, IZA, CESifo, IDEP and CEPR. Financial support from Institut Universitaire de France and Center for Equitable Growth at University of California Berkeley are gratefully acknowledged.

§Email: johannes@bu.edu. Address: Boston University, Department of Economics, 270 Bay State Road, Boston, MA 02215, USA.

Contents

List of Tables

List of Figures

A Theory

In this section, we consider first various specialization of our model to clarify how some optimal tax models in the literature are nested into our general framework (I.1). There we discuss how our optimal tax formula (10) can be simplified in the specific environments. We then propose some extensions of our model (I.2) to discuss how our general optimal formula (10) is affected when we relax some of the key assumptions.

I.1 Special cases

The various specialization we successively consider are, the case with exogenous unemployment rate and constant returns to scale technology (I.1.a), to extend Saez (2004) to the case with exogenous unemployment rate, the job ratio

I.1.a The case without unemployment responses

In this subsection, we consider the case where wages can freely adjust, but the conditional employment probability is exogenous at $p_i \in (0,1]$ (so $\frac{\mathbf{d}\mathscr{P}}{\mathbf{dT}} = \mathbf{0}$) and where the different types of labor are substitutable. More specifically, we assume that the different types of labor h_i and capital Z produce a numeraire good sold in a perfectly competitive product market under a constant returns to scale technology $F(h_1,...,h_I,Z)$. We hence generalize Saez (2002) who considered perfect substitution across the difference types of labor through the production function: $F(h_1,...,h_I) = \sum_{i=1}^{I} w_i h_i$, where w_i stands both for the productivity of labor in occupation i and for the wage in the corresponding labor market. We also extend Saez (2004) for exogenous unemployment rates $1 - p_i \leq 1$. We furthermore assume the rate of return to capital, $r > 0$, is exogenous. The latter assumption can be viewed either by considering a small open economy and assuming perfect capital mobility, or by considering the steady state of a closed economy with infinite horizon savers. The assumptions of exogenous unemployment rates and constant returns to scale seem plausible in the long run, even though they ruled out job rationing considered by Landais et al. (2018b) which are plausible in the short run. We then get that:

Proposition A.1. *If the unemployment rates are exogenous, the production function exhibits constant returns to scale and* $\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}$ *is invertible, the optimal tax schedule is given by:*

$$0 = (1-g_j)h_j + \sum_{i=1}^{I}(T_i+b)\left.\frac{\partial \mathscr{H}_i}{\partial T_j}\right|^{Micro} \tag{A.1}$$

and depends only on microeconomic employment responses.

Proof: In the absence of unemployment responses to taxation $\frac{\partial \mathscr{P}_i}{\partial T_j} = 0$, the corrective terms $\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{\partial \mathscr{P}_i}{\partial T_j}\frac{u(c_i)-d_i-u(b)}{p_i u'(c_i)}$ in Equation (4) simplify to $\frac{\partial \mathscr{C}_i}{\partial T_j}$. We thus get: $\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}} = -\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \left.\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}}\right|^{\text{Micro}}$. Following Lemma 1, we thus get that: $\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}} = -\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \left.\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}\right|^{\text{Micro}}$. Equation (6) then successively leads to: $\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}} = -\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \left.\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}}\right|^{\text{Micro}}$. The optimal tax formula (10) then successively leads to:

$$\begin{aligned}
\mathbf{0} &= \mathbf{h} - \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \left.\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}}\right|^{\text{Micro}} \cdot (\mathbf{T}+\mathbf{b}) + \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \left.\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}\right|^{\text{Micro}} \cdot \left(\left.\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}\right|^{\text{Micro}}\right)^{-\mathbf{1}} \cdot (\mathbf{g}\,\mathbf{h}) \\
\mathbf{0} &= \mathbf{h} - \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \left.\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}}\right|^{\text{Micro}} \cdot (\mathbf{T}+\mathbf{b}) + \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot (\mathbf{g}\,\mathbf{h}) \\
\mathbf{0} &= \left(\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}\right)^{-\mathbf{1}} \cdot \mathbf{h} - \left.\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}}\right|^{\text{Micro}} \cdot (\mathbf{T}+\mathbf{b}) + \mathbf{g}\,\mathbf{h}
\end{aligned} \tag{A.2}$$

where the last equality requires the matrix $\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}$ to be invertible.

Moreover, the firm's profit function verifies $\Pi(w_1,...,w_I,r) \stackrel{\text{def}}{\equiv} \max_{h_1,...,h_I,Z} \quad F(h_1,...,h_I,Z) - \sum_{i=1}^{I} w_i h_i - rZ$. Applying the envelope theorem leads to $\frac{\partial \Pi}{\partial w_i} = -h_i$, thereby $d\Pi = -\sum_{i=1}^{I} h_i\, dw_i - Z\, dR$. Because of perfect competition and constant returns to scale, we get that $d\Pi = 0$, which together with the assumption of an inelastic return of capital (which implies $dR = 0$) leads to $0 = \sum_{i=1}^{I} h_i \frac{\partial \mathscr{W}_i}{\partial T_j}$. In matrix notation, this implies that $\mathbf{h}$ is an eigenvector of Matrix $\frac{\mathbf{d}\mathscr{W}}{\mathbf{dT}}$ associated to eigenvalue 0. Hence, $\mathbf{h}$ is an eigenvector of Matrix $\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}$ associated to eigenvalue -1, so $\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \mathbf{h} = -\mathbf{h}$ and eventually $\left(\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}\right)^{-\mathbf{1}} \cdot \mathbf{h} = -\mathbf{h}$. Therefore Equation (A.2) simplifies to:

$$\mathbf{0} = \mathbf{1} - \mathbf{g}\,\mathbf{h} + \left.\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}}\right|^{\text{Micro}} \cdot (\mathbf{T}+\mathbf{b})$$

which corresponds to (A.1).

This result may look surprising and is also due to the specific representation of the labor supply responses along the intensive margin in the occupation model of Saez (2002). Stiglitz (1982), Naito (1999) propose alternatively a two-skills version of the Mirrlees model with intensive labor supply responses where low skilled and high skilled labor are imperfect substitutes. Stiglitz (1982) shows that the labor supply of the high skilled workers needs to be upward distorted (negative marginal tax rate for high skilled workers), unless the elasticity of substitution across the two types of labor is infinite. This result of Stiglitz (1982) looks at odds with the result above. Saez (2004) explains this discrepancy by the fact that in Stiglitz (1982) when a high skill worker earns the gross income intended to a low-skilled one, he does so keeping her high skill productivity. In other words, a worker's skill is portable across the different income levels in Stiglitz

(1982) but not in Saez (2004). Therefore, a change in the low skilled gross wage affects the self-selection incentive constraint in Stiglitz (1982) and Naito (1999), as well as in the continuous income model of Rothschild and Scheuer (2013), while in the occupation model of Saez (2004) and Lee and Saez (2008), when an individual works in a low-skilled job, she has a low productivity. The occupation model captures not only extensive (participation) responses but also educational choice along the intensive margin in the long-run while the models of Stiglitz (1982) and Naito (1999) focus on the short-run hours of work and in-work effort responses along the intensive margin. □

I.1.b Job rationing under decreasing returns to scale technology

An older tradition in economics has proposed job rationing to explain unemployment. In contrast to the matching framework, the job-rationing framework assumes search frictions away and considers that each type of labor exhibits decreasing marginal productivity. In each labor market, employment is determined by the equality between the marginal product and the wage. Unemployment occurs whenever the wage is set above its market-clearing level. This theory of unemployment that Keynes (1936) attributed to Pigou was formalized in the disequilibrium theory (Barro and Grossman, 1971) and further developed in models that allowed for wages being set endogenously above the market clearing level (McDonald and Solow, 1981, Shapiro and Stiglitz, 1984, Akerlof and Yellen, 1990).[1]

To develop some intuition about the macro-micro participation gap in job-rationing models, we now consider a model with labor supply responses concentrated along the extensive margin, a single type of labor that exhibits a decreasing marginal productivity and a fixed gross wage w. This can occur for instance as a result of a minimum wage regulation. The fixed wage determines the level of employment h, independently of the number of participants.[2] We assume that individuals who participate face a heterogeneous participation cost χ that is sunk upon participation. The k participants face the same probability $p = h/k$ to be employed, whatever the participation cost χ they incur if they participate. In such a framework, a tax cut in T triggers a rise in participation at the micro level. However, provided that this tax cut occurs for a fixed wage, employment does not change, so the macro employment response is nil. Therefore, as the number of participants increases, the probability to be employed is reduced, which attenuates the participation responses at the macro level, as compared to the micro one. As a result, the optimal employment tax on the working poor is more likely to be positive in this job-rationing model without cross effect than in the pure extensive case.

There are different job-rationing models in the literature. For instance, in Lee and Saez (2012), there are different types of labor that are perfect substitutes, the minimum wage policy is explicitly an additional policy instrument and efficient rationing is assumed, so that the probability to be employed varies across participants as a function of their private cost upon working. Wages can also be made endogenous through union bargaining (McDonald and Solow, 1981) or through efficiency wages (Shapiro and Stiglitz, 1984, Akerlof and Yellen, 1990). Job rationing can also be analyzed within a search-matching framework if decreasing returns to scale is assumed for the production function, as in Michaillat (2012). As in a job-rationing model without matching, the macro employment effect would be dampened compared to the micro one and conditional employment probabilities would fall in response to a tax decrease. This in turn generates

[1] The Keynesian and New Keynesian theories of unemployment in addition assume nominal rigidities to give a transitional role to aggregate demand management policies. See also Michaillat and Saez (2015) for an extension of the new Keynesian model in which disequilibrium due to price rigidity are smoothed by matching functions on both the labor and the product market.

[2] Note that with a fixed wage, it is no longer equivalent whether the firm or the worker pays the tax. If the firm pays the tax, then a tax cut reduces the cost of labor and increases labor demand. In this case, the government controls not only the total tax liability in an occupation, but also the cost of labor and thereby the employment level. Lee and Saez (2012) provides conditions where the government finds it optimal to set the cost of labor above the market-clearing level, thereby generating unemployment in a job-rationing model.

a gap in the micro and macro participation response that captures the spillover effect on the labor market. While decreasing returns to scale may not be realistic in the long run, it may be plausible at least in the short-run during recessions with aggregate demand shortfalls. Landais et al. (2018b) discuss this possibility as a possible reason that the effect of unemployment insurance benefits on employment may be larger when the labor market is tight than when it is slack and thus the moral hazard associated with UI may be less severe during a crisis. For the same reason it may be that reductions in tax levels may have a larger effect on employment in recessions than in booms and the optimal policy during recessions may look more like an NIT.

I.1.c Search and Matching model with Proportional Bargaining

We now consider the case of a Diamond-Mortensen-Pissarides (DMP) search-and-matching model with a linear production function and proportional bargaining. Following (Diamond, 1982, Pissarides, 1985, Mortensen and Pissarides, 1999, Pissarides, 2000) we assume that for for each occupation i, a constant returns to scale matching function gives the employment level h_i as a function $\mathscr{M}_i(v_i,k_i)$ of the number v_i of vacancies posted and the number k_i of participating job seekers (Pissarides and Petrongolo, 2001). Creating a jobs costs $\kappa_i > 0$ and generates output $y_i > \kappa_i$ when a worker is recruited. Hence, the different types of labor are perfect substitutes.

Each vacancy is matched with probability $q_i = Q_i(\theta_i) \stackrel{\text{def}}{\equiv} \mathscr{M}_i(v_i,k_i)/v_i = \mathscr{M}_i(1,1/\theta_i)$, which is decreasing in tightness $\theta_i \stackrel{\text{def}}{\equiv} v_i/k_i$. Firms create jobs whenever the expected profit $q_i(y_i - w_i) - \kappa_i$ is positive. As more vacancies are created, tightness decreases until the free entry condition $q_i(y_i - w_i) = \kappa_i$ is verified. The conditional employment probability is an increasing function of tightness through $p_i = P(\theta_i) \stackrel{\text{def}}{\equiv} \mathscr{M}_i(v_i,k_i)/k_i = \mathscr{M}_i(\theta_i,1)$. Therefore, the conditional probability p_i is a decreasing function of the gross wage through:

$$p_i = \mathscr{L}_i(w_i) \stackrel{\text{def}}{\equiv} P_i\left(Q_i^{-1}\left(\kappa_i/(y_i - w_i)\right)\right) \tag{A.3}$$

which determines the labor demand function $p_i = \mathscr{L}_i(w_i)$.

Firms employ more workers the lower the gross wage (which makes it more rewarding for firms to hire a worker) and the more numerous job-seekers there are (which decreases the search congestions from the firm's viewpoint thereby easing their recruitment). In the model, the conditional employment probability p_i is a decreasing function $\mathscr{L}_i(\cdot)$ of the gross wage and is independent of the number of job-seekers. Therefore, a policy reform that increases labor supply, without affecting the gross wage, leads to a rise in employment in the same proportion as the rise in labor supply, but does not affect the employment probability.

If we consider a version of the matching model where wages are fixed, then the conditional employment probabilities are fixed, so the macro participation responses are equal to the micro ones. If we instead consider a version of the matching model where wage setting is based on wage bargaining, taxes may affect the outside option for workers as well as the match surplus and thus equilibrium wages and in turn conditional employment probabilities. To build intuition, consider the case with risk neutral workers (hence $u(c) \equiv c$) and proportional bargaining. In such a setting, workers receive an exogenous share $\beta_i \in (0,1)$ of the total match surplus $y_i - T_i - b$, so the wage is given by:[3]

$$w_i = \mathscr{W}_i(T_i,b) \equiv \beta_i\, y_i + (1-\beta_i)(T_i + b) \tag{A.4}$$

Combining the labor demand relation $p_i = \mathscr{L}_i(w_i)$ with the wage equation (A.4) and the assumption that labor supply responses are concentrated along the extensive margin provides a complete search-matching micro-foundation for the no-cross effect economy. The following proposition shows that the macro-micro

[3] A similar expression for wage bargaining appears in Jacquet et al. (2014) and in Landais et al. (2018b).

participation gap is directly linked to the bargaining weights and the elasticity of the matching function with respect to the number of job-seekers $\mu_i \in (0,1)$:

Proposition A.2. *In the search-matching economy with proportional bargaining (A.4), the micro and macro participation responses are equal either when the workers have full bargaining power so there is no wage responses, or when the Hosios (1990) condition $\beta_i = \mu_i$ is verified. If $\beta_i < \mu_i$ the macro response is lower then micro one. If $\mu_i < \beta_i < 1$ the macro response is larger then micro one.*

Proof: Under risk neutrality and proportional bargaining (A.4), one has for any $j \neq i$ that $\frac{\partial \mathscr{W}_i}{\partial T_j} = 0$, thereby $\frac{\partial \mathscr{P}_i}{\partial T_j} = 0$ from $p_i = \mathscr{L}_i(w_i)$, and finally $\frac{\partial \mathscr{U}_i}{\partial T_j} = 0$ from (4). Moreover, we get from $p_i = \mathscr{L}_i(w_i)$ and (4) that:

$$\frac{\partial \mathscr{U}_i}{\partial T_i} = \left[-1 + \frac{\partial \mathscr{W}_i}{\partial T_i}\left(1 + \frac{w_i}{p_i}\frac{\partial \mathscr{P}_i}{\partial w_i}\frac{w_i - T_i - b}{w_i}\right)\right] p_i$$

As $\mu_i \in (0,1)$ denote the elasticity of the matching function with respect to the number of job-seekers, we get $\frac{dp_i}{p_i} = (1-\mu_i)\frac{d\theta_i}{\theta_i}$ and $\frac{dq_i}{q_i} = -\mu_i \frac{d\theta_i}{\theta_i}$, so $\frac{dp_i}{p_i} = -\frac{1-\mu_i}{\mu_i}\frac{dq_i}{q_i}$. Log-differentiating the free-entry condition $k_i = q_i\,(y_i - w_i)$ leads to $\frac{dq_i}{q_i} = \frac{w_i}{y_i - w_i}\frac{dw_i}{w_i}$. So, we get $\frac{dp_i}{p_i} = -\frac{1-\mu_i}{\mu_i}\frac{w_i}{y_i-w_i}\frac{dw_i}{w_i}$, i.e: $\frac{w_i}{p_i}\frac{\partial \mathscr{P}_i}{\partial w_i} = -\frac{1-\mu_i}{\mu_i}\frac{w_i}{y_i-w_i}$. Moreover, when $\beta_i < 1$, Equation (A.4) implies that $\frac{w_i - T_i - b}{y_i - w_i} = \frac{\beta_i}{1-\beta_i}$ and $\frac{\partial \mathscr{W}_i}{\partial T_i} = 1 - \beta_i$. We thus finally get:

$$\frac{\partial \mathscr{U}_i}{\partial T_i} = \left[-1 + (1-\beta_i)\left(1 - \frac{1-\mu_i}{\mu_i}\frac{\beta_i}{1-\beta_i}\right)\right] p_i = \frac{\beta_i}{\mu_i}\left.\frac{\partial \mathscr{U}_i}{\partial T_i}\right|^{\text{Micro}} \tag{A.5}$$

□

An increase in tax liability has three effects on expected utility, thereby on participation decisions. First, absent wage and conditional employment response, a rise in T_i has a *direct* negative impact at the micro level (holding w_i and p_i constant) as it reduces the net wage and thus incentives to work and to participate. Second, at the macro level, the gross wage increases (through bargaining) attenuating the direct labor supply effect. Finally, the gross wage increase triggers a reduction in labor demand that amplifies the direct effect at the macro level. If the workers get all of the surplus (i.e. if $\beta_i = 1$), wages do not respond to taxation ($\frac{\partial \mathscr{W}_i}{\partial T_i} = 0$), the conditional employment probabilities are not affected so the micro and macro responses to participation are identical. On the other hand, if $\beta_i < 1$, the conditional employment probability effect dominates (is dominated by) the wage effect whenever the labor demand elasticity is (not) sufficiently elastic, which happens when the matching elasticity μ_i is higher (lower) than the bargaining share β_i. Propositions 2 and A.2 imply that the optimal employment tax rate on the working poor is more likely to be negative in the no-cross effect DMP case than in the pure extensive case if the workers' bargaining power is inefficiently high, i.e, is higher than the bargaining power prescribed by the Hosios (1990) condition.[4] Therefore, in the DMP model the macro-micro participation gap can be higher or lower than one, attenuating or reinforcing the arguments in favor of a negative participation tax at the bottom.[5]

---

[4]As $\frac{\pi_j}{\pi_j^m} = \frac{\beta_j}{\mu_j}$ from (4), Equation (12) becomes $\frac{T_j + b}{c_j - b} = \frac{1 - \frac{\beta_j}{\mu_j} g_j}{\eta_j}$ which corresponds to (19b) in Jacquet et al. (2014).

It is worth noting that under the Hosios (1990) condition $\beta_i = \mu_i$, while the macro and the micro *participation* elasticities are equal, this does not imply that the macro *employment* elasticities is equal to the micro *employment* elasticity. At the micro level, for fixed wages and tightness, a 1% increase in tax reduces employment only through the reduction in participation. The micro employment elasticity is therefore equal to the micro participation elasticity. Under the Hosios (1990) condition, the latter is equal to the macro participation elasticity. However, as a 1% increase in tax also decreases tightness because of the wage response to taxes, the conditional employment probability is also reduced, so the macro employment response is larger than the macro participation response.

[5]By extending this model with intensive labor supply decision, the present model can include the central mechanism of Golosov et al. (2013) where firms have different productivity and individuals direct their search.

I.1.d Search and Matching model with Nash Bargaining

Another strand in the literature has stressed the possibility that increases in tax progressivity may actually increase employment. For example in the monopoly union model, unions set the wage to maximize the expected utility of its members, which is increasing in the net wage and in the level of employment. Since the level of employment is decreasing in the gross wage, unions do not want to push the wage too high. If the tax schedule becomes more progressive, the wedge between net and gross wages increases more rapidly with the wage. Therefore, a one unit increase in the net wage will have to be traded off against a larger loss in employment. Thus, unions may actually accept a lower gross wage in response to an increase in tax progressivity, which may increase employment.[6] The main consequence of introducing the wage moderating effect of tax progressivity into the model is to make the matrix $\frac{\mathbf{d}\mathscr{W}}{\mathbf{dT}}$ and therefore the matrices $\frac{\mathbf{d}\mathscr{P}}{\mathbf{dT}}$, $\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}}$, $\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}$ and $\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}}$ non-diagonal.

To understand how the wage moderating effects of tax progressivity affects the optimal tax schedule, let us consider a matching model with two occupations $I = 2$ and a linear production function, so that the conditional employment probability in one labor market is the decreasing function $p_i = \mathscr{L}_i(w_i)$ of the gross wage on that labor market, as described in Equation (A.3). Assume that labor supply responses are concentrated along the extensive margin. Assume that the wage functions $\mathscr{W}_i$ not only verify $\frac{\partial \mathscr{W}_i}{\partial T_i} > 0$ for $i = 1,2$, as in the proportional bargaining case, but also that the marginal tax rate, as approximated by $T_2 - T_1$, has a wage moderating and unemployment reducing effect. This implies that $\frac{\partial \mathscr{W}_2}{\partial T_1} \geq 0 \geq \frac{\partial \mathscr{W}_1}{\partial T_2}$, with at least one strict inequality. Then we have $\frac{\partial \mathscr{P}_i}{\partial T_i} > 0$ and $\frac{\partial \mathscr{H}_i}{\partial T_i} < 0$ for $i = 1,2$ and $\frac{\partial \mathscr{P}_2}{\partial T_1} \leq 0 \leq \frac{\partial \mathscr{P}_1}{\partial T_2}$ with at least one strict inequality, thereby $\frac{\partial \mathscr{H}_2}{\partial T_1} < 0 < \frac{\partial \mathscr{H}_1}{\partial T_2}$, with at least one strict inequality. Suppose that only one of the latter inequalities is strict and that only the welfare of the non employed is valued, so that $g_0 > g_1 = g_2 = 0$. For notational compactness, we assume $\frac{\partial \mathscr{H}_i}{\partial T_j} < 0 = \frac{\partial \mathscr{H}_j}{\partial T_i}$ with $j \in \{1,2\} \setminus \{i\}$. Inverting the optimal tax formula (10) leads to:

$$T_i + b = -\frac{h_i}{\frac{\partial \mathscr{H}_i}{\partial T_i}} > 0 \qquad \text{and} \qquad T_j + b = -\underbrace{\frac{h_j}{\frac{\partial \mathscr{H}_j}{\partial T_j}}}_{>0} + \frac{\partial \mathscr{H}_i}{\partial T_j} \underbrace{\frac{h_i}{\frac{\partial \mathscr{H}_i}{\partial T_i}\frac{\partial \mathscr{H}_j}{\partial T_j}}}_{>0}$$

Therefore, when $\frac{\partial \mathscr{W}_1}{\partial T_2} = 0 < \frac{\partial \mathscr{W}_2}{\partial T_1}$, so that $\frac{\partial \mathscr{H}_1}{\partial T_2} = 0 > \frac{\partial \mathscr{H}_2}{\partial T_1}$, the wage moderating effect of tax progressivity captured by the negative term $\frac{\partial \mathscr{H}_2}{\partial T_1}$ has no effect on the optimal tax for the high-skilled and tends to reduce the optimal tax for the low-skilled. Conversely when $\frac{\partial \mathscr{W}_1}{\partial T_2} < 0 = \frac{\partial \mathscr{W}_2}{\partial T_1}$, so that $\frac{\partial \mathscr{H}_1}{\partial T_2} > 0 = \frac{\partial \mathscr{H}_2}{\partial T_1}$, the wage moderating effect of tax progressivity captured by the positive term $\frac{\partial \mathscr{H}_2}{\partial T_1}$ has no effect on the optimal tax for low-skilled and tends to increase the optimal tax for the high-skilled. In these two simplistic cases, we retrieve the general result shown by Hungerbühler et al. (2006) Lehmann et al. (2011) in more specialized search matching models, that compared to the proportional bargaining case, the case with a wage moderating/unemployment reducing effect of tax progressivity leads to a more progressive optimal tax schedule.

[6] This result has been obtained in a Monopoly unions model with job rationing by Hersoug (1984), in a union bargaining model by Lockwood and Manning (1993) or in the competitive directed search model (or wage posting) of Moen (1997) by Lehmann et al. (2011). A very similar result can also hold in the efficiency wage model of Pisauro (1991) or within the matching framework with Nash bargaining (Pissarides, 1985, 1998), or with the bargaining model of top income earners of Piketty et al. (2014). Evidence for this wage moderating effect of tax progressivity can be found in Malcomson and Sartor (1987), Holmlund and Kolm (1995), Hansen et al. (2000) and Brunello and Sonedda (2007), while Manning (1993) and Lehmann et al. (2016) provide some empirical support for the unemployment reducing effect of tax progressivity.

I.2 Extensions

We now consider various extensions of our model to explore how our optimal tax formula is affected. We first consider in I.2.a the case where unemployed worker in occupation i receive a different benefit b_i than non participants z. Next, we consider in I.2.b an extension where we allow search intensity to be continuous instead of dichotomous and where profitd can be taxed. That search intensity becomes a continuous choice does not modify the optimal tax formula. Conversely the tax on profits enriches the optimal tax formula (10) by a term capturing the effects of labor taxation on the corporate tax base. Finally, in section I.2.c, we retrieve the optimal UI formula of Landais et al. (2018b,a) as the latter model is then a special case of the extension we considered in I.2.b.

I.2.a Different Unemployment and Welfare Benefits

We now extend our baseline model to allow for different benefits for the unemployed and the non participants. To differentiate the benefits provided to the non-employed, the government needs to perfectly observe the participation decisions of individuals. In such a case, the government provides non-participants with welfare benefits z and provides unemployed in labor market i with unemployment benefits b_i. The policy vector therefore becomes $\mathbf{t} = (T_1, ..., T_i, b_1, ..., b_I, z)$ and the budget constraint (1) becomes:[7]

$$\sum_{i=1}^{n} T_i\, h_i = z\, k_0 + \sum_{i=1}^{n} (k_i - h_i) b_i + E \qquad \Leftrightarrow \qquad \sum_{i=1}^{n} (T_i + b_i)\, h_i + \sum_{i=1}^{n} (z - b_i) k_i = z + E \tag{A.6}$$

For a fixed participation level k_i in labor market i, an additional employed worker pays tax T_i and no longer receives unemployment benefits b_i. Therefore, for each additional employed worker in labor market i holding the number of participants constant, the government's revenue increases by the *employment tax* $T_i + b_i$. Symmetrically, for a fixed employment level h_i in labor market i, an additional participant receives the unemployment benefits b_i instead of the welfare benefits z. Therefore, the government's revenue changes by $z - b_i$ for each additional participant in labor market i holding the level of employment constant. The Lagrangian (7) of the government's problem thus becomes:

$$\Lambda(\mathbf{t}) \stackrel{\text{def}}{\equiv} \sum_{i=1}^{I} (T_i + b_i) \mathscr{H}_i(\mathbf{t}) + \sum_{i=1}^{I} (z - b_i) \mathscr{K}_i(\mathbf{t}) + \frac{1}{\lambda} \Omega \left(\mathscr{U}_1(\mathbf{t}), ..., \mathscr{U}_I(\mathbf{t}), u(z) \right)$$

Using (8), the first-order condition with respect to tax liability T_j in labor market j becomes:

$$0 = h_j + \sum_{i=1}^{I} (T_i + b_i) \frac{\partial \mathscr{H}_i}{\partial T_i} + \sum_{i=1}^{I} (z - b_i) \frac{\partial \mathscr{K}_i}{\partial T_i} - \sum_{i=1}^{I} \frac{\partial \mathscr{U}_i}{\partial T_j} \left(\left. \frac{\partial \mathscr{U}_i}{\partial T_j} \right|^{\text{Micro}} \right)^{-1} g_i h_i$$

Compared to (9), different benefits for the unemployed and the non-participants generates a new fiscal externality: for a fixed level h_i of employment in labor market i, each additional participant increases the number of unemployment benefits b_i recipients and reduces the number of welfare benefits recipients z. Moreover, the employment tax now depends on the unemployment benefits b_i. As Lemma 1 continues to hold, the optimal tax formula (10) becomes:

$$\mathbf{0} = \mathbf{h} + \frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}} \cdot (\mathbf{T} + \mathbf{b}) + \frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}} \cdot (\mathbf{z} - \mathbf{b}) - \frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}} \cdot \left(\left. \frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}} \right|^{\text{Micro}} \right)^{-\mathbf{1}} \cdot (\mathbf{g}\, \mathbf{h}) \tag{A.7}$$

[7] We here use that $k_0 = 1 - \sum_{i=1}^{I} k_i$.

Therefore, the same sufficient statistics, namely the macro employment responses $\frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}}$, the macro participation responses $\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}$ and the micro participation responses $\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}\Big|^{\text{Micro}}$ have to be estimated to implement the optimal tax formula. Clearly, when unemployment and welfare benefits are equal, we retrieve formula (10).

We now derive the formulas for the optimal unemployment and welfare benefit levels. Let g_i^b denote the welfare weights on the unemployed and let it be defined from the microeconomic effect of a rise in unemployment benefits. We thus get:

$$g_i^b \stackrel{\text{def}}{\equiv} \frac{1}{\lambda(k_i-h_i)} \frac{\partial \mathscr{U}_i}{\partial b_i}\bigg|^{\text{Micro}} \frac{\partial \Omega}{\partial U_i} \qquad \Leftrightarrow \qquad \frac{1}{\lambda}\frac{\partial \Omega}{\partial U_i} = \left(\frac{\partial \mathscr{U}_i}{\partial b_i}\bigg|^{\text{Micro}}\right)^{-1} (k_i-h_i) g_i^b$$

The first-order condition with respect to b_j writes:

$$0 = -(k_j-h_j) + \sum_{i=1}^{I}(T_i+b_i)\frac{\partial \mathscr{H}_i}{\partial b_j} + \sum_{i=1}^{I}(z-b_i)\frac{\partial \mathscr{K}_i}{\partial b_j} + \sum_{i=1}^{I}\frac{\partial \mathscr{U}_i}{\partial b_j}\left(\frac{\partial \mathscr{U}_i}{\partial b_i}\bigg|^{\text{Micro}}\right)^{-1} g_i^b(k_i-h_i)$$

The optimal unemployment benefits formula in matrix term is therefore very similar to the corresponding optimal tax formula (A.7):

$$\mathbf{0} = -(\mathbf{k}-\mathbf{h}) + \frac{\mathbf{d}\mathscr{H}}{\mathbf{db}}\cdot(\mathbf{T}+\mathbf{b}) + \frac{\mathbf{d}\mathscr{K}}{\mathbf{db}}\cdot(\mathbf{z}-\mathbf{b}) - \frac{\mathbf{d}\mathscr{K}}{\mathbf{db}}\cdot\left(\frac{\mathbf{d}\mathscr{K}}{\mathbf{db}}\bigg|^{\text{Micro}}\right)^{-\mathbf{1}}\cdot(\mathbf{g^b}\ (\mathbf{k}-\mathbf{h}))$$

Finally, the first-order condition on the welfare benefit z is simply:

$$0 = -k_0 + \sum_{i=1}^{I}(T_i+b_i)\frac{\partial \mathscr{H}_i}{\partial z} + \sum_{i=1}^{I}(z-b_i)\frac{\partial \mathscr{K}_i}{\partial z} + \frac{\partial \Omega}{\partial u(z)}\frac{u'(z)}{\lambda}$$

I.2.b Continuous Search Intensity and tax on profits

Up to now, profits did not appear in our model. We assumed that if firms make profits, these profits are untaxed and these profits are received by some "capital owners" whose welfare is not included in the social welfare function. Alternatively, the public finance literature has considered a polar assumption where profits are fully taxed, or, equivalently, where all production is controlled by the government (Diamond and Mirrlees, 1971). It is therefore important to consider an extension of our model where profits are taxed at an exogenous rate denoted $\tau \in [0,1]$. Our baseline model corresponds to the case where $\tau = 0$ while the case of fully taxed profits corresponds to $\tau = 1$.

To introduce a tax on profits, we need to specify how profits appear, so we have to specify the production technology and the matching technology. We consider a model where a representative firm produces a numeraire good using the different types of labor under the technology $F(h_1,...,h_I)$ which is increasing and weakly concave in each of its I arguments so $F_i > 0 \geq F_{ii}$. We assume that creating a vacancy costs $\kappa_i > 0$ to the firm.

To have a general model that includes Landais et al. (2018b) as a special case, we also introduce a continuous search intensity denoted e. Firms open v_i vacancies on labor market i, while the total amount of search units provided by the k_i participants is denoted S_i. In particular, in equilibrium, all participants choose the same amount of search units, in which case $S_i = e_i k_i$. The employment level h_i in labor market i is given by the matching function $\mathscr{M}_i(v_i,S_i)$ which is increasing in each of its two arguments and exhibits constant returns to scale. Let $\theta_i \stackrel{\text{def}}{\equiv} v_i/S_i$ be the "vacancy search units ratio". Each vacancy is filled with a probability $q_i = \mathscr{M}_i(v_i,S_i)/v_i = \mathscr{M}_i(1,1/\theta_i) = Q_i(\theta_i)$, where $Q_i(\cdot)$ is decreasing in the vacancy search units ratio. Symmetrically, the probability of finding a job per unit of search in labor market i is $a_i = \mathscr{M}_i(v_i,S_i)/S_i =$

$\mathscr{M}_i(\theta_i, 1) = P_i(\theta_i)$, where $P_i(\cdot)$ is increasing in the vacancy search units ratio. We define tightness in labor market i by the probability a_i of finding a job per unit of search.[8] The conditional probability to find a job is $p_i = e_i a_i$.

To hire one more worker, the firm has to post $1/Q_i(\theta_i)$ vacancies, which costs $\kappa_i/Q_i(\theta_i)$. Let $J_i(a) \overset{\text{def}}{\equiv} \kappa_i/Q_i\left(P_i^{-1}(a)\right)$ denote the hiring cost for the firm as an increasing function of tightness a and let $\mathscr{A}_i(\mathbf{t})$ be the reduced form describing how the tax policy affects tightness a_i in labor market i at the general equilibrium. The representative firm chooses labor demand $h_1, ..., h_I$ to maximize profits, taking wages $\mathbf{w} = (w_1, ..., w_I)$ and tightnesses $\mathbf{a} = (a_1, ..., a_I)$ as given:

$$\Pi(\mathbf{t}) \overset{\text{def}}{\equiv} \max_{h_1,...,h_I} \qquad F(h_1, ..., h_I) - \sum_{i=1}^{I} \left(\mathscr{W}_i(\mathbf{t}) + J_i\left(\mathscr{A}_i(\mathbf{t})\right)\right) h_i \tag{A.8}$$

The labor demand first-order conditions are:

$$F_i(h_1, ..., h_I) = w_i + J_i(a_i) \qquad \forall i \in \{1, ..., I\}$$

For a participant in labor market i, searching for a job with intensity e induces a search cost equal to $D_i(e)$ where function $D_i(\cdot)$ is assumed increasing and convex. Each participant chooses search intensity taking the wage w_i and the tightness a_i as given. Individual m expects $U_i - \chi_i(m)$ by searching a job in labor market i where:

$$U_i \overset{\text{def}}{\equiv} \max_{e} \qquad e\, a_i\, u(w_i - T_i) + (1 - e\, a_i)\, u(b) - D_i(e) \tag{A.9}$$

Let:

$$\mathscr{U}_i(\mathbf{t}) \overset{\text{def}}{\equiv} \max_{e} \qquad e\mathscr{A}_i(\mathbf{t})\, u\left(\mathscr{C}_i(\mathbf{t})\right) + \left(1 - e\mathscr{A}_i(\mathbf{t})\right) u(b) - D_i(e)$$

Denoting $e_i = \mathscr{E}_i(\mathbf{t})$ the effort choice made, the first-order condition for optimal search is:[9]

$$D'(e_i) = a_i\left[u(c_i) - u(b)\right] \qquad \Leftrightarrow \qquad e_i\, D'(e_i) - D(e_i) = U_i - u(b)$$

which eventually implies:[10]

$$\mathscr{E}_i(\mathbf{t})\, D'\left(\mathscr{E}_i(\mathbf{t})\right) - D\left(\mathscr{E}_i(\mathbf{t})\right) = \mathscr{U}_i(\mathbf{t}) - u(b) \tag{A.10a}$$

Using the envelope theorem, we get:

$$\frac{\partial \mathscr{U}_i}{\partial T_j}^{\text{Micro}} = \left.\frac{\partial \mathscr{U}_i}{\partial T_j}\right|_{\mathbf{w},\mathbf{a}} = -e_i\, a_i\, u'(c_i)\, \mathbb{1}_{i=j} \tag{A.10b}$$

$$\begin{aligned} \frac{\partial \mathscr{U}_i}{\partial T_j} &= e_i\, a_i\, u'(c_i)\left[\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{\partial \mathscr{A}_i}{\partial T_j}\frac{u(c_i) - u(b)}{a_i\, u'(c_i)}\right] \\ &= \left[\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{\partial \mathscr{A}_i}{\partial T_j}\frac{u(c_i) - u(b)}{a_i\, u'(c_i)}\right]\frac{\partial \mathscr{U}_i}{\partial T_i}^{\text{Micro}} \end{aligned} \tag{A.10c}$$

[8] It is usual in the matching literature to define instead tightness as the vacancy search units ratio. However, as there is a one-to-one increasing relation between the two, there is no loss of generality in choosing either of the two definitions.

[9] We here used that: $U_i - u(b) + D(e) = e_i a_i \left[u(c_i) - d_i - u(b)\right]$

[10] The derivative of $e \mapsto e\, D'(e) - D(e)$ is $eD''(e)$ which is positive from the convexity of $D(\cdot)$. Therefore, Equation (A.10a) uniquely determines the search intensity $\mathscr{E}_i(\mathbf{t})$ in labor market i.

From Equation (A.10a), search intensity in labor market i is a function denoted $\mathscr{E}_i$ of gross expected utility: $e_i = \mathscr{E}_i(U_i)$, so that $\mathscr{E}_i(\mathbf{t}) \stackrel{\text{def}}{\equiv} \mathscr{E}_i(\mathscr{U}_i(\mathbf{t}))$. We have that:

$$\begin{aligned}
\frac{\partial \mathscr{E}_i}{\partial T_j}^{\text{Micro}} &= -e_i\, a_i\, u'(c_i)\, \mathbb{1}_{i=j}\, \frac{\partial \mathscr{E}_i}{\partial U_i} \\
\frac{\partial \mathscr{E}_i}{\partial T_j} &= \frac{\partial \mathscr{U}_i}{\partial T_j}\frac{\partial \mathscr{E}_i}{\partial U_i} = \left[\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{\partial \mathscr{A}_i}{\partial T_j}\frac{u(c_i)-u(b)}{a_i\, u'(c_i)}\right] e_i\, f_i\, u'(c_i)\frac{\partial \mathscr{E}_i}{\partial U_i} \\
&= -\left[\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{\partial \mathscr{A}_i}{\partial T_j}\frac{u(c_i)-u(b)}{a_i\, u'(c_i)}\right]\frac{\partial \mathscr{E}_i}{\partial T_i}^{\text{Micro}}
\end{aligned}$$

Using (5), we thus get in matrix terms:

$$\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}} \cdot \left(\left.\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}}\right|^{\text{Micro}}\right)^{-\mathbf{1}} = \frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}} \cdot \left(\left.\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}\right|^{\text{Micro}}\right)^{-\mathbf{1}} = \frac{\mathbf{d}\mathscr{E}}{\mathbf{dT}} \cdot \left(\left.\frac{\mathbf{d}\mathscr{E}}{\mathbf{dT}}\right|^{\text{Micro}}\right)^{-\mathbf{1}} \qquad \text{(A.10d)}$$

According to Equation (2), the number of participants in labor market i depends on taxation only through the responses of gross expected utility U_i to taxation. Lemma 1 thus continues to hold.

With the additional tax revenues from profits, and assuming the same benefits for all the non-employed to save on notations, the Lagrangian (7) becomes:

$$\Lambda(\mathbf{t}) \stackrel{\text{def}}{\equiv} (T_i + b)\mathscr{H}_i(\mathbf{t}) + \tau\, \Pi(\mathbf{t}) + \frac{1}{\lambda}\Omega\left(\mathscr{U}_1(\mathbf{t}), ..., \mathscr{U}_I(\mathbf{t}), u(b)\right)$$

Using Hotelling's lemma[11] and Equation (8), the condition for the optimal tax liability T_j is:

$$0 = h_j + \sum_{i=1}^{I}(T_i + b)\frac{\partial \mathscr{H}_i}{\partial T_j} - \tau\sum_{i=1}^{I}\left(\frac{\partial \mathscr{W}_i}{\partial T_j} + J_i'(a_i)\frac{\partial \mathscr{A}_i}{\partial T_j}\right)h_i - \sum_{i=1}^{I}\frac{\partial \mathscr{U}_i}{\partial T_j}\left(\left.\frac{\partial \mathscr{U}_i}{\partial T_i}\right|^{\text{Micro}}\right)^{-1} g_i h_i \qquad \text{(A.11)}$$

Using matrix notations and Equation (5) in Lemma 1 (Equivalently (A.10d)), this condition becomes:

$$\mathbf{0} = \mathbf{h} + \frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}} \cdot (\mathbf{T} + \mathbf{b}) - \tau\left[\frac{\mathbf{d}\mathscr{W}}{\mathbf{dT}} \cdot \mathbf{h} + \frac{\mathbf{d}\mathscr{A}}{\mathbf{dT}} \cdot (\mathbf{J_i'}(\mathbf{a_i})\ \mathbf{h_i})\right] - \frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}} \cdot \left(\left.\frac{\mathbf{d}\mathscr{K}}{\mathbf{dT}}\right|^{\text{Micro}}\right)^{-\mathbf{1}} \cdot (\mathbf{g}\ \mathbf{h}) \qquad \text{(A.12)}$$

A first striking feature of (A.12) is that search intensity responses do not appear explicitly. This is because the way taxation affects search intensity is very similar to the way taxation affect participation decisions, as shown in Equation (A.10d). Consequently, behavioral effects in terms of search intensity are encapsulated in the macro employment responses. Moreover, provided that $\frac{\mathbf{d}\mathscr{E}}{\mathbf{dT}}^{\text{Micro}}$ is invertible, which is not the case when search intensity is exogenous, the optimal tax formula (A.12) can be equivalently expressed as follows:

$$\mathbf{0} = \mathbf{h} + \frac{\mathbf{d}\mathscr{H}}{\mathbf{dT}} \cdot (\mathbf{T} + \mathbf{b}) - \tau\left[\frac{\mathbf{d}\mathscr{W}}{\mathbf{dT}} \cdot \mathbf{h} + \frac{\mathbf{d}\mathscr{A}}{\mathbf{dT}} \cdot (\mathbf{J_i'}(\mathbf{a_i})\ \mathbf{h_i})\right] - \frac{\mathbf{d}\mathscr{E}}{\mathbf{dT}} \cdot \left(\left.\frac{\mathbf{d}\mathscr{E}}{\mathbf{dT}}\right|^{\text{Micro}}\right)^{-\mathbf{1}} \cdot (\mathbf{g}\ \mathbf{h}) \qquad \text{(A.13)}$$

This is related to results in Chetty (2008) who shows that one can use search effort behavioral responses to unemployment benefits to value Unemployment Insurance. However, since search intensity is typically unobserved, $\frac{d\mathscr{E}}{dT}$ and $\frac{d\mathscr{E}}{dT}^{\text{Micro}}$ cannot be estimated unless one imposes a normalization.[12]

[11] That is, applying the envelope theorem to (A.8) implies that: $\frac{\partial \Pi}{\partial T_j} = -\sum_{i=1}^{I}\left(\frac{\partial \mathscr{W}_i}{\partial T_j} + J_i(a_i)\frac{\partial \mathscr{A}_i}{\partial T_j}\right)h_i$.

[12] Chetty (2008) assumes that search effort is equal to the hazard rate out of unemployment, which can be estimated using labor market flows.

The second noteworthy feature of Equation (A.12) is that compared to optimal tax formulas (9) or (10), several new terms are present when profits are taxed. This is because a change in tax on labor of type j affects wages w_i and recruitment costs $J_i(a_i)$, which triggers a change in the profit tax base. The optimal tax formula (A.12) shows that one needs to additionally account for the macro response of wages to taxation $\frac{\mathbf{d}\mathscr{W}}{\mathbf{dT}}$, and market tightness, $\frac{\mathbf{d}\mathscr{A}}{\mathbf{dT}}$. Compared to labor supply responses, it is more difficult to identify these responses. For example, one needs to account for selection effects when trying to estimate the wage response to taxes.

There are two cases where these additional fiscal spillover responses do not appear in the optimal tax formula (A.12). The first case is when the tax on profits iz zero ($\tau = 0$) in which case we retrieve formula (**??**). The second case is when profits are fully taxed ($\tau = 1$) and when the behavioral responses to tax reforms are parameterized with respect to changes in utility levels $\blacksquare = (\Delta_1 = u(c_1) - u(b), ..., \Delta_I = u(c_I) - u(b)$ instead of being parameterized with respect to changes in tax liabilities $(T_1, ..., T_I)$. In such a case, the macro responses of wages to taxation disappear from the optimal tax formula because the resource constraint depends only on employment and on after-tax incomes, pre-tax incomes being absent. This argument was made in Landais et al. (2018b). The wage responses that show up when behavioral responses are parameterized with respect to changes in tax liabilities are now encapsulated in the behavioral responses parameterized with respect to changes in after-tax income.

I.2.c Retrieving Landais et al. (2018b,a)

Step 1: Specifying the version of the present model consistent with Landais et al. (2018b)

To retrieve the optimal policy rule of Landais et al. (2018b), we assume in this Appendix that $I = 1$, there are full participation $k = 1$the social welfare objective to be unweighted utilitarian, so Equation (8) simplifies to $-\left(\frac{\partial \mathscr{U}}{\partial T}\Big|^{\text{micro}}\right)^{-1} gh = \frac{1}{\lambda}$ and profits are fully taxed so that $\tau = 1$. As job search intensity is endogenous, individuals solve:

$$\mathscr{U}(T,b) = \max_e \qquad e\mathscr{A}(T,b)u(\mathscr{W}(T,b) - T) + (1 - e\mathscr{A}(T,b))\,u(b) - D(e) \tag{A.14}$$

Considering the derivatives of $\mathscr{W}$ and $\mathscr{A}$ are nil, we get:

$$\frac{\partial \mathscr{U}}{\partial T}\Bigg|^{\text{Micro}} = -h\,u'(c) \qquad \frac{\partial \mathscr{U}}{\partial b}\Bigg|^{\text{Micro}} - u'(b) = -h\,u'(b) \tag{A.15}$$

The first-order condition associated to (A.14) is:

$$D'(e) = \mathscr{A}(T,b)\left[u(w-T) - u(b)\right]$$

which leads to

$$\mathscr{E}(T,b)D'(\mathscr{E}(T,b)) - D(\mathscr{E}(T,b)) = \mathscr{U}(T,b) - u(b)$$

and so:

$$\frac{\frac{\partial \mathscr{E}(T,b)}{\partial T}}{\frac{\partial \mathscr{E}(T,b)}{\partial T}\Big|^{\text{Micro}}} = \frac{\frac{\partial \mathscr{U}(T,b)}{\partial T}}{\frac{\partial \mathscr{U}(T,b)}{\partial T}\Big|^{\text{Micro}}} \qquad , \qquad \frac{\frac{\partial \mathscr{E}(T,b)}{\partial b}}{\frac{\partial \mathscr{E}(T,b)}{\partial b}\Big|^{\text{Micro}}} = \frac{\frac{\partial \mathscr{U}(T,b)}{\partial b} - u'(b)}{\frac{\partial \mathscr{U}(T,b)}{\partial b}\Big|^{\text{Micro}} - u'(b)}$$

Combing the latter equations with (A.15) leads to:

$$\frac{\partial \mathscr{U}}{\partial T} = -h\,u'(c)\frac{\frac{\partial \mathscr{E}(T,b)}{\partial T}}{\frac{\partial \mathscr{E}(T,b)}{\partial T}\Big|^{\text{Micro}}} \qquad , \qquad \frac{\partial \mathscr{U}}{\partial b} = -h\,u'(b)\frac{\frac{\partial \mathscr{E}(T,b)}{\partial b}}{\frac{\partial \mathscr{E}(T,b)}{\partial b}\Big|^{\text{Micro}}} + u'(b) \tag{A.16}$$

The government's Lagrangian is:

$$\Lambda(T,b) \stackrel{\text{def}}{\equiv} (T+b)\, \mathscr{H}(T,b) - b - E + \Pi(T,b) + \frac{1}{\lambda}\mathscr{U}(T,b)$$

where

$$\Pi(T,b) \stackrel{\text{def}}{\equiv} \max_{h} \qquad F(h) - \left(\mathscr{W}(T,b) + J\left(\mathscr{A}(T,b)\right)\right)h$$

with first-order and envelope conditions:

$$F'(h) = w + J \qquad \frac{\partial \Pi}{\partial T} = -\left(\frac{\partial \mathscr{W}}{\partial T} + J'(a)\frac{\partial \mathscr{A}}{\partial T}\right)h \qquad \frac{\partial \Pi}{\partial b} = -\left(\frac{\partial \mathscr{W}}{\partial b} + J'(a)\frac{\partial \mathscr{A}}{\partial b}\right)h$$

The first-order conditions associated to the government's program are:[13]

$$T \quad : \qquad 0 = h + (T+b)\frac{\partial \mathscr{H}}{\partial T} - \left(\frac{\partial \mathscr{W}}{\partial T} + J'(a)\frac{\partial \mathscr{A}}{\partial T}\right)h + \frac{1}{\lambda}\frac{\partial \mathscr{U}}{\partial T} \tag{A.17a}$$

$$b \quad : \qquad 0 = h - 1 + (T+b)\frac{\partial \mathscr{H}}{\partial b} - \left(\frac{\partial \mathscr{W}}{\partial b} + J'(a)\frac{\partial \mathscr{A}}{\partial b}\right)h + \frac{1}{\lambda}\frac{\partial \mathscr{U}}{\partial b} \tag{A.17b}$$

Using (A.16) and $\frac{\partial \mathscr{C}}{\partial T} = \frac{\partial \mathscr{W}}{\partial T} - 1$, the system (A.17) can be rewritten as:

$$T: \quad 0 \quad = \quad (T+b)\frac{\partial \mathscr{H}}{\partial T} - \frac{\partial \mathscr{C}}{\partial T}h - J'(a)\frac{\partial \mathscr{A}}{\partial T}h - \frac{u'(c)}{\lambda}\frac{\frac{\partial \mathscr{E}}{\partial T}}{\left.\frac{\partial \mathscr{E}}{\partial T}\right|^{\text{Micro}}}h \tag{A.18a}$$

$$b: \quad 0 \quad = \quad -(1-h) + (T+b)\frac{\partial \mathscr{H}}{\partial b} - \left(\frac{\partial \mathscr{W}}{\partial b} + J'(a)\frac{\partial \mathscr{A}}{\partial b}\right)h + \frac{u'(b)}{\lambda}\left(1 - h\frac{\frac{\partial \mathscr{E}}{\partial b}}{\left.\frac{\partial \mathscr{E}}{\partial b}\right|^{\text{Micro}}}\right) \tag{A.18b}$$

Step 2: Relating behavioral responses with respect to T and b to the responses in Landais et al. (2018b) in terms of $\Delta = u(c) - u(b)$

Landais et al. (2018b) assume that T and b affect the wage only through the difference Δ of current utility:

$$\Delta(T,b) \stackrel{\text{def}}{\equiv} u\left(\mathscr{C}(T,b)\right) - u(b) = u\left(\mathscr{W}(T,b) - T\right) - u(b) \tag{A.19}$$

We now try to rewrite the behavioral responses in (A.18) in terms of responses to Δ. We denote the latter with a hat. Differentiating (A.19) implies:

$$d\Delta \quad = \quad u'(c)\left(\frac{\partial \mathscr{W}}{\partial T} - 1\right)dT + \left(u'(c)\,\frac{\partial \mathscr{W}}{\partial b} - u'(b)\right)db \tag{A.20}$$

Let $\hat{\mathscr{W}}(\cdot)$ be the function expressing the gross wage as a function of Δ, we get for any dT and db:

$$dw = \frac{\partial \mathscr{W}}{\partial T}\,dT + \frac{\partial \mathscr{W}}{\partial b}db = \frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\,d\Delta = \frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\left[u'(c)\left(\frac{\partial \mathscr{W}}{\partial T} - 1\right)dT + \left(u'(c)\frac{\partial \mathscr{W}}{\partial b} - u'(b)\right)db\right]$$

[13] Equation (A.17a) can be directly retrieved from Equation (A.11) as $I = 1$, $\tau = 1$ and $\Omega \equiv U$ so that $-\left(\left.\frac{\partial \mathscr{U}}{\partial T}\right|^{\text{micro}}\right)^{-1} gh = \frac{1}{\lambda}$. Equation (A.17b) follows the extension of (**??**) for the inclusion of profits.

As these equalities have to hold for each dT and db, we get:[14]

$$\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) \left(\frac{\partial \mathscr{W}}{\partial T} - 1\right) = \frac{\partial \mathscr{W}}{\partial T} \quad \Rightarrow \quad \frac{\partial \mathscr{W}}{\partial T} = \frac{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \quad \frac{\partial \mathscr{C}}{\partial T} = \frac{1}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \tag{A.21a}$$

$$\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\left[u'(c)\, \frac{\partial \mathscr{W}}{\partial b} - u'(b)\right] = \frac{\partial \mathscr{W}}{\partial b} \quad \Rightarrow \quad \frac{\partial \mathscr{W}}{\partial b} = \frac{\partial \mathscr{C}}{\partial b} = \frac{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \tag{A.21b}$$

We get from (A.20):

$$\begin{aligned} d\Delta &= u'(c)\left(\frac{\partial \mathscr{W}}{\partial T} - 1\right) dT + \left(u'(c)\, \frac{\partial \mathscr{W}}{\partial b} - u'(b)\right) db \\ &= \frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} dT + \frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} db \end{aligned}$$

Moreover, for $m = a, h, e$, as $\mathscr{M} = \mathscr{A}, \mathscr{H}, \mathscr{E}$ depends on T and b only through changes in Δ as described by functions $\hat{\mathscr{M}} = \hat{\mathscr{A}}, \hat{\mathscr{H}}, \hat{\mathscr{E}}$, we get:

$$dm = \frac{\partial \mathscr{M}}{\partial T} dT + \frac{\partial \mathscr{M}}{\partial b} db \quad = \quad \frac{\partial \hat{\mathscr{M}}}{\partial \Delta} d\Delta = \frac{\partial \hat{\mathscr{M}}}{\partial \Delta}\left[\frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} dT + \frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} db\right]$$

So

$$\frac{\partial \mathscr{H}}{\partial T} = \frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \frac{\partial \hat{\mathscr{H}}}{\partial \Delta} \qquad \frac{\partial \mathscr{H}}{\partial b} = \frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \frac{\partial \hat{\mathscr{H}}}{\partial \Delta} \tag{A.21c}$$

$$\frac{\partial \mathscr{A}}{\partial T} = \frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \frac{\partial \hat{\mathscr{A}}}{\partial \Delta} \qquad \frac{\partial \mathscr{A}}{\partial b} = \frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \frac{\partial \hat{\mathscr{A}}}{\partial \Delta} \tag{A.21d}$$

$$\frac{\partial \mathscr{E}}{\partial T} = \frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \frac{\partial \hat{\mathscr{E}}}{\partial \Delta} \qquad \frac{\partial \mathscr{E}}{\partial b} = \frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1} \frac{\partial \hat{\mathscr{E}}}{\partial \Delta} \tag{A.21e}$$

We repeat this exercise for microeconomic responses. Differentiating (A.19) implies:

$$d\Delta^{\text{Micro}} = -u'(c)\, dT - u'(b)\, db$$

so

$$\begin{aligned} de^{\text{Micro}} = \left.\frac{\partial \mathscr{E}}{\partial T}\right|^{\text{Micro}} dT + \left.\frac{\partial \mathscr{E}}{\partial b}\right|^{\text{Micro}} db &= \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} d\Delta^{\text{Micro}} \\ &= \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} \left[-u'(c)\, dT - u'(b)\, db\right] \end{aligned}$$

so

$$\left.\frac{\partial \mathscr{E}}{\partial T}\right|^{\text{Micro}} = -u'(c)\, \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} \qquad \left.\frac{\partial \mathscr{E}}{\partial b}\right|^{\text{Micro}} = -u'(b)\, \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} \tag{A.21f}$$

[14] To get the usual property that $\frac{\partial \mathscr{W}}{\partial T} > 0 > \frac{\partial \mathscr{C}}{\partial T}$, one needs $\frac{\partial \hat{\mathscr{W}}}{\partial \Delta} < 0$. Hence, we have $\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}\, u'(c) - 1 < 0$.

Step 3: Rewriting our optimal tax formulas in Step 1 using behavioral responses of Step 2

We now rewrite the equations in the system (A.18) using (A.21). For Equation (A.18a), we successively get:

$$
\begin{aligned}
0 &= (T+b)\frac{\partial \mathscr{H}}{\partial T} - \frac{\partial \mathscr{C}}{\partial T}h - J'(a)\frac{\partial \mathscr{A}}{\partial T}h - \frac{u'(c)}{\lambda}\frac{\frac{\partial \mathscr{E}}{\partial T}}{\left.\frac{\partial \mathscr{E}}{\partial T}\right|^{\text{Micro}}}h \\
&= (T+b)\frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{\partial \hat{\mathscr{H}}}{\partial \Delta} - \frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{h}{u'(c)} - J'(a)\frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{\partial \hat{\mathscr{A}}}{\partial \Delta}h - \frac{u'(c)}{\lambda}\frac{\frac{u'(c)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}}{-u'(c)\left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}}}h \\
&= (T+b)\frac{\partial \hat{\mathscr{H}}}{\partial \Delta} - \frac{h}{u'(c)} - J'(a)\frac{\partial \hat{\mathscr{A}}}{\partial \Delta}h + \frac{1}{\lambda}\frac{\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}}{\left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}}}h
\end{aligned}
$$

which leads to (A.22a) below. For Equation (A.18b), we successively get:

$$
\begin{aligned}
0 &= -(1-h) + (T+b)\frac{\partial \mathscr{H}}{\partial b} - \left(\frac{\partial \mathscr{W}}{\partial b} + J'(a)\frac{\partial \mathscr{A}}{\partial b}\right)h + \frac{u'(b)}{\lambda}\left(1 - h\frac{\frac{\partial \mathscr{E}}{\partial b}}{\left.\frac{\partial \mathscr{E}}{\partial b}\right|^{\text{Micro}}}\right) \\
1-h &= (T+b)\frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{\partial \hat{\mathscr{H}}}{\partial \Delta} - \frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}h - J'(a)\frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{\partial \hat{\mathscr{A}}}{\partial \Delta}h \\
&+ \frac{u'(b)}{\lambda}\left(1 - h\frac{\frac{u'(b)}{\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1}\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}}{-u'(b)\left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}}}\right)
\end{aligned}
$$

$$
\left(\frac{1-h}{u'(b)} - \frac{1}{\lambda}\right)\left(\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1\right) = (T+b)\frac{\partial \hat{\mathscr{H}}}{\partial \Delta} - \frac{\partial \hat{\mathscr{W}}}{\partial \Delta}h - J'(a)\frac{\partial \hat{\mathscr{A}}}{\partial \Delta}h + \frac{1}{\lambda}\frac{\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}}{\left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}}}h
$$

which leads to (A.22b)

$$
(T+b)\frac{\partial \hat{\mathscr{H}}}{\partial \Delta} - J'(a)\frac{\partial \hat{\mathscr{A}}}{\partial \Delta}h + \frac{1}{\lambda}\frac{\frac{\partial \mathscr{E}}{\partial \Delta}}{\left.\frac{\partial \mathscr{E}}{\partial \Delta}\right|^{\text{Micro}}}h = \frac{h}{u'(c)} \tag{A.22a}
$$

$$
= \left(\frac{1-h}{u'(b)} - \frac{1}{\lambda}\right)\left(\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1\right) + \frac{\partial \hat{\mathscr{W}}}{\partial \Delta}h \tag{A.22b}
$$

These two equations are mutually consistent only if

$$
\begin{aligned}
\frac{h}{u'(c)} &= \left(\frac{1-h}{u'(b)} - \frac{1}{\lambda}\right)\left(\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1\right) + \frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)\frac{h}{u'(c)} \\
0 &= \left(\frac{h}{u'(c)} + \frac{1-h}{u'(b)} - \frac{1}{\lambda}\right)\left(\frac{\partial \hat{\mathscr{W}}}{\partial \Delta}u'(c)-1\right)
\end{aligned}
$$

As $\frac{\partial \mathscr{W}}{\partial \Delta} u'(c) < 0$ and so $\frac{\partial \mathscr{W}}{\partial \Delta} u'(c) - 1 < 0$, we must have:

$$\frac{h}{u'(c)} + \frac{1-h}{u'(b)} = \frac{1}{\lambda} \tag{A.23}$$

which is Equation (12) in Landais et al. (2018b), i.e. the inverse Euler equation which is usual in moral hazards models and in New Dynamic Public Finance models whenever utility function are additively separable between consumption and effort, as this is here the case.

Step 4:

Following LMS, we define :

$$R \stackrel{\text{def}}{\equiv} 1 - \frac{c-b}{w} = \frac{T+b}{w} \qquad \varepsilon^{\text{Micro}} \stackrel{\text{def}}{\equiv} \frac{\Delta}{1-h} \left.\frac{\partial \hat{\mathscr{H}}}{\partial \Delta}\right|^{\text{Micro}} \qquad \varepsilon^{\text{Macro}} \stackrel{\text{def}}{\equiv} \frac{\Delta}{1-h} \frac{\partial \hat{\mathscr{H}}}{\partial \Delta} \qquad \varepsilon^{\text{f}} \stackrel{\text{def}}{\equiv} \frac{a}{e} \frac{\partial e}{\partial a} \tag{A.24}$$

Remark 1. *Rewriting the first-order condition on individuals' job search intensity as $D'(e) = a\ \Delta$, and using $h = a\ e$, Landais et al. (2018b) should have $\varepsilon^f = \varepsilon^{Micro} \frac{1-h}{h}$.*

Proof: The first-order condition $D'(e) = a\ \Delta$ on individuals' job search intensity from program (A.14) implies $\log(D'(e)) = \log(a) + \log(\Delta)$. Differentiating implies:

$$\begin{aligned}
\frac{D''(e)\ de}{D'(e)} &= \frac{da}{a} + \frac{d\Delta}{\Delta} \\
\frac{D''(e)\ e}{D'(e)} \frac{de}{e} &= \frac{da}{a} + \frac{d\Delta}{\Delta} \\
\frac{de}{e} &= \frac{D'(e)}{D''(e)\ e} \left(\frac{da}{a} + \frac{d\Delta}{\Delta} \right)
\end{aligned}$$

Hence:

$$\frac{D'(e)}{D''(e)\ e} = \frac{\Delta}{e} \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} = \frac{\Delta}{h} \left.\frac{\partial \hat{\mathscr{H}}}{\partial \Delta}\right|^{\text{Micro}} = \frac{1-h}{h} \varepsilon^{\text{Micro}} = \frac{a}{e} \frac{\partial e}{\partial a} = \varepsilon^{\text{f}}$$

Hence

$$\varepsilon^{\text{f}} = \frac{1-h}{h} \varepsilon^{\text{Micro}} \tag{A.25}$$

□

Let use denote:

$$\alpha \stackrel{\text{def}}{\equiv} \frac{\Delta}{a} \frac{\partial a}{\partial \Delta}$$

From $h = a\ e$, thereby $\frac{dh}{h} = \frac{da}{a} + \frac{de}{e}$, we get:

$$\frac{\Delta}{h} \left.\frac{\partial \hat{\mathscr{H}}}{\partial \Delta}\right|^{\text{Micro}} = \frac{\Delta}{e} \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} \qquad \Rightarrow \qquad \frac{\Delta}{e} \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} = \frac{1-h}{h} \varepsilon^{\text{Micro}}$$

and

$$\frac{\Delta}{h} \frac{\partial \hat{\mathscr{H}}}{\partial \Delta} = \frac{\Delta}{a} \frac{\partial \hat{\mathscr{A}}}{\partial \Delta} + \frac{\Delta}{e} \frac{\partial \hat{\mathscr{E}}}{\partial \Delta} \qquad \Rightarrow \qquad \varepsilon^{\text{Macro}} = \frac{h}{1-h} \alpha + \frac{h}{1-h} \frac{\Delta}{e} \frac{\partial \hat{\mathscr{E}}}{\partial \Delta}$$

Finally, applying the chain rule, we get:

$$\frac{\partial \hat{\mathscr{E}}}{\partial \Delta} = \frac{\partial \hat{\mathscr{E}}}{\partial a}\frac{\partial \hat{A}}{\partial \Delta} + \left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}} \qquad \Rightarrow \qquad \frac{\Delta}{e}\frac{\partial \hat{\mathscr{E}}}{\partial \Delta} = \varepsilon^{\text{f}}\,\alpha + \frac{1-h}{h}\varepsilon^{\text{Micro}}$$

Hence:

$$\begin{aligned}
\varepsilon^{\text{Macro}} &= \frac{h}{1-h}\alpha + \frac{h}{1-h}\frac{\Delta}{e}\frac{\partial \hat{\mathscr{E}}}{\partial \Delta} = \frac{h}{1-h}\alpha + \frac{h}{1-h}\varepsilon^{\text{f}}\,\alpha + \varepsilon^{\text{Micro}} = \frac{h}{1-h}\alpha(1+\varepsilon^{\text{f}}) + \varepsilon^{\text{Micro}} \\
&\Rightarrow \qquad \alpha = \frac{1-h}{h}\frac{\varepsilon^{\text{Macro}} - \varepsilon^{\text{Micro}}}{1+\varepsilon^{\text{f}}} \qquad \text{(A.26)} \\
\frac{\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}}{\left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}}} &= \frac{\frac{h}{1-h}\frac{\Delta}{e}\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}}{\frac{h}{1-h}\frac{\Delta}{e}\left.\frac{\partial \hat{\mathscr{E}}}{\partial \Delta}\right|^{\text{Micro}}} = \frac{\varepsilon^{\text{f}}\frac{h}{1-h}\alpha + \varepsilon^{\text{Micro}}}{\varepsilon^{\text{Micro}}} = 1 + \frac{\varepsilon^{\text{f}}}{1+\varepsilon^{\text{f}}}\left(\frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}} - 1\right) \qquad \text{(A.27)}
\end{aligned}$$

Therefore, we have to retrieve Equation (23) in Landais et al. (2018b) from (A.22a):

$$\begin{aligned}
\frac{T+b}{h}\frac{\partial \hat{\mathscr{H}}}{\partial \Delta} &= \frac{1}{u'(c)} - \frac{1}{\lambda}\frac{\frac{\partial \mathscr{E}}{\partial \Delta}}{\left.\frac{\partial \mathscr{E}}{\partial \Delta}\right|^{\text{Micro}}} + J'(a)\frac{\partial \hat{\mathscr{A}}}{\partial \Delta} \\
\frac{T+B}{w}\frac{w}{\Delta}\frac{1-h}{h}\frac{\Delta}{1-h}\frac{\partial \hat{\mathscr{H}}}{\partial \Delta} &= \frac{1}{u'(c)} - \frac{1}{\lambda} + \frac{1}{\lambda}\frac{\varepsilon^{\text{f}}}{1+\varepsilon^{\text{f}}}\left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right] + \frac{a\,J'(a)}{\Delta}\frac{\Delta}{a}\frac{\partial \hat{\mathscr{A}}}{\partial \Delta} \\
R\frac{w}{\Delta}\frac{1-h}{h}\varepsilon^{\text{Macro}} &= \frac{1}{u'(c)} - \frac{h}{u'(c)} - \frac{1-h}{u'(b)} + \frac{1}{\lambda}\frac{\varepsilon^{\text{f}}}{1+\varepsilon^{\text{f}}}\left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right] + \frac{a\,J'(a)}{\Delta}\frac{1-h}{h}\frac{\varepsilon^{\text{Macro}} - \varepsilon^{\text{Micro}}}{1+\varepsilon^{\text{f}}} \\
R\frac{w}{\Delta}\frac{1-h}{h}\varepsilon^{\text{Macro}} &= (1-h)\left[\frac{1}{u'(c)} - \frac{1}{u'(b)}\right] + \frac{1}{\lambda}\frac{\varepsilon^{\text{f}}}{1+\varepsilon^{\text{f}}}\left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right] + \frac{a\,J'(a)}{\Delta}\frac{1-h}{h}\frac{\varepsilon^{\text{Macro}} - \varepsilon^{\text{Micro}}}{1+\varepsilon^{\text{f}}} \\
R\frac{w}{\Delta}\varepsilon^{\text{Macro}} &= h\left[\frac{1}{u'(c)} - \frac{1}{u'(b)}\right] + \frac{1}{\lambda}\frac{h}{1-h}\frac{\varepsilon^{\text{f}}}{1+\varepsilon^{\text{f}}}\left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right] + \frac{a\,J'(a)}{\Delta}\frac{\varepsilon^{\text{Macro}} - \varepsilon^{\text{Micro}}}{1+\varepsilon^{\text{f}}} \\
R\frac{w}{\Delta}\varepsilon^{\text{Macro}} &= h\left[\frac{1}{u'(c)} - \frac{1}{u'(b)}\right] + \frac{1}{\lambda}\frac{h}{1-h}\frac{\varepsilon^{\text{f}}}{1+\varepsilon^{\text{f}}}\left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right] - \frac{a\,J'(a)}{\Delta}\frac{\varepsilon^{\text{Micro}}}{1+\varepsilon^{\text{f}}}\left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right] \\
R\frac{w}{\Delta}\varepsilon^{\text{Macro}} &= h\left[\frac{1}{u'(c)} - \frac{1}{u'(b)}\right] + \left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right]\left[\frac{1}{\lambda}\frac{h}{1-h}\frac{\varepsilon^{\text{f}}}{1+\varepsilon^{\text{f}}} - \frac{a\,J'(a)}{\Delta}\frac{\varepsilon^{\text{Micro}}}{1+\varepsilon^{\text{f}}}\right] \\
R\frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}} &= \frac{h}{\varepsilon^{\text{Micro}}}\frac{\Delta}{w}\left[\frac{1}{u'(c)} - \frac{1}{u'(b)}\right] + \left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right]\frac{1}{1+\varepsilon^{\text{f}}}\left[\frac{\Delta}{w\,\lambda}\frac{h}{1-h}\frac{\varepsilon^{\text{f}}}{\varepsilon^{\text{Micro}}} - \frac{a\,J'(a)}{w}\right]
\end{aligned}$$

Using (A.25) we get:

$$\begin{aligned}
R\frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}} &= \frac{h}{\varepsilon^{\text{Micro}}}\frac{\Delta}{w}\left[\frac{1}{u'(c)} - \frac{1}{u'(b)}\right] + \left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right]\frac{1}{1+\varepsilon^{\text{f}}}\left[\frac{\Delta}{w\,\lambda} - \frac{a\,J'(a)}{w}\right] \\
R &= \frac{h}{\varepsilon^{\text{Micro}}}\frac{\Delta}{w}\left[\frac{1}{u'(c)} - \frac{1}{u'(b)}\right] + \left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}}\right]\frac{1}{1+\varepsilon^{\text{f}}}\left[\frac{\Delta}{w\,\lambda} + (1+\varepsilon^{\text{f}})R - \frac{a\,J'(a)}{w}\right] \qquad \text{(A.28)}
\end{aligned}$$

which differs from Equation (23) of Landais et al. (2018b) only by the term $\frac{a\,J'(a)}{w}$ instead of $\frac{\eta}{1-\eta}\tau(\theta)$ in Landais et al. (2018b). Both terms corresponds to recruiting costs for the firm. However, in our model (as in most of matching models), recruiting costs are cost per vacancy posted. In Landais et al. (2018b), these costs are hiring cost of recruiting workers. So both are conceptually identical although slightly different in the details.

Rewriting Landais et al. (2018b) in terms of sufficient stat

Using our Equation (A.25), Equation (A.28) becomes:

$$R = \frac{h}{\varepsilon^{\text{Micro}}} \frac{\Delta}{w} \left[\frac{1}{u'(c)} - \frac{1}{u'(b)} \right] + \left[1 - \frac{\varepsilon^{\text{Macro}}}{\varepsilon^{\text{Micro}}} \right] \frac{h}{h + (1-h)\varepsilon^{\text{Micro}}} \left(\frac{\Delta}{w\,\lambda} + \frac{h + (1-h)\varepsilon^{\text{Micro}}}{h} R - \frac{a\,J'(a)}{w} \right)$$

B Simulations

We simulate the optimal tax schedule using a similar approach as Saez (2002). We denote the current tax system with the vector of occupation tax rates T_i^0. The corresponding density weights in the observed economy are given as $h_i^0 = \mathscr{H}_i(T_i^0)$.

II.1 System of Equations

The system of equations the determines the optimal tax schedule is given by the budget constraint:

$$b + E = \sum_{i=1}^{I} (T_i + b)\ \mathscr{H}_i(\mathbf{t}) \tag{B.1}$$

and the first order condition given by equation (15) in the main text for each of the I income groups set to zero. Finally we follow Saez (2002) and assume no income effects, in which case we can use the normalization: $\sum_i g_i h_i = 0$. With income effects one could alternatively use the first order condition for the nonemployment benefit b shown in the paper.

In order to solve the system of equations we also have to parameterize $g_i(T_i)$ and $h_i(T_i)$. For the former we follow Saez (2002) and assume that $g_i = \frac{1}{\lambda c_i^{\nu}}$ with the curvature parameter $\nu = 0.5$ - the version in the paper - and $\nu = 0.25$, $\nu = 1$ and $\nu = 4$ shown in the appendix. However, there is a complication, since $c_i = w_i(\mathbf{t}) - T_i$, but we do not have an estimate of how taxes affect pre-tax earnings. Therefore for the purpose of calculating the welfare weights, we will keep pre-tax earnings fixed at the observed levels and calculate c_i as $c_i = w_i(\mathbf{t_0}) - T_i$.

For the density weights in the optimal tax formula $h_i(T_i)$ we use the same approach as Saez (2002) and ignore intensive margin responses:

$$h_i(T_i) = h_i^0 * \left(\frac{w_i - T_i - (w_i - T_i^0)}{-T_0 + T_0^0} \right)^{\eta_i}$$

We calibrate the density weights in the observed economy $h_i^0 = \mathscr{H}_i(T_i^0)$ using the earnings distribution in the March 2011 CPS for single workers age 18 to 55 (corresponding to our empirical analysis). We use income bins of \$500 between 0 and \$25,000 income and bins of \$2000 for incomes up to \$100,000. We truncate the distribution at \$100,000. Due to round number bunching the resulting distribution is quite choppy which makes the simulated marginal tax rates also quite choppy. To deal with this we run a lowess smoother for incomes above 0 and below \$100,000. This has almost no effects on the resulting post-tax vs. pre-tax income figures, but smoothes out the MTR figures.

II.2 Calibrating the intensive margin mobility elasticity ζ_i

To calibrate the intensive margin mobility elasticity ζ_i, we adapt the approach of Saez (2002) to the case where $\mu \neq 0$.

Adapting Equation (12) in Saez (2002) to the case where $\mu \neq 0$

Consider a small increase[15] dT in tax liabilities in occupations $j = i, i+1, ..., I$. We thus have: $dT = dT_i = dT_{i+1} = ... = dT_I$. This tax change

1. raises $[h_i + h_{i+1} + ...h_I]dT$ additional taxes through the mechanical effect

$$[hi + h_{i+1} + ... + h_I]dT$$

by the government.

2. also induces extensive and income responses that change the level of employment in occupations $j = i, i+1, ..., I$ by $dh_j = (\eta_j h_j / (c_j - b))dT$. These responses in turn modify the government's revenue by

$$(T_j + b)dh_j = \frac{T_j + b}{c_j - b} \eta_j h_j dT$$

for all $j = i, i+1, ..., I$.

3. Finally, the marginal tax rate between occupation $i-1$ and i increases. This induce behavioral responses that can be decomposed in two terms. First, as in Saez (2002), a flow $(\zeta_i / (c_i - c_{i-1}))h_i$ of workers switch from occupation i to occupation $i-1$, which modifies government's revenue by

$$-\frac{T_i - T_{i-1}}{c_i - c_{i-1}} \zeta_i \, h_i \, dT$$

Moreover, some additional jobs in occupation $i-1$ may be created through the employment enhancing effect, which implies a change in tax revenue equal to

$$\frac{T_{i-1} + b}{c_i - c_{i-1}} \mu_i \, h_i \, dT$$

Equation (12) of Saez (2002) is only about the third effect, i.e. the response to change in $T_i - T_{i-1}$ and no to changes in T_j for all $j \geq i$. Hence, instead of equation (12), we get:

$$\left[-\frac{T_i - T_{i-1}}{c_i - c_{i-1}} \zeta_i + \frac{T_{i-1} + b}{c_i - c_{i-1}} \mu_i\right] h_i \, dT = \left[-\frac{\tau_i}{1 - \tau_i} (w_i - w_{i-1}) \zeta_i + \frac{T_{i-1} + b}{1 - \tau_i} \mu_i\right] h_i \, d\tau_i \tag{B.2}$$

where the second equality is due to denoting $\tau_i = (T_i - T_{i-1}) / (w_i - w_{i-1})$ and $dT = (w_i - w_{i-1}) d\tau_i$

Adapting Equation (13) in Saez (2002) to the case where $\mu \neq 0$

We now adopt a model with a continuous earnings distribution. We consider the effect of a small increase in marginal tax rate around wage w_i. This also triggers a uniform change in tax libability above w. However we here only concentrate on the effects of the change in the marginal tax rate around wage w. This first induces a compensated response of earnings by $dw = -\varepsilon_i \, w_i \frac{d\tau_i}{1-\tau_i}$ As there are h_i individuals with income w_i, the total effect on tax revenue is equal to:

$$\tau_i \, dw_i \, h_i = -\frac{\tau_i}{1 - \tau_i} w_i \, \varepsilon_i \, h_i \, d\tau_i$$

[15] The case $dT < 0$ where tax liabilities are decreased is obviously symmetric.

Second, this change in marginal tax rate modifies employment by $-\frac{\partial \ln(1-u_i)}{\partial \ln(1-\tau_i)} h_i \frac{d\tau_i}{1-\tau_i}$ with a total effect on tax revenue equal to:

$$-\frac{T_i + b}{1-\tau_i} \frac{\partial \ln(1-u_i)}{\partial \ln(1-\tau_i)} h_i \, d\tau_i$$

Hence, instead of Equation (13), we have:

$$\left[-\frac{\tau_i}{1-\tau_i} w_i \, \varepsilon_i - \frac{T_i + b}{1-\tau_i} \frac{\partial \ln(1-u_i)}{\partial \ln(1-\tau_i)} \right] h_i \, d\tau_i \tag{B.3}$$

Equations (B.2) and (B.3) coincide if

$$(w_i - w_{i-1}) \zeta_i = w_i \, \varepsilon_i \qquad \text{and} \qquad \mu_i = -\frac{\partial \ln(1-u_i)}{\partial \ln(1-\tau_i)} = \frac{h_i}{1-\tau_i} \frac{\partial(1-u_i)}{\partial \tau_i}$$

if one accepts the approximation $T_i + b \simeq T_{i-1} + b$

II.3 Calibrating μ_i

Lehmann et al. (2016) used the following notations. The tax schedule is denoted: $w \mapsto T(w)$ were w stands for labor earnings and $T(w)$ stands for tax liability. The retention rate is defined as:

$$ret(w) \stackrel{\text{def}}{\equiv} 1 - \frac{T(w)}{w}$$

The local CRIP (Coefficient of Residual Income Progression) at earnings w is defined as:

$$\Psi(w) \stackrel{\text{def}}{\equiv} \frac{1 - T'(w)}{1 - \frac{T(w)}{w}} = \frac{\partial \log(w - T(w))}{\partial \log w} = \frac{\partial \log ret(w)}{\partial \log w} + 1$$

while the global CRIP between earnings w_0 and w_1 is defined as:

$$\Psi_{w_0}^{w_1} \stackrel{\text{def}}{\equiv} \log\left(\frac{ret(w_1)}{ret(w_0)}\right) = \log\left(\frac{1 - \frac{T(w_1)}{w_1}}{1 - \frac{T(w_0)}{w_0}}\right) = \int_{w_0}^{w_1} (\Psi(t) - 1) \frac{dt}{t}$$

The global CRIP measures the reduction factor of the net wage ratio compared to the gross wage ratio between w_0 and w_1.

Table 3 Colmun 3 in Lehmann et al. (2016) provides an estimate of the elasticity of the global CRIP between 67% and 167% of the average wage on the employment to population ratio. Let denote this estimate as β; with $\beta = 0.2$ under OLS and $\beta = 1.05$ under IV. We have

$$\frac{dh}{h} = \beta \, d\Psi_{167\%}^{67\%}$$

Neglecting the variations of the local CRIP between 67% and 167% of the average wage, we get

$$\Psi_{167\%}^{67\%} = (\Psi - 1) \int_{167\%}^{67\%} d\log(t) = (1 - \Psi) \log\left(\frac{167}{67}\right) = (1 - \psi) \log\left(\frac{5}{2}\right) \simeq 0.91(1 - \Psi)$$

We thus get

$$\frac{dh}{h} \simeq -0.91 \, \beta \, d\Psi$$

Conversely, μ is about the compensated response of employment to change in marginal tax rate. For a compensated change, $dT' = -\left(1 - \frac{T(w)}{w}\right) d\Psi$. As the mean of $1 - \frac{T(w)}{w}$ in Lehmann et al. (2016) is equal to 0.62, we get:

$$\frac{dh}{h} \simeq \frac{0.91}{1 - \frac{T(w)}{w}} \beta \, dT' \simeq \frac{0.91}{0.62} \beta \, \frac{d(T_i - T_{i-1})}{w_i - w_{i-1}} \simeq 1.47 \, \beta \, \frac{d(T_i - T_{i-1})}{w_i - w_{i-1}}$$

So

$$\frac{dh_{i-1}}{h_{i-1}} \simeq \frac{dh_i}{h_i} \simeq \frac{1.47}{w_i - w_{i-1}} \beta \, d(T_i - T_{i-1})$$

$$dh_{i-1} + dh_i \simeq 1.47 \frac{h_i + h_{i-1}}{w_i - w_{i-1}} \beta \, d(T_i - T_{i-1})$$

$$\mu_i \simeq 1.47 \, \frac{c_i - c_{i-1}}{w_i - w_{i-1}} \, \frac{h_i + h_{i-1}}{h_i} \, \beta$$

i.e. to get correct order of magnitude,

$$\mu_i \simeq 1.47 \, (1 - T'(w)) \, 2 \, \beta$$

As the mean of $1 - T'(w)$ is close to 0.5 in Lehmann et al. (2016), this finally leads to

$$\mu_i \simeq 1.5 \;\; \beta \tag{B.4}$$

i.e. μ_i around 0.3 using OLS estimates of Lehmann et al. (2016).

C Description of Data Sources and Cleaning Steps

III.1 Data Sources

The empirical analysis combines information from several sources. This subsection describes each of the data sources used in this paper. In the subsections below, we describe how each of these are used to construct our final dataset.

1. Current Population Survey (CPS): The CPS is a monthly survey, sponsored by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), and is the main source of labor market statistics for the United States. The CPS sample is an overlapping panel of households that are randomly selected to participate in the survey. Information (including labor force status) is asked about each member of the household. For the first four months after their selection, households are surveyed monthly on the calendar week of the 19^{th} of each month about their labor market activities for the previous week. After their four months, households are not surveyed for eight consecutive months. Following the eight month of not being surveyed, households are surveyed again for four additional consecutive months. This is sometimes referred to as a $4-8-4$ sampling scheme. Households are asked about their regular weekly earnings and hours of work only in their fourth or eighth month of interviews. These households form the outgoing rotation group (ORG). Every March, the CPS supplements its standard questionnaire with additional questions on demographic characteristics and annual income, among others.[16] This supplement is referred to as the March annual data or the March Supplement.

[16] While questions about labor force status (the *empstat* variable described in more detail below) are the same for the ORG and March supplement, some variables are not. For example, as we discuss below, annual earnings (the *incwage* CPS variable) are only available for those in the March Supplement. We use this information to impute earnings for all ORG and March Supplement households in year-by-education group cells.

The March Supplement includes those scheduled to be interviewed in the March monthly CPS survey, as well as non-Hispanic White households with children 18 or younger and minority (Hispanic and non-Hispanic non-White) households drawn from CPS households that are in their eight month "off-period". We choose to supplement the ORG data with the March annual data because it increases our sample of households with children, especially lower income-households.

Our individual (and aggregate) employment and labor force participation data comes from the monthly ORG and the March annual data of the CPS. In addition to the labor market variables, we extract demographic information on state of residence, education attainment, marital status and number of children for CPS respondents. The March annual data spans the time period 1984-2011, while the ORG data (from IPUMS) spans 1994-2010. Thus, each observation in the ORG and March annual data corresponds to a unique individual that is in a given month and year. Approximately 40 percent of our observations are interviewed in March, with the remaining observations (from the ORG) being equally distributed across the remaining months.[17]

2. Survey of Income and Program Participation (SIPP): We use information from the 1985 to 2008 SIPP panel's to construct AFDC/TANF and food stamp take-up rates for households with various numbers of children and income levels in each local labor market. We describe this procedure in detail in the following subsection.[18]

3. Federal Reserve Economic Data (FRED): We inflate all dollar amounts to 2010 levels using the national Consumer Price Index for All Urban Consumers (CPI) from the FRED. In some specifications, we also control for the seasonally-adjusted state unemployment rate. This information is also obtained from the FRED.

4. NBER TAXSIM software: Given the year, a household's state of residence, number of children and earnings, we calculate their net tax liability using the NBER TAXSIM software.[19]

5. Welfare Benefit Calculator: We use our own calculator constructed from the Welfare Rules Database. Given the year, a household's state of residence, number of children and earnings, we approximate welfare (AFDC and TANF) and food-stamps benefits.

III.2 Data Cleaning

III.2.a CPS Data

The CPS data cleaning process is divided into the following steps:

1. Correctly assign the number of children to the mother of a household

2. Keep only non-military single women

3. Drop observations with illogical responses

1. We first pool the ORG and March annual CPS cross-sections and merge this data to the FRED CPI and unemployment data. At this stage, we have 29,916,758 person-month-year observations spanning the 1984 to 2011 period. Each observation represents a unique individual. Next, we assign the number of children

[17]From 1984 to 1993 we only have data from the March Supplement, so all observations for this period are for the month of March.

[18]We sometimes refer to the AFDC/TANF and food stamps programs as "welfare" programs.

[19]See Feenberg and Coutts (1993) for a detailed description of the TAXSIM software.

a mother is responsible for. This number is different for welfare benefit eligibility than for tax purposes. Specifically, welfare benefits vary with the number of children under the age of 18 in the household, whereas for tax purposes a child must be under the age of 19, or younger than 24 but in school. The key input in the raw CPS data for this calculation is the *momloc* variable. This variable indicates whether a respondent's mother is living in the household, as well as her "person number" if she is living in the household. For example, if there an individual's mother is not living in the household the value of the *momloc* variable would be equal to "00"; if the mother is the head of household, the value of the *momloc* variable would be "1".

To determine the number of children in the household for welfare benefit purposes, we sort the pooled CPS data by households and count the number of children under 18 living in the household. We assign this number to the head of household. Note that this number will include those that are not biological children of the household head, consistent with the way welfare benefits are typically calculated. See Appendix B (below) for more details about welfare benefit calculations. For respondents between the ages of 16 and 24, the CPS variable *schlcoll* indicates whether the respondent was in high school or college during the previous week. The CPS variable *empstat* indicates the respondent's labor force status. We assign those that report not being in the labor force because they are in school ($empstat = 33$) or who report being in college or university full time ($schlcoll = 3$) and who are between the ages of 18 and 24 as children of the head of household. We add this count to the number of minors above in order to calculate the correct number of children for tax purposes.

2. After having assigned children to female household heads, we restrict the sample to non-military single women between the ages of 18 and 55 in the ORG and March annual supplement. Specifically, dependent children (7,449,217 observations), males (10,674,890), married women (7,093,086), those who report being less than 10 years older than their youngest child (1,977), those not in the ORG or March data (2,908,023), those under the age of 18 or over the age of 55 (600,843), those in the military (924) are dropped from the sample. At this state, we have 1,187,798 person-year observations spanning the 1984 to 2011 period.

3. We also drop observations where there is evidence that the data are contaminated. The CPS variable *wkswork*1 (available in the March Supplement only) indicates the number of weeks the respondent worked for pay in the previous year. The *incwage* (also available in the March Supplement only) variable captures the respondent's reported pre-tax earnings.[20] We drop women that claim positive earnings for the previous year (i.e. $incwage > 0$) yet report not working ($wkswork1 = 0$) (9,771 observations).

4. In the final data cleaning step we exclude those who report being full-time students (149,472 observations), those with more than seven children (215), those that report having negative non-employment (other) income (1,464), those that are the only person in their state-year-month education category (562). Dropping this final group is necessary for specifications where we estimate models with state-by-year-by-month fixed effects. Finally, we exclude those with a Bachelor's degree or higher, as they are unlikely to be affected by the tax-schedule at the bottom of the income distribution (234,343 observations).

The number of children assigned to a mother is an important input into eligibility for welfare benefits and for net tax liabilities. We assess how our measure of the number of children a mother is responsible for compares with the reported value in the CPS (the *nchild* variable in the CPS) in the cleaned sample. The following table reports the difference between our calculation and the reported number of children in the CPS. A value of 1 means that we calculate a female head of household to be responsible for one more child than she claims to be her own. For example, a respondent might fail to count any non-biological children she is responsible for. A value of 0 means that our measures are identical, while a value of-1 means the female

[20]In contrast to the labor force status questions that are asked each month for all CPS (ORG and March Supplement) respondents, the *wkswork*1 and *incwage* variables are only available for the March Supplement. This information is used below to estimate annual earnings for tax and welfare purposes.

head of household claims more of her own children in the CPS than we calculate. An example of this case could occur if a respondent counts a non-school age child living at home; our calculations would exclude this child for both welfare eligibility and tax purposes.In the overwhelmingly majority of case (90.23 percent), our calculated number matches the number reported in the CPS.

$\Delta kids$	Count	Percent	Cumulative Percent
-7	4	0.00	0.00
-6	9	0.00	0.00
-5	46	0.00	0.01
-4	245	0.02	0.03
-3	1,969	0.20	0.23
-2	14,442	1.43	1.66
-1	70,288	6.97	8.63
0	909,305	90.23	98.86
1	8,028	0.80	99.66
2	2,253	0.22	99.88
3	803	0.08	99.96
4	256	0.03	99.99
5	83	0.01	100.00
6	24	0.00	100.00
7	5	0.00	100.00
Total	1,007,760	100.00	

III.2.b SIPP Data

We use information from the SIPP to calculate welfare (AFDC/TANF) and food stamp take up rates. The SIPP data cleaning process is divided into the following steps:

1. Extracting raw SIPP data
2. Ensure the data are comparable across SIPP panels
3. Calculate the number of children (under 18) in a family
4. Keep only single, non-military women age 18 to 55
5. Drop observations with illogical responses
6. Calculate welfare (AFDC/TANF/food stamps) take-up rates

1. We first pool cross sections from the 1985 to 2008 SIPP panels that span the years 1985 to 2012.[21] Respondents in each SIPP panel are interviewed every four months (a wave) for a two to four years.[22] Thus, each observation in our pooled cross-section is a person-month; the raw data include 24,401,516 such observations. We do not use the 1984 panel since it does not include individuals from Alaska, Montana, Nevada, New Hampshire, North Dakota, Utah and Vermont. Also, the 1984 panel does not differentiate between children's full time and part-time student status that is important for calculating welfare benefit eligibility.

[21] At the time we extracted the raw data the most recent wave of the 2008 SIPP panel was wave 13 that covered the September 2012 to December 2012 period. As discussed below, we only use data up to 2011 to be consistent with the CPS data. At the time of writing, the most recent wave of the 2008 SIPP panel is wave 16, which covers the September 2013 to December 2013 perod.

[22] There are 14 SIPP panels; annual, overlapping panels from 1984 to 1993, 1996, 2001, 2004 and 2008.

2. Some variable names and response values differ across SIPP waves. For example, the variable indicating the age of the respondent is called *age* in the 1990 to 1993 SIPP panels, but is called *tage* beginning in the 1996 panel. Also, total family unemployment income is called $funemp$ in the 1990 to 1993 SIPP panels; the variable name changes to $tfunemp$ beginning in 1996. Thus, the next step in the data cleaning process ensures that the data are comparable across SIPP panels. We use the code and crosswalk from the Centre for Economic Policy Research (CEPR) website that makes the 1990 to 2008 SIPP panels comparable.[23] We borrow from this code for earlier panels to ensure the comparability.

3. We calculate the number of children in a family as follows. We use information in the SIPP to designate women as family heads. Family heads can be living in the same household as their parents. In these cases, the woman would be designated as a sub-family head if she also has a dependent child. We classify all female family or sub-family heads as heads of household. Each person-month observation in the SIPP has common "family-level (or sub-family level)" variables, such as the number of children in the family/sub-family. We use this common family-level variable to calculate the number of children (that are under the age of 18, reside in the same household, and are related through birth or adoption) a female family or sub-family head is responsible for. [24]

4. Next, we restrict the sample to single non-military women between the ages of 18 and 55, as with the CPS data. First, we drop observations from the 2012 calendar year (116,624 observations). We drop males (11,640,919), those under 18 or over 55 (6,062,223), married women (3,959,793), those that are not heads of household (825,927), full-time students (120,822), those in the military (2,570), as well as a small number of those with more than seven children due to a lack of program data on these households (467).

5. As with the CPS data, we drop observations where there is evidence that the data are contaminated. We drop women who claim positive earnings for the previous year yet report not working. We also drop those that report working the previous year but have zero earnings (86,892 observations). The resulting sample size is 1,585,279.

6. We calculate AFDC/TANF and food stamps recipiency rates based on cells defined by an individual's year of observation, education group, and number of children. We calculate recipiency rates for each of these programs separately as follows. Using the cleaned SIPP data, we define our cells as follows. The four education groups are: less than a high school diploma (or equivalent), high school diploma, some college (or an associate's degree), and a college degree. The number of children groups are $\{0,1,2,3+\}$. The year of observation groups are $\{1984-1988, 1989-1993, 1994-1998, 1999-2003, 2004-2008, 2009-2011\}$. The interaction of these groups leads to 96 cells. Thus, each observation in the SIPP will we an element of one of these cells. We calculate the fraction of individuals receiving AFDC/TANF and food stamps by calculating the fraction of women in each cell that report receiving benefit income.[25] Since women with no children are ineligible for AFDC/TANF benefits, the recipiency rate is zero in one quarter of the cells. In the empirical section we collapse the recipiency rates for the pre- and post-1996 years (after major welfare reform) for each education group. This leads to eight recipiency rates, one for each education group before 1996, and one for each education group after 1996.

III.3 Dependent Variables

Our dependent variables of interest are (a) the micro labor force participation rate; (b) the macro participation rate; and (c) the macro employment rate. We use information on the reported labor force and employment status from ORG and March CPS respondents to construct these three variables. The $empstat$ variable (available for both the ORG and March Supplement) in the CPS indicates a respondent's employ-

[23] http://ceprdata.org/sipp-uniform-data-extracts/

[24] Since we only use the SIPP for welfare take-up rates we don't need to worry about children over 18 that are still dependents.

[25] The person-month probability weights in the SIPP are used to calculate these averages.

ment status for the previous week.[26] The possible values for this variable are (i) "Not in labor force", (ii) "Unemployed", and (iii) "Employed".[27] For some years additional detail on a respondent's labor force status is available, but we do not use it in this paper. For example, information on whether those out of the labor force are unable to work is available for most years in the time period we study. In other years, reasons for being out of the labor force due to being in school full time is also available.

From the *empstat* variable we define an indicator variable equal to one if a CPS respondent is in the labor force and zero otherwise. Specifically, those that are coded as being "Unemployed" or "Employed" are in the labor force. Our macro measure of labor force participation aggregates this variable to the state, year and education group level (our definition of a local labor market). Similarly, we define an employment status indicator equal to one if a CPS respondent reports being "Employed" and zero otherwise; the employment/population rate. The macro employment status variable aggregates the employment status dummy variable to the state, year and education group level.

III.4 Tax and Benefit Variables

Our independent variables of interest are the net tax liability, after-tax income and welfare benefits of respondents. We assign each person in our CPS sample, the net tax liability and benefit amount corresponding to their state, year, education group, number of children and imputed earnings level. The first step is to impute earnings.

III.4.a Preliminaries: Imputed Earnings

We impute earnings as follows. Let w_i be the reported earnings by individuals in the March Supplement of the CPS, which indicates each respondent's pre-tax wage and salary income for the previous calendar year. For those with positive earnings, we take the natural logarithm of this variable $log(w_i)$. Next, for each year and education group (high school dropouts, high school graduates, and some college), we estimate the following model separately by education group e and year t:

$$log(w_i) = X_i\beta_{e,t} + \varepsilon_i,$$

where X_i are a set of demographic variables: a linear and quadratic term in age, dummies for race (hispanic and black) and urban/rural status and state fixed effects. The predicted values from these regressions (for each year and education group) are converted back into levels and assigned to all CPS respondents, regardless of their work status:

$$\hat{w}_i = exp(\widehat{log(w_i)} = exp(X_i\hat{\beta}_{e,t})$$

This amount is inflated (or deflated) to 2010 dollars.

III.4.b Calculating Tax and Welfare Benefit Variables

Given imputed earnings, as well as a the TANF/AFDC and food stamps take-up rates, calculate the net tax liability and welfare benefits. We use the Urban Institute's Welfare Rules Database[28] and TRIM3[29]

[26] The monthly CPS interviews (including those for the March Supplement) occur during the week of the 19^{th} of the month. The baseline labor force status questions for each month (and therefore apply to the ORG and March samples) ask respondents about whether they were working, working but temporarily absent, searching for a job or not working and not searching for a job during the previous week, referred to as the "reference week" (i.e. the week of the 12^{th} of the month).

[27] An individual is employed if he or she reports working or temporarily absent from a job during the CPS reference week. An individual is unemployed if they report not being employed but actively searching for a job during the reference week.

[28] http://anfdata.urban.org/wrd/WRDWelcome.cfm

[29] http://trim3.urban.org/

program rules to create an AFDC/TANF benefit calculator. For tax credits and liabilities we use the NBER's TAXSIM9 software[30].

Micro Tax and Benefit Variables: The micro tax and benefit variables are calculated as follows:

1. We group individual imputed earnings $\hat{w}_i$ into a grid with 200 dollar bins: $\{200, 400, 600,..., 120000\}$. We call this the binned imputed earnings: $\tilde{w}_i$. The reason for doing this is simply to ease the computational burden for calculating taxes and transfers.[31]

2. Let $G_\tau(w,s,t,n)$ be the tax policy function for tax/transfer program τ, for $\tau \in \{$federal taxes, state taxes, payroll taxes, AFDC, TANF, food stamps$\}$, that maps earnings into tax liability depending on state, year and number of (dependent) children. We calculate G_τ separately for federal, state or payroll tax liabilities, as well as AFDC, TANF and food stamp benefit levels using our welfare calculator and TAXSIM9. Rather than using actual earnings we compute tax liability using the imputed binned earnings: $G_\tau(\tilde{w}_i, s_i, t_i, n_i)$

3. Let the take-up rate for tax/transfer τ be ρ_τ, which will be a function of education, number of children and period: $\rho_\tau(e,n,t)$

4. Let's define the Micro tax variable for individual i: $T_\tau^{micro}(i) = \rho_\tau(e_i,n_i,t_i)G_\tau(\tilde{w}_i,s_i,t_i,n_i)$, where $\rho_\tau = 0$ for $\tau \in \{$federal taxes, state taxes, payroll taxes$\}$.

5. After-tax income conditional on working, c_i, for each individual in the CPS is calculated as follows:

$$\begin{aligned} c_i \;=\; & \tilde{w}_i \\ & -G_{Federal}(\tilde{w}_i,s_i,t_i,n_i) \\ & -G_{state}(\tilde{w}_i,s_i,t_i,n_i) \\ & -G_{Fica}(\tilde{w}_i,s_i,t_i,n_i) \\ & +\rho_{TANF/AFDC}(e,n,t)G_{TANF/AFDC}(\tilde{w}_i,s_i,t_i,n_i) \\ & +\rho_{FoodStamps}(e,n,t)G_{FoodStamps}(\tilde{w}_i,s_i,t_i,n_i) \end{aligned}$$

   where $G_{TANF/AFDC}(\tilde{w},s,t,n)$ and $G_{FoodStamps}(\tilde{w},s,t,n)$ are the annual levels of benefits for women with n children, binned predicted income $\tilde{w}$, living in state s, in year t, multiplied by the welfare take-up rate for groups defined by year, education and number of children. This accounts for the fact that the take up of these programs is less than 100 percent.

Macro Tax and Benefit Variables: The macro tax and benefit variables are calculated as follows.

1. First we collapse the individual tax variables $T_\tau^{micro}(i) = \rho_\tau(e_i,n_i,t_i)G_\tau(\tilde{w}_i,s_i,t_i,n_i)$ to the state X year X NumChildren X education level. Call this collapsed tax liability $T_\tau^{collapsed}(e,n,s,t)$.

2. Let $N_{e,n}$ be the number of individuals with education e and n children, let N_e be the number of individuals with education e and define $\alpha_{e,n} = \frac{N_{e,n}}{N_e}$ to be the share of women with n children in education group e over the entire sample period and all states.

3. The Macro tax variable is constructed by integrating over the collapsed micro tax variables but using a constant distribution of children across all cells:

$$T_\tau^{macro}(e,s,t) = \sum_n \alpha_{e,n} T_\tau^{collapsed}(e,n,s,t)$$

[30] http://users.nber.org/ taxsim/taxsim9/

[31] Those with predicted earnings greater than \$120,000 are topcoded at \$120,000.

III.4.c Instruments

Welfare benefits and tax liabilities, including tax credits such as the EITC, are endogenous to a taxpayer's earnings. We deal with this endogeneity using a simulated instrumental variables strategy. Our strategy exploits changes in tax and benefit rules across states over time between those with different numbers of children. Identification relies on holding fixed the distribution of income, which may be endogenous to tax policy. Our instruments are calculated as follows:

1. Calculate empirical CDF of real earnings w_i^r for each year and education group $F_{e,t}(\omega)$.

   We approximate the empirical CDF using centiles:

   First, we inflate the imputed income variable $w_{m,e,s,t,n}$ (see above) to 2010 dollars using the CPI. Using these imputed real incomes for all individuals from 1984 to 2011, we construct the percentiles of the empirical earnings distribution. We record the income cutoffs for the lower and upper bounds of each centile.

   To get the CDF by education and year, we compute the percentage of individuals in each centile. **taxes_simInst.ado** 78-127

   For each year we compute the mean nominal earnings in each centile, conditional on real earnings in that year being within the bounds of the centile from step 1. **taxes_simInst.ado** 144-166.

2. We then calculate the micro instruments by using our policy functions and the empirical CDF using the centiles:
$$T_\tau^{micro,Instrument}(t,s,e,n) = \int \rho_\tau(e_i,n_i,t_i)G_\tau(\tilde{\omega}_i,s_i,t_i,n_i)dF_{e,t}(\omega)$$

3. We then collapse the micro instrument: $T_\tau^{micro,Instrument}(t,s,e,n)$ to the state X year X NumChildren X education level. Call this: $T_\tau^{collapsed,Instrument}(e,n,s,t)$:
$$T_\tau^{collapsed,Instrument}(e,n,s,t)$$

4. The Macro Instrument is then calculated by aggregating across number of children, so that it only varies on the education X state X year level:
$$T_\tau^{macro}(e,s,t) = \sum_n \alpha_{e,n} T_\tau^{collapsed,Instrument}(e,n,s,t)$$

III.5 Variable List

For convenience, this subsection provides a list of all variables used in the empirical analysis. Since we use information from several sources, we record which dataset each variable originated from. Definitions for each variable are also included.

CPS Variables:

- *age*: age of CPS respondent
- *sex*: gender of CPS respondent (1 for males and 2 for females)
- *hisp*, *nonwhite*, *black*: race dummy variables from the CPS
- *marst*: marital status of CPS respondent (7 categories); singles are either divorced, widowed or never married

- *momloc*: indicates whether a CPS respondent's mother lives in the household. A value of 00 indicates that the mother is not in the household. Otherwise, the CPS person number of the respondent is coded. For example, if a CPS respondent's mother is the head of household, her person number would be 1.

- $statefip$: state of residence of CPS respondent

- *schlcoll*: Indicates whether CPS respondent's between the ages of 16 and 24 are in school. The acceptable responses are (CPS coded values in parenthesis): NIU (0), high school full time (1), high school part time (2), college or university full time (3), college or university part time (4), does not attend school, college or university (5)

- *educ*: a respondent's education attainment. The categories are (along with their coded values in the CPS in parenthesis):
  - NIU or no schooling: separate categories for no information available (001) or preschool/kindergarten (002), as well as a summary category (000)
  - Grades 1-4 inclusive: separate categories for each of grades 1 to 4 (011 to 014), along with a summary grades 1 to 4 category (010)
  - Grades 5 or 6: separate categories for grades 5 and 6 (021 to 022), along with a summary grades 5 to 6 category (020)
  - Grades 7 or 8: separate categories for grades 7 and 8 (031 to 032), along with a summary grades 7 to 8 category (030)
  - Grade 9: CPS respondent completed grade 9 (040)
  - Grades 10: CPS respondent completed grade 10 (050)
  - Grade 11: CPS respondent completed grade 11 (060)
  - Grade 12: separate categories for 12th grade completed with no diploma (071), 12th grade completed by diploma status unknown (072), 12th grade completed with a high school diploma or equivalent (073), as well as a summary variable for any one of these three categories (070)
  - 1 year of college: CPS respondent completed one year of college and did not earn a degree (080 to 081)
  - 2 years of college: separate categories for Associate's degree, occupational or vocational program (091), Associate's degree, academic program (092), as well as a summary variable for each of these two categories (090)
  - 3 years of college: CPS respondent completed three years of college (no bachelor degree) (100)
  - 4 years of college: CPS respondent completed four years of college and earned a bachelor's degree (110 to 111)
  - 5+ years of college: separate categories for 5 years of college (121), 6 years of college (122), completed a Master's degree (123), completed a professional school degree (124), completed a doctorate (125), as well as a summary variable for any one of these categories (120)

- *hsDrop*: dummy variable equal to 1 if a CPS respondent has less than a high school diploma (value of $educ < 72$); 0 otherwise (constructed variable)

- *hsGrad*: dummy variable equal to 1 if a CPS respondent has a high school diploma (value of $educ \geq 72$ and $educ \leq 73$); 0 otherwise (constructed variable)

- *college*: dummy variable equal to 1 if a CPS respondent has an associate's degree, vocational certificate or attended some college but did not complete a certificate or degree program (value of $educ > 73$ and $educ < 110$); 0 otherwise (constructed variable)
- *bachelor*: dummy variable equal to 1 if a CPS respondent has a bachelor's degree or higher (value of $educ \geq 110$); 0 otherwise (constructed variable)
- $wkswork1$: number of weeks a CPS respondent worked during the past calendar year
- *yearWork*: dummy variable equal to 1 if $wkswork1 > 0$; 0 otherwise (constructed variable)
- *incwage*: reported pre-tax wage and salary income
- *hrswork*: reported number of hours worked during the previous week
- *weekWork*: dummy variable equal to 1 if CPS respondent worked a positive number of hours during the previous week; 0 otherwise (constructed variable)
- *uhrswork*: number of hours a CPS respondent normally works during the week
- *hoursWork*: estimated number of hours worked last year; equal to $wkswork1 * uhrswork$ (constructed variable)
- *empstat*: a CPS respondent's employment status. The categories are (along with their coded values in the CPS in parenthesis):
  - NIU (00)
  - CPS respondent in the armed forces
  - CPS respondent's labor force status, conditional on being in the labor force: separate categories for employed at at work (10), employed but was temporarily not at work during the reference week (12), unemployed and an experienced worker (21), unemployed and a new worker (22) and a summary unemployed variable (20)
  - CPS respondent's status (not in the labor force): separate categories for does housework (31), unable to work (32), in school full time (33), other (34), does unpaid work (35)
- lfp_ind: Labor force participation status dummy variable; equal to one if respondent is in the labor force ($empstat \geq 10$ and $empstat \leq 22$); zero otherwise (constructed variable)
- emp_ind: Employment status dummy variable; equal to one if respondent is employed ($empstat \geq 10$ and $empstat \leq 12$); zero otherwise (constructed variable)

D Description of Welfare Program Rules and Calculation of Benefits

In this Appendix, we provide a brief description of the transfer programs that low-income families are eligible for. In particular, we summarize the following programs: Aid to Families with Dependent Children (TANF), Temporary Assistance to Needy Families (TANF), and the Supplemental Nutrition Assistance Program (SNAP). The SNAP program is often referred to as "food stamps". For simplicity, we refer to these programs collectively as "welfare". After describing these programs, we describe how we calculate individual welfare benefits using the rules published in the Welfare Rules Database[32] and TRIM3[33], managed by the Urban Institute.

[32] http://anfdata.urban.org/wrd/WRDWelcome.CFM
[33] http://trim3.urban.org

IV.1 Description of Welfare Program Rules

IV.1.a Aid to Families with Dependent Children (AFDC)

The AFDC program was introduced in 1936 to provide financial assistance to children from low-income families. The program was replaced in 1997 by the TANF program following the passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA), which we describe below. AFDC benefits were administered by the federal government, through the Department of Health and Human Services, although states shared in the program's costs and rule-making authority. In particular, states were able to determine individual eligibility and benefit levels, subject to federal guidelines and program requirements.

Families with children under the age of 18 that are residents of the state and whose children are living with them were eligible for AFDC benefits if they met the state's standard of need. A family was considered needy, if their monthly income was below a specified level; some types of income, such as child support payments, the EITC, and allowances for child care expenses, were disregarded for the purposes of determining eligibility.[34] As income increased above the disregard, a family's AFDC benefit was reduced until they were no longer eligible for benefits. Families that were eligible for the AFDC were automatically eligible for other entitlements, such as Medicaid and food stamps.

IV.1.b Temporary Assistance to Needy Families (TANF)

Two criticisms of the AFDC program was that the high claw-back rates on benefits and no duration limit on benefits provided a disincentive to work. These criticisms, among others, led to the replacement of the AFDC by the TANF program in 1997 as part of the PRWORA. In general, the primary difference between the AFDC and TANF programs is that the latter provides states with much more flexibility in choosing eligibility requirements, benefit levels, work requirements and phase-out rates. Under TANF, states are provided with block grants to finance their own programs, provided that they help achieve four goals set forth in the PRWORA.[35] The four goals are: (i) provide assistance to children from needy families, (ii) end the dependence of needy parents on government benefits by promoting job preparation, work and marriage, (iii) reduce out-of-marriage pregnancies, and (iv) encourage the formation and maintenance of two-parent families. States must ensure that TANF benefit recipients meet work requirements to remain eligible for benefits, with some exceptions.[36] The work requirements are that recipients: (a) must work as soon as they are job ready and no later than two years after initially receiving benefits and (b) work a minimum number of hours per week. Federal TANF rules also impose time limits on the receipt of (cash) benefits. Income (and asset) cutoffs for TANF eligibility varies significantly across states.

IV.1.c Supplemental Nutrition Assistance Program (SNAP or food stamps)

The Supplemental Nutrition Assistance Program (SNAP or food stamps) provides assistance to low- and moderate-income families to purchase food items. Rules for the food stamp program are determined by the federal government and is funded through United States Department of Agriculture. The program is administered by states that have some discretion in setting household income reporting requirements and choosing what the program is called in their state. SNAP benefits are delivered each month to households via a magnetically encoded payment card, known as an Electronic Benefits Transfer (EBT) card. After applying and getting approved for benefits, recipients receive their EBT card. States credit EBT cards for

[34] A household's eligibility also depended on meeting asset tests set by the federal and state governments.

[35] The basic (nominal dollar) block grant for each state was set in 1996. States with faster population growth are eligible for larger block grants, and states can be eligible for more funding to deal with increased case loads during recessions.

[36] The activities that fulfill the work requirement varies by state.

eligible households monthly. This card, similar to a debit card or a bank card, is accepted to purchase food items.

Eligibility for food stamps is primarily determined by a household's monthly income. The income test is increasing in family size. For households with one individual in 2015, the monthly income cutoff is $1,265. The monthly income cutoff for households with two, three and four members is $1,705, $2,144 and $2,584 respectively. A household's monthly allotment is calculated as $FS = (MaxBen - 0.3 * [(1 - EIDed) * EI + OtherInc - StDed - Shelt])$ where $MaxBen$ is the maximum allotment determined annually and dependant on the household size, $EIDed$ is the earned income deduction, $OtherInc$ is unearned income, which includes AFDC or TANF benefits, $StDed$ is a standard deduction and $Shelt$ is a shelter expense deduction[37].

IV.2 Calculating Individual Welfare Benefits

We calculate expected annual AFDC, TANF and SNAP benefits for each woman in our CPS sample using two databases of ruels. For every state and for each year from 1996 to 2013, the Welfare Rules Database contains detailed information on benefit levels (by household size), eligibility requirements, income disregards, work requirements and other details. For years prior to 1996 we use the AFDC rules from the Urban Institute's TRIM3 program structured similarly to the Welfare Rules Database. We assume that households have not exhausted their welfare eligibility throughout the analysis. We model the initial parameters of the welfare programs, some of the income disregards expire or change after extended periods of sustained earnings. We use this information to construct separate welfare calculators for AFDC/TANF and food stamps. For each year and state, this calculator takes income, state, year and number of children and uses state disregards, claw-back rates and income tests to compute a household's monthly level of benefits. We multiply the level of monthly benefits by twelve as our measure of annual benefits for the OLS regressions.

Figure A-1 provides some example budget sets that our welfare / tax calculator generates. The figures show the different components that create the difference between pre- and post-tax income: food stamps, TANF/AFDC, state taxes and federal taxes. Both panels show the budget set of a single individual with 2 dependent children. As can be seen in the two examples (California and New York), food stamps have a structure like a negative income tax but with a cliff at the end, leading to a notch in the tax schedule. TANF pays a large amount at zero income and is then phased out though at different rates in different states (much slower in California for example). State taxes are essentially absent in California in the relevant range, but the federal EITC creates a sizeable bump in the 8 to 15 000 income range. In New York, state taxes create a small positive transfer at low incomes due to a state EITC, but have a negative effect above 30 000. The two figures highlight that there is substantial heterogeneity in these programs across states.

Figure A-2 shows the variation in the overall budget sets across number of children, time and states. Panels (a), (b) and (c) show how the budget sets by number of children have evolved in Ohio from 1984 to 2000, highlighting how the transfers have become more EITC-like with lower phase-out rates and somewhat smaller transfers at the bottom. Panels (c) to (f) show different states in the year 2000, revealing substantial heterogeneity in the shape and structure of these schedules. For example California's transfer schedule implies tax rate close to zero at low incomes up to around 10,000 but then the tax rate due to phase out of various programs is close to 100 percent between 10,000 and 30,00 for a single parent with two children. Compared to this Texas provides much higher work incentives (and much lower transfers at zero income). Overall these figures highlight the type of variation that identifies our micro responses (within labor market differential changes in taxes across children) and macro responses (across state and year changes on the labor market level).

[37] There is also an asset test of $2,250 in financial resources. Recipients between the ages of 18 and 50 without dependent children also face work requirements. In particular, they are only eligible to receive SNAP benefits for three months in a 36 month period if they do not participate in a workfare or employment training program.

References

Akerlof, George A and Janet L Yellen, "The Fair Wage-Effort Hypothesis and Unemployment," *The Quarterly Journal of Economics*, 1990, *105* (2), 255–83.

Barro, Robert J and Herschel I Grossman, "A General Disequilibrium Model of Income and Employment," *American Economic Review*, 1971, *61* (1), 82–93.

Brunello, Giorgio and Daniela Sonedda, "Progressive taxation and wage setting when unions strategically interact," *Oxford Economic Papers*, 2007, *59* (1), 127–140.

Chetty, Raj, "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 2008, *116* (2), 173–234.

Diamond, Peter, "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies*, 1982, *49* (2), 217–27.

_ and James Mirrlees, "Optimal Taxation and Public Production," *American Economic Review*, 1971, *61*, 8–27 also 261–278.

Golosov, Mikhail, Pricila Maziero, and Guido Menzio, "Taxation and Redistribution of Residual Income Inequality," *Journal of Political Economy*, 2013, *121* (6), 1160 – 1204.

Hansen, Claus Thustrup, Lars Haagen Pedersen, and Torsten Sløk, "Ambiguous effects of tax progressivity – theory and Danish evidence," *Labour Economics*, 2000, *7* (3), 335–347.

Hersoug, Tor, "Union Wage Responses to Tax Changes," *Oxford Economic Papers*, 1984, *36* (1), 37–51.

Holmlund, B and A. S. Kolm, "Progressive taxation, wage setting and unemployment: theory and Swedish evidence," *Swedish Economic Policy Review*, 1995, *2*, 423–460.

Hosios, Arthur J, "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 1990, *57* (2), 279–98.

Hungerbühler, Mathias, Etienne Lehmann, Alexis Parmentier, and Bruno Van der Linden, "Optimal Redistributive Taxation in a Search Equilibrium Model," *Review of Economic Studies*, 2006, *73* (3), 743–767.

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden, "Optimal income taxation with Kalai wage bargaining and endogenous participation," *Social Choice and Welfare*, 2014, *42* (2), 381–402.

Keynes, John Maynard, *The General Theory of Employment, Interest and Money*, New York: Macmillan, 1936.

Landais, Camille, Pascal Michaillat, and Emmanuel Saez, "A Macroeconomic Approach to Optimal Unemployment Insurance: Applications," *American Economic Journal: Economic Policy*, 2018, *10* (2), 182–216.

_ , _ , and _ , "A Macroeconomic Approach to Optimal Unemployment Insurance: Theory," *American Economic Journal: Economic Policy*, 2018, *10* (2), 152–181.

Lee, David and Emmanuel Saez, "Optimal Minimum Wage Policy in Competitive Labor Markets," *NBER Working Papers 14 320*, 2008.

__ **and** __ , "Optimal minimum wage policy in competitive labor markets," *Journal of Public Economics*, 2012, *96* (9 - 10), 739 – 749.

Lehmann, Etienne, Alexis Parmentier, and Bruno Van der Linden, "Optimal income taxation with endogenous participation and search unemployment," *Journal of Public Economics*, 2011, *95* (11), 1523–1537.

__ **, Claudio Lucifora, Simone Moriconi, and Bruno Van der Linden**, "Beyond the Labour Income Tax Wedge: The Unemployment-Reducing Effect of Tax Progressivity," *International Tax and Public Finance*, 2016, *23* (3), 454–489.

Lockwood, Ben and Alan Manning, "Wage setting and the tax system theory and evidence for the United Kingdom," *Journal of Public Economics*, 1993, *52* (1), 1–29.

Malcomson, James M. and Nicola Sartor, "Tax push inflation in a unionized labour market," *European Economic Review*, 1987, *31* (8), 1581 – 1596.

Manning, Alan, "Wage Bargaining and the Phillips Curve: The Identification and Specification of Aggregate Wage Equations," *Economic Journal*, 1993, *103* (416), 98–118.

McDonald, Ian M and Robert M Solow, "Wage Bargaining and Employment," *American Economic Review*, 1981, *71* (5), 896–908.

Michaillat, P., "Do matching frictions explain unemployment? Not in bad times," *The American Economic Review*, 2012, *102* (4), 1721–1750.

Michaillat, Pascal and Emmanuel Saez, "Aggregate Demand, Idle Time, and Unemployment," *The Quarterly Journal of Economics*, 2015, *130* (2), 507–569.

Moen, Espen R, "Competitive Search Equilibrium," *Journal of Political Economy*, 1997, *105* (2), 385–411.

Mortensen, Dale T. and Christopher A. Pissarides, "New developments in models of search in the labor market," 1999, *3*, 2567–2627.

Naito, H, "Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency," *Journal of Public Economics*, 1999, *71*, 165–188.

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, 2014, *6* (1), 230–71.

Pisauro, Giuseppe, "The effect of taxes on labour in efficiency wage models," *Journal of Public Economics*, 1991, *46* (3), 329–345.

Pissarides, Christopher A., "Taxes, Subsidies and Equilibrium Unemployment," *Review of Economic Studies*, 1985, *52* (1), 121–134.

__ , "The impact of employment tax cuts on unemployment and wages; The role of unemployment benefits and tax structure," *European Economic Review*, 1998, *42* (1), 155–183.

__ , *Equilibrium Unemployment Theory, 2nd Edition*, Vol. 1 of MIT Press Books, The MIT Press, 2000.

__ **and Barbara Petrongolo**, "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 2001, *39* (2), 390–431.

Rothschild, Casey and Florian Scheuer, “Redistributive Taxation in the Roy Model,” *The Quarterly Journal of Economics*, 2013, *128* (2), 623–668.

Saez, Emmanuel, “Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses,” *The Quarterly Journal of Economics*, 2002, *117* (3), pp. 1039–1073.

_ , “Direct or Indirect Tax Instruments for Redistribution: short-run versus long-run,” *Journal of Public Economics*, 2004, *88*, 503–518.

Shapiro, Carl and Joseph E Stiglitz, “Equilibrium Unemployment as a Worker Discipline Device,” *American Economic Review*, 1984, *74* (3), 433–44.

Stiglitz, Joseph E., “Self-selection and Pareto efficient taxation,” *Journal of Public Economics*, 1982, *17* (2), 213–240.

E Appendix Tables

Table A-1: Recipiency Rates of Transfer Programs

Period	(1) 1984-1996	(2) 1997-2011
Panel A: Food Stamps		
HS Dropout	0.414	0.406
HS Graduate	0.187	0.225
Some College	0.101	0.146
College Graduate	0.012	0.022
Panel B: AFDC/TANF		
HS Dropout	0.489	0.209
HS Graduate	0.230	0.100
Some College	0.170	0.062
College Graduate	0.030	0.011

Notes: Recipiency rates are calculated using the Survey of Income and Program Participation. These data reflect the recipiency rates of single women aged 18-55 who are not full time students or in the military, consistent with the data used for the empirical analysis from the CPS. An individual is counted as a recipient of either food stamps or AFDC/TANF if they received a transfer in any amount from the program. The recipiency rates for food stamps include single women without children. The recipiency rates for AFDC/TANF include only single mothers (single women without children are not eligible for the benefit).

Table A-2: OLS Regressions

LHS Variable	(1) Participation	(2) Employment
Panel A: Micro Response	$\frac{\partial \mathscr{K}_i}{\partial T_i}^{micro}$	$\frac{\partial \mathscr{H}_i}{\partial T_i}^{micro}$
Taxes Plus Benefits $(T_i + b)$	-0.006	-0.006
	[0.001]***	[0.001]***
Num. Obs	1816065	1816065
Panel B: Macro Response	$\frac{\partial \mathscr{K}_i}{\partial T_i}$	$\frac{\partial \mathscr{H}_i}{\partial T_i}$
Avg Taxes Plus Benefits within Labor Market	0.007	0.009
	[0.001]***	[0.001]***
Num. Obs	8568	8568

Table A-3: Reduced Form Regressions

LHS Variable	(1) Participation	(2) Employment
Panel A: Micro Response	$\frac{\partial \mathscr{K}_i}{\partial T_i}^{micro}$	$\frac{\partial \mathscr{H}_i}{\partial T_i}^{micro}$
Taxes Plus Benefit with takeup: sim	-0.053	-0.051
	[0.003]***	[0.003]***
Num. Obs	1816065	1816065
Panel B: Macro Response	$\frac{\partial \mathscr{K}_i}{\partial T_i}$	$\frac{\partial \mathscr{H}_i}{\partial T_i}$
Avg Taxes Plus Benefit with takeup: sim	-0.031	-0.026
	[0.015]*	[0.016]
Num. Obs	8568	8568

Table A-4: Simulating the Optimal Tax and Transfer Schedule - Low Redistributive Preferences $\nu = 0.25$

	η	μ	$\frac{\frac{d\mathscr{H}}{dT}}{\frac{d\mathscr{H}}{dT}\Big\vert^{Micro}}$	Demogrant b	Emp. Tax $T_1 + b$	ATR (6k-0)	Avg. MTR	Break-even $T(w) = 0$	Solution
Panel A: Alternative η - Macro/Micro ratio = 1, $\mu = 0$									
	0.01	0.00	1.00	7100	500	79.28	0.47	9600	Yes
Macro Est.	0.42	0.00	1.00	4600	-200	29.02	0.40	12000	Yes
Micro Est.	0.60	0.00	1.00	4000	-300	18.90	0.38	12600	Yes
	1.00	0.00	1.00	3100	-300	7.44	0.35	13600	Yes
Panel B: Comparing different Macro-Micro Participation Ratios									
	0.42	0.00	0.75	8800	300	66.35	0.57	12900	Yes
Benchmark	0.42	0.00	0.90	6600	100	48.32	0.47	12400	Yes
	0.42	0.00	1.00	4600	-200	29.02	0.40	12000	Yes
	0.42	0.00	1.25	500	-2600	-22.01	0.32	11300	Yes
Panel C: Comparing different values for μ									
	0.42	0.00	1.00	4600	-200	29.02	0.40	12000	Yes
	0.42	0.10	1.00	5500	-300	22.60	0.45	14400	Yes
	0.42	0.20	1.00	6600	-200	15.73	0.52	17100	Yes
	0.42	0.30	1.00	7900	-200	9.12	0.61	20200	Yes
Panel D: Optimal Tax Schedule over Business Cycle - with Wage and Unemployment Responses									
Recession	0.34	0.00	0.73	9200	300	72.44	0.59	12700	Yes
Normal	0.42	0.00	0.90	6600	100	48.32	0.47	12400	Yes
Boom	0.48	0.00	1.06	3100	-700	10.38	0.36	11900	Yes
Panel E: Optimal Tax Schedule over Business Cycle - without Wage and Unemployment Responses									
Recession	0.34	0.00	1.00	4900	-100	34.76	0.41	11700	Yes
Normal	0.42	0.00	1.00	4600	-200	29.02	0.40	12000	Yes
Boom	0.48	0.00	1.00	4400	-200	25.35	0.40	12200	Yes

Notes: The table shows simulations of the optimal tax schedule based on equation 15 in the text. All simulations are based on the single worker earnings distribution in the March 2011 CPS and assume the parameter values $\nu = 0.25$, $\varepsilon = 0.25$ and - for wages above \$20,000 - $\eta = 0$. In Pandel D, both η and the Macro/micro ratio vary between boom and recession, while in Panel E the Macro/micro ratio is held constant at 1.

Table A-5: Simulating the Optimal Tax and Transfer Schedule - Medium High Redistributive Preferences $\nu = 1.00$

	η	μ	$\frac{\frac{d\mathscr{K}}{dT}}{\frac{d\mathscr{K}}{dT}\vert^{Micro}}$	Demogrant b	Emp. Tax $T_1 + b$	ATR (6k-0)	Avg. MTR	Break-even $T(w) = 0$	Solution
Panel A: Alternative η - Macro/Micro ratio = 1, $\mu = 0$									
	0.01	0.00	1.00	11100	500	89.14	0.64	13500	Yes
Macro Est.	0.42	0.00	1.00	9300	-200	53.58	0.61	14600	Yes
Micro Est.	0.60	0.00	1.00	8900	-300	43.44	0.60	15000	Yes
	1.00	0.00	1.00	8200	-400	27.32	0.58	15800	Yes
Panel B: Comparing different Macro-Micro Participation Ratios									
	0.42	0.00	0.75	10700	300	72.78	0.67	14100	Yes
Benchmark	0.42	0.00	0.90	10000	100	62.84	0.64	14300	Yes
	0.42	0.00	1.00	9300	-200	53.58	0.61	14600	Yes
	0.42	0.00	1.25	7100	-2100	15.23	0.57	15500	Yes
Panel C: Comparing different values for μ									
	0.42	0.00	1.00	9300	-200	53.58	0.61	14600	Yes
	0.42	0.10	1.00	10300	-300	46.19	0.67	16200	Yes
	0.42	0.20	1.00	11300	-400	36.71	0.73	18000	Yes
	0.42	0.30	1.00	10400	-2000	-2.50	0.76	20700	No
Panel D: Optimal Tax Schedule over Business Cycle - with Wage and Unemployment Responses									
Recession	0.34	0.00	0.73	11000	300	77.37	0.68	14000	Yes
Normal	0.42	0.00	0.90	10000	100	62.84	0.64	14300	Yes
Boom	0.48	0.00	1.06	8700	-500	42.62	0.60	14900	Yes
Panel E: Optimal Tax Schedule over Business Cycle - without Wage and Unemployment Responses									
Recession	0.34	0.00	1.00	9600	-0	58.73	0.62	14400	Yes
Normal	0.42	0.00	1.00	9300	-200	53.58	0.61	14600	Yes
Boom	0.48	0.00	1.00	9200	-200	50.07	0.61	14700	Yes

Notes: The table shows simulations of the optimal tax schedule based on equation 15 in the text. All simulations are based on the single worker earnings distribution in the March 2011 CPS and assume the parameter values $\nu = 1.00$, $\varepsilon = 0.25$ and - for wages above \$20,000 - $\eta = 0$. In Pandel D, both η and the Macro/micro ratio vary between boom and recession, while in Panel E the Macro/micro ratio is held constant at 1. Note that for high values of μ (around 0.3 and higher), the optimizer does not converge to a solution for large values of ν (see last column).

Table A-6: Simulating the Optimal Tax and Transfer Schedule - High Redistributive Preferences $\nu = 4.00$

	η	μ	$\frac{\frac{d\mathscr{K}}{dT}}{\frac{d\mathscr{K}}{dT}\big\vert^{Micro}}$	Demogrant b	Emp. Tax $T_1 + b$	ATR (6k-0)	Avg. MTR	Break-even $T(w) = 0$	Solution
Panel A: Alternative η - Macro/Micro ratio = 1, $\mu = 0$									
	0.01	0.00	1.00	13500	500	94.34	0.75	15100	Yes
Macro Est.	0.42	0.00	1.00	12100	100	76.53	0.74	14700	Yes
Micro Est.	0.60	0.00	1.00	11800	-0	71.46	0.74	14600	Yes
	1.00	0.00	1.00	11400	-100	64.29	0.74	14500	Yes
Panel B: Comparing different Macro-Micro Participation Ratios									
	0.42	0.00	0.75	12300	300	81.87	0.75	14500	Yes
Benchmark	0.42	0.00	0.90	12200	200	78.82	0.74	14600	Yes
	0.42	0.00	1.00	12100	100	76.53	0.74	14700	Yes
	0.42	0.00	1.25	11800	-200	69.19	0.73	15000	Yes
Panel C: Comparing different values for μ									
	0.42	0.00	1.00	12100	100	76.53	0.74	14700	Yes
	0.42	0.10	1.00	13100	-100	71.46	0.80	16000	Yes
	0.42	0.20	1.00	14100	-200	63.36	0.87	17400	Yes
	0.42	0.30	1.00	10000	-4600	-26.72	0.82	20700	No
Panel D: Optimal Tax Schedule over Business Cycle - with Wage and Unemployment Responses									
Recession	0.34	0.00	0.73	12500	400	84.44	0.75	14500	Yes
Normal	0.42	0.00	0.90	12200	200	78.82	0.74	14600	Yes
Boom	0.48	0.00	1.06	11900	0	73.13	0.74	14800	Yes
Panel E: Optimal Tax Schedule over Business Cycle - without Wage and Unemployment Responses									
Recession	0.34	0.00	1.00	12200	200	79.16	0.74	14700	Yes
Normal	0.42	0.00	1.00	12100	100	76.53	0.74	14700	Yes
Boom	0.48	0.00	1.00	12000	100	74.76	0.74	14700	Yes

Notes: The table shows simulations of the optimal tax schedule based on equation 15 in the text. All simulations are based on the single worker earnings distribution in the March 2011 CPS and assume the parameter values $\nu = 4.00$, $\varepsilon = 0.25$ and - for wages above \$20,000 - $\eta = 0$. In Pandel D, both η and the Macro/micro ratio vary between boom and recession, while in Panel E the Macro/micro ratio is held constant at 1. Note that for high values of μ (around 0.3 and higher), the optimizer does not converge to a solution for large values of ν (see last column).

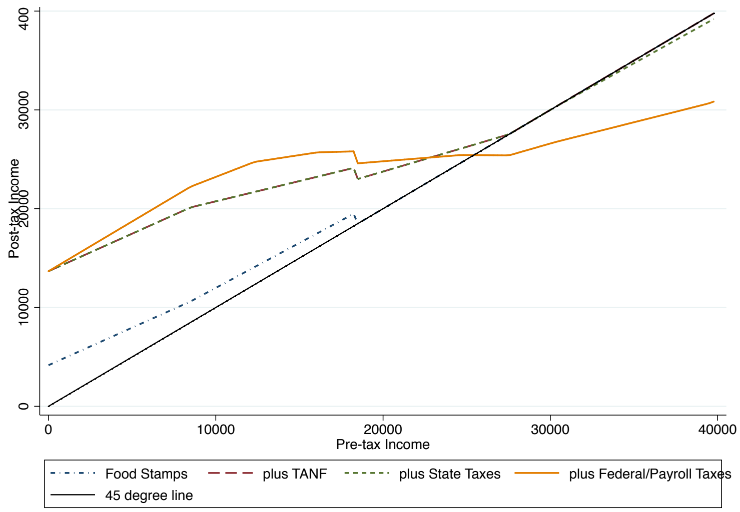
Table A-7: Simulating the Optimal Tax and Transfer Schedule - $\nu = 0.50$ and and Higher Intensive Margin Elasticity for Workers with less than

	η	μ	$\frac{\frac{d\mathscr{H}}{dT}}{\frac{d\mathscr{H}}{dT}\Big\vert^{Micro}}$	Demogrant b	Emp. Tax $T_1 + b$	ATR (6k-0)	Avg. MTR	Break-even $T(w) = 0$	Solution
Panel A: Alternative η - Macro/Micro ratio = 1, $\mu = 0$									
	0.01	0.00	1.00	8300	500	79.04	0.52	11800	Yes
Macro Est.	0.42	0.00	1.00	6600	100	46.81	0.48	13500	Yes
Micro Est.	0.60	0.00	1.00	6100	-100	35.95	0.47	14300	Yes
	1.00	0.00	1.00	5100	-300	19.49	0.43	15700	Yes
Panel B: Comparing different Macro-Micro Participation Ratios									
	0.42	0.00	0.75	9200	400	71.88	0.59	13400	Yes
Benchmark	0.42	0.00	0.90	7800	300	59.29	0.53	13400	Yes
	0.42	0.00	1.00	6600	100	46.81	0.48	13500	Yes
	0.42	0.00	1.25	3100	-2100	-5.43	0.41	14500	Yes
Panel C: Comparing different values for μ									
	0.42	0.00	1.00	6600	100	46.81	0.48	13500	Yes
	0.42	0.10	1.00	7500	0	42.09	0.53	15800	Yes
	0.42	0.20	1.00	8400	-100	36.44	0.59	18500	Yes
	0.42	0.30	1.00	9600	-100	29.65	0.67	21200	Yes
Panel D: Optimal Tax Schedule over Business Cycle - with Wage and Unemployment Responses									
Recession	0.34	0.00	0.73	9500	400	76.36	0.60	13300	Yes
Normal	0.42	0.00	0.90	7800	300	59.29	0.53	13400	Yes
Boom	0.48	0.00	1.06	5600	-200	32.51	0.45	14000	Yes
Panel E: Optimal Tax Schedule over Business Cycle - without Wage and Unemployment Responses									
Recession	0.34	0.00	1.00	6900	200	52.09	0.49	13200	Yes
Normal	0.42	0.00	1.00	6600	100	46.81	0.48	13500	Yes
Boom	0.48	0.00	1.00	6500	0	43.11	0.48	13800	Yes

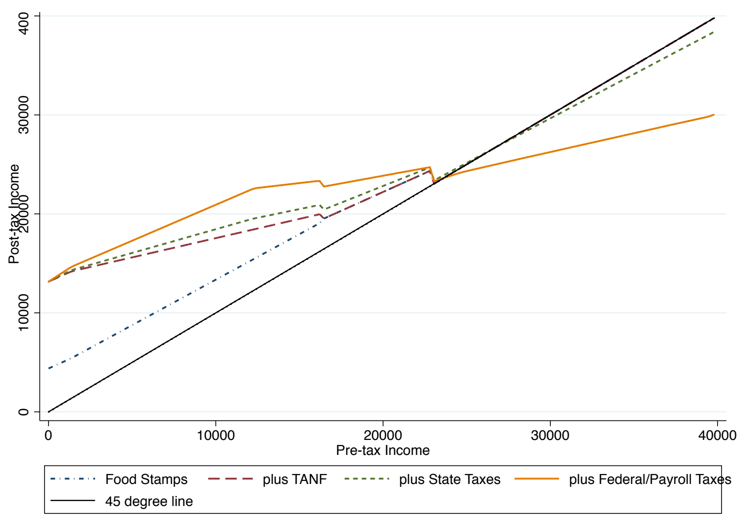
Notes: The table shows simulations of the optimal tax schedule based on equation 15 in the text. All simulations are based on the single worker earnings distribution in the March 2011 CPS and assume the parameter values $\nu = 0.50$, $\varepsilon = 0.50$ and - for wages above \$20,000 - $\eta = 0$. In Pandel D, both η and the Macro/micro ratio vary between boom and recession, while in Panel E the Macro/micro ratio is held constant at 1.

F Appendix Figures

Figure A-1: Budget Set Components



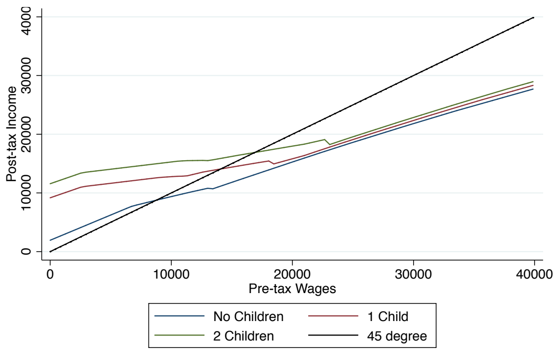
(a) California, Year 2000



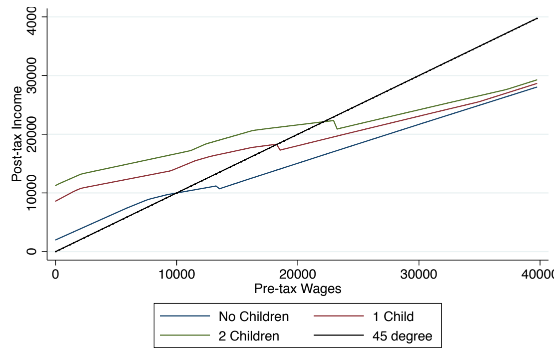
(b) New York, Year 2000

Notes: The figure shows the budget sets of a person with 2 children broken up by the individual components. The 45 degree line would be post-tax income in the absence of any taxes. The dashed blue line is pre-tax income plus foodstamps. The red line adds TANF, the green line adds state taxes and finally the yellow line adds federal taxes (including the EITC) and FICA taxes. Panel (a) shows the budget set for California in the year 2000. Panel (b) shows the budget for New York in the year 2000. The x-axis corresponds to pre-tax earnings, and the y-axis to post-tax and transfer income. Each line corresponds to the budget set of a single individual with either zero, one or two kids. The black line represents the 45 degree line.

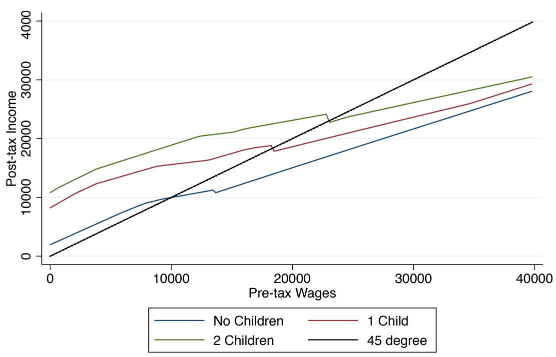
Figure A-2: Example Budget Sets for Selected States and Years



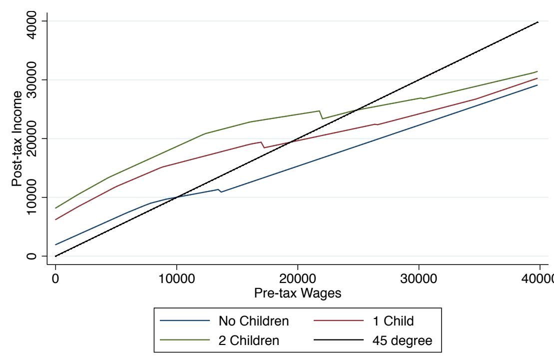
(a) Ohio, Year 1984



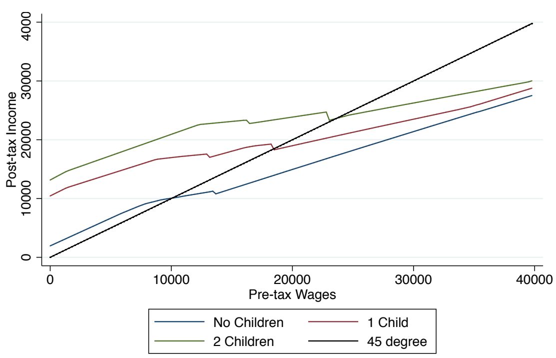
(b) Ohio, Year 1994



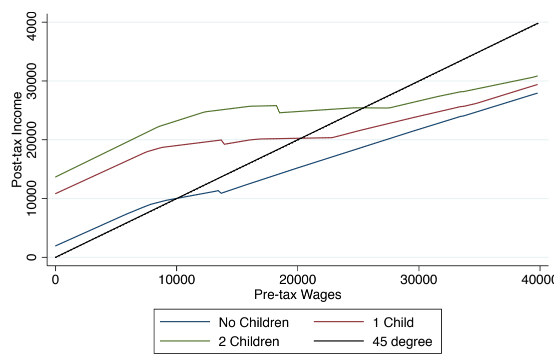
(c) Ohio, Year 2000



(d) Texas, Year 2000



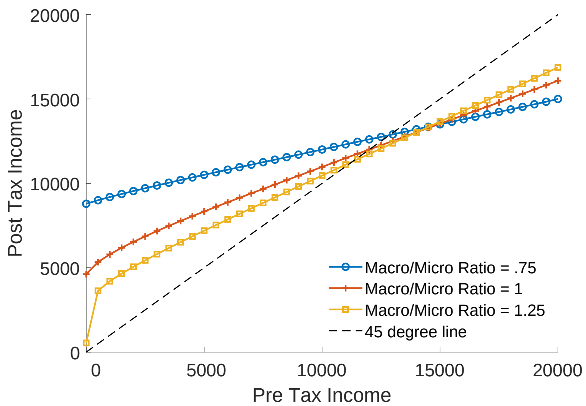
(e) New York, Year 2000



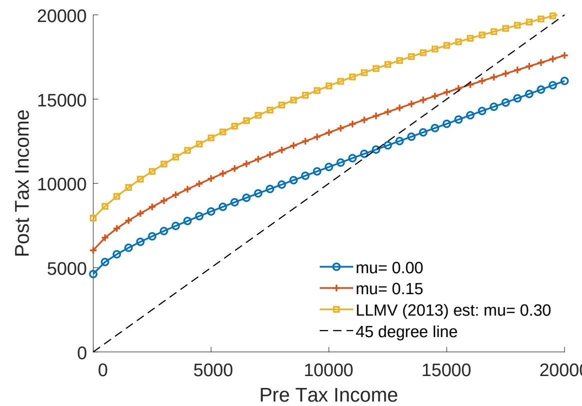
(f) California, Year 2000

Notes: The figure shows the budget sets of individuals in our sample by number of children for a selected sample of states and years. The x-axis corresponds to pre-tax earnings, and the y-axis to post-tax and transfer income. Each line corresponds to the budget set of a single individual with either zero, one or two kids. The black line represents the 45 degree line.

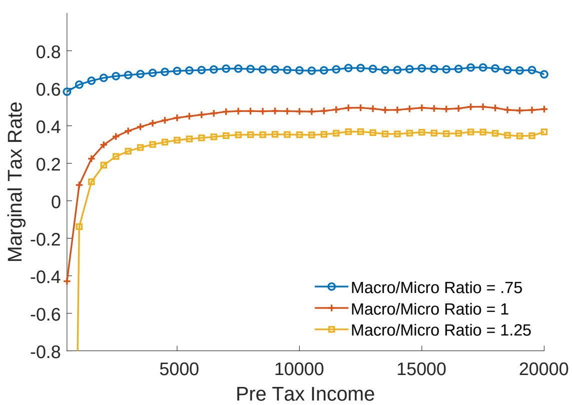
Figure A-3: The Effect of the Macro-Micro Participation Ratio and μ on the Optimal Tax and Transfer Schedule - Low Redistributive Preferences $\nu = 0.25$



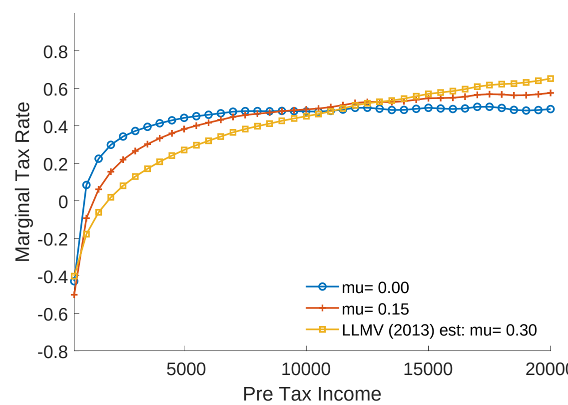
(a) Post vs. Pre-tax income, Alternative Macro/Micro ratio



(b) Post vs. Pre-tax income, Alternative μ



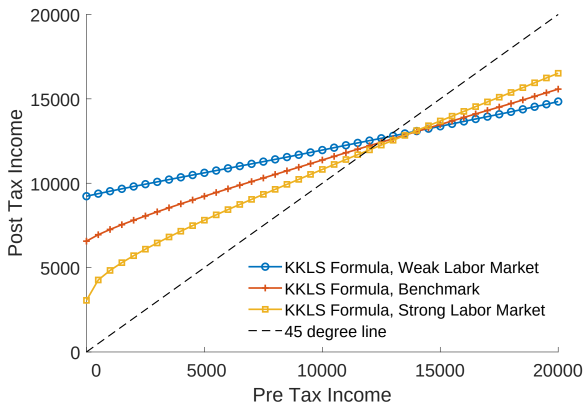
(c) Marginal tax rates, Alternative Macro/Micro ratio



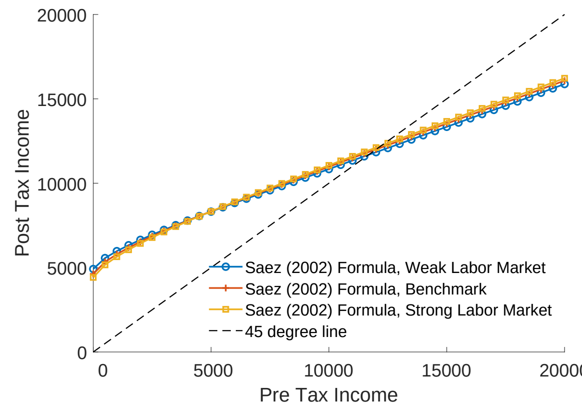
(d) Marginal tax rates, Alternative μ

Simulations of the optimal tax and transfer schedule using eqn (15) in the paper. Panels a) and c) contrast the optimal tax and transfer schedule under different values for the Macro/Micro participation response ratio while setting $\mu = 0$. Panels b) and d) contrast the optimal schedule for different values of μ while setting the Macro/Micro participation response ratio to 1. All simulations assume: $\varepsilon = 0.25$, $\eta_{low} = 0.42$, $\eta_{high} = 0$ and the redistributive taste parameter $\nu = 0.25$.

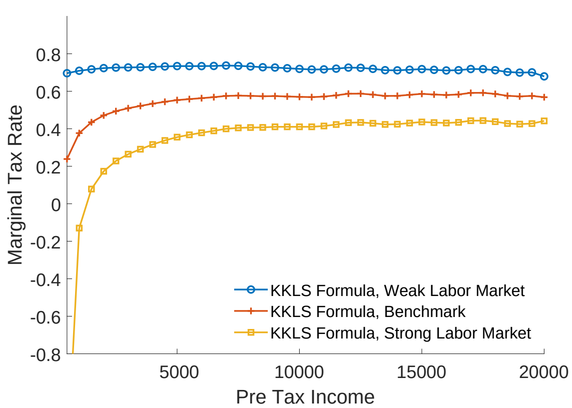
Figure A-4: Optimal Tax and Transfer Schedule in Weak vs. Strong Labor Markets - Low Redistributive Preferences $\nu = 0.25$



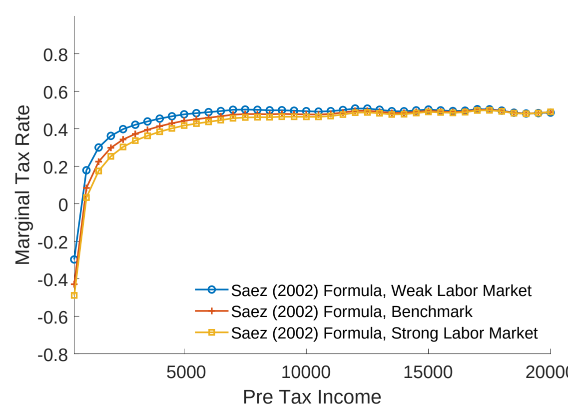
(a) KKLS formula: Post vs. Pre-tax income



(b) Saez (2002) formula: Post vs. Pre-tax income



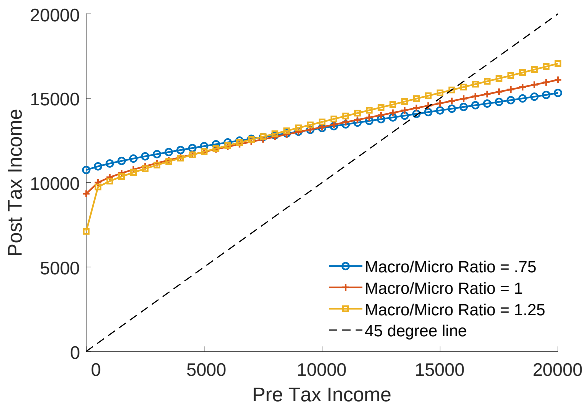
(c) KKLS formula: Marginal tax rates



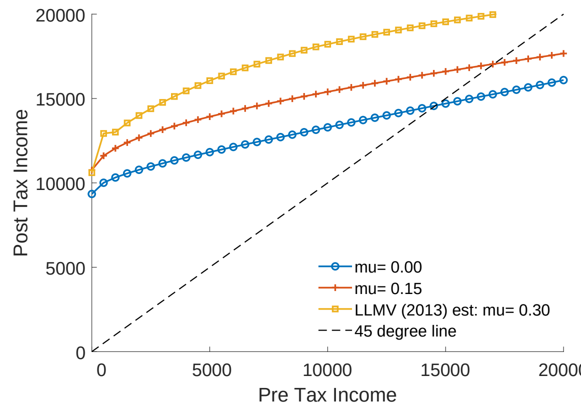
(d) Saez (2002) formula: Marginal tax rates

Notes: Simulations of the optimal tax and transfer schedule using eqn (15) in the paper. Panels a) and c) contrast the optimal tax and transfer schedule during weak and strong labor markets using the empirical estimates in the paper. Panels a) and c) use the optimal tax formula in the paper where both the employment elasticity η_{low} varies between boom and recessions as well as the Macro/Micro participation response. Panels b) and d) set the Macro/Micro participation response ratio to 1 and let's only the employment elasticity vary over the cycle and thus corresponds to the Saez (2002) formula . The solid line shows the tax schedule using the weak labor market estimates from Table 4 based on the 6 month change in the unemployment rate. The line with plus signs shows the tax schedule for the corresponding strong labor market estimates from Table 4. All simulations assume $\varepsilon = 0.25$, $\eta_{high} = 0$, $\mu = 0$ and the redistributive taste parameter $\nu = 0.25$.

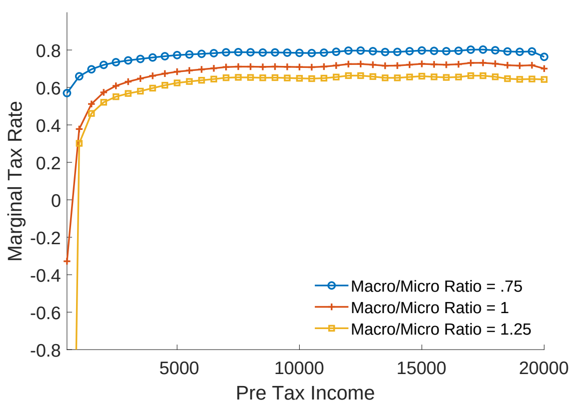
Figure A-5: The Effect of the Macro-Micro Participation Ratio and μ on the Optimal Tax and Transfer Schedule - Medium High Redistributive Preferences $\nu = 1$



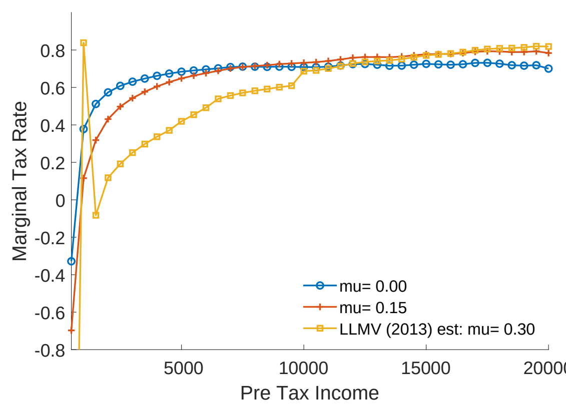
(a) Post vs. Pre-tax income, Alternative Macro/Micro ratio



(b) Post vs. Pre-tax income, Alternative μ



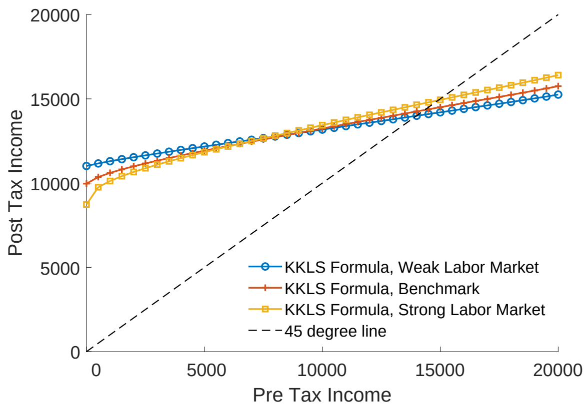
(c) Marginal tax rates, Alternative Macro/Micro ratio



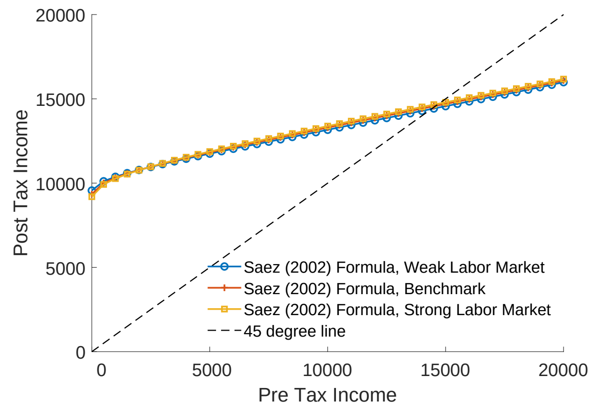
(d) Marginal tax rates, Alternative μ

Simulations of the optimal tax and transfer schedule using eqn (15) in the paper. Panels a) and c) contrast the optimal tax and transfer schedule under different values for the Macro/Micro participation response ratio while setting $\mu = 0$. Panels b) and d) contrast the optimal schedule for different values of μ while setting the Macro/Micro participation response ratio to 1. All simulations assume: $\varepsilon = 0.25$, $\eta_{low} = 0.42$, $\eta_{high} = 0$ and the redistributive taste parameter $\nu = 1$. Note that for high values of μ (around 0.3 and higher), the optimizer does not converge to a solution for large values of ν, which is why the yellow line (LLMV) in figure (b) and (d) is so choppy.

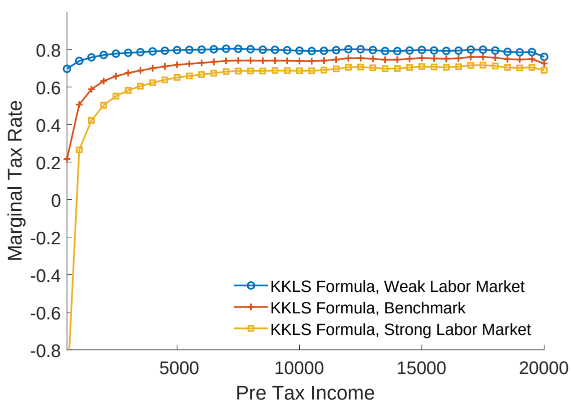
Figure A-6: Optimal Tax and Transfer Schedule in Weak vs. Strong Labor Markets - Medium High Redistributive Preferences $\nu = 1$



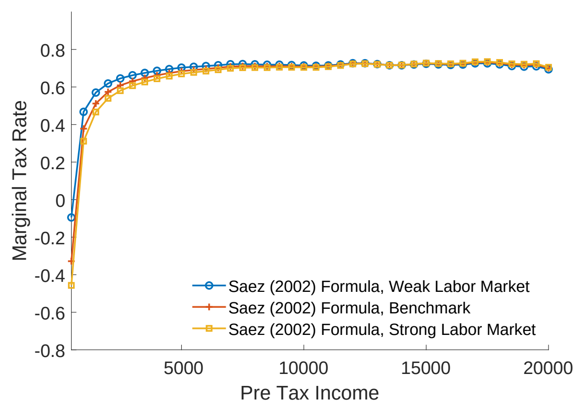
(a) KKLS formula: Post vs. Pre-tax income



(b) Saez (2002) formula: Post vs. Pre-tax income



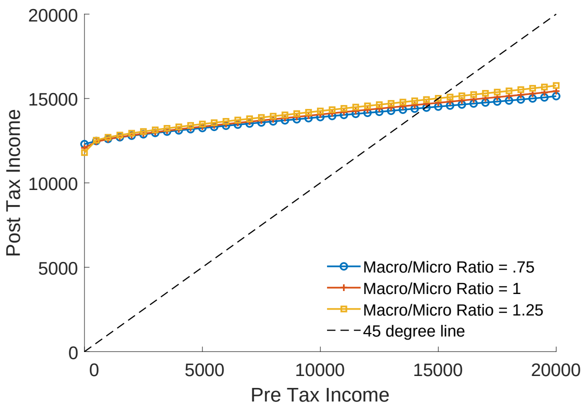
(c) KKLS formula: Marginal tax rates



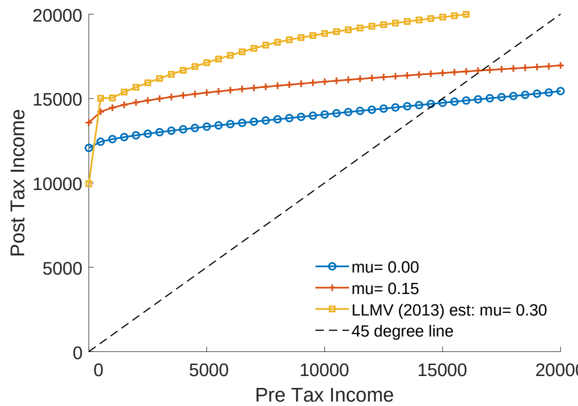
(d) Saez (2002) formula: Marginal tax rates

Notes: Simulations of the optimal tax and transfer schedule using eqn (15) in the paper. Panels a) and c) contrast the optimal tax and transfer schedule during weak and strong labor markets using the empirical estimates in the paper. Panels a) and c) use the optimal tax formula in the paper where both the employment elasticity η_{low} varies between boom and recessions as well as the Macro/Micro participation response. Panels b) and d) set the Macro/Micro participation response ratio to 1 and let's only the employment elasticity vary over the cycle and thus corresponds to the Saez (2002) formula . The solid line shows the tax schedule using the weak labor market estimates from Table 4 based on the 6 month change in the unemployment rate. The line with plus signs shows the tax schedule for the corresponding strong labor market estimates from Table 4. All simulations assume $\varepsilon = 0.25$, $\eta_{high} = 0$, $\mu = 0$ and the redistributive taste parameter $\nu = 1$.

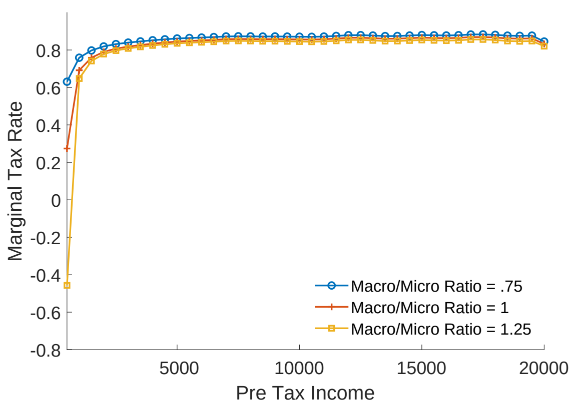
Figure A-7: The Effect of the Macro-Micro Participation Ratio and μ on the Optimal Tax and Transfer Schedule - High Redistributive Preferences $\nu = 4$



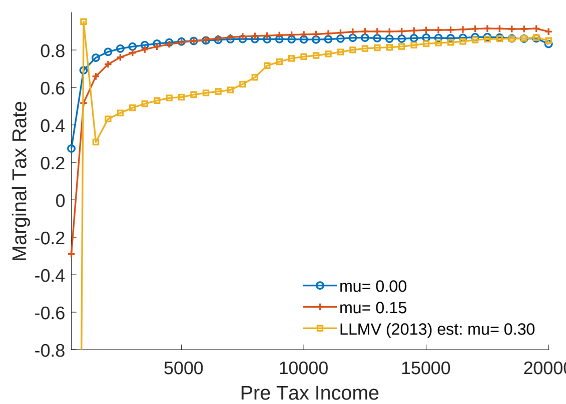
(a) Post vs. Pre-tax income, Alternative Macro/Micro ratio



(b) Post vs. Pre-tax income, Alternative μ



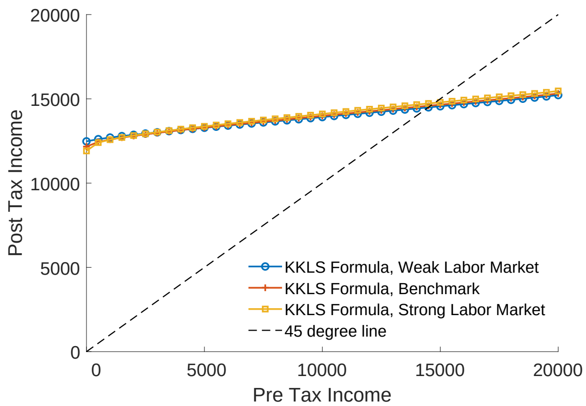
(c) Marginal tax rates, Alternative Macro/Micro ratio



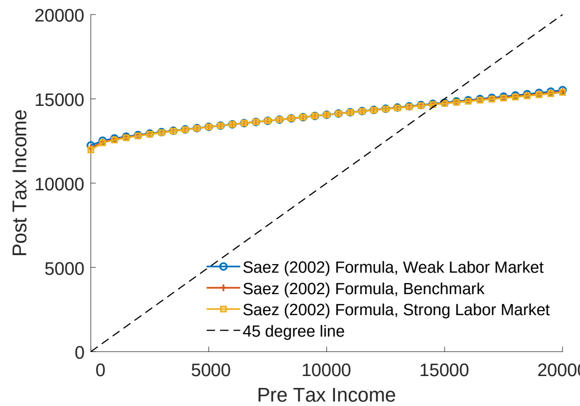
(d) Marginal tax rates, Alternative μ

Simulations of the optimal tax and transfer schedule using eqn (15) in the paper. Panels a) and c) contrast the optimal tax and transfer schedule under different values for the Macro/Micro participation response ratio while setting $\mu = 0$. Panels b) and d) contrast the optimal schedule for different values of μ while setting the Macro/Micro participation response ratio to 1. All simulations assume: $\varepsilon = 0.25$, $\eta_{low} = 0.42$, $\eta_{high} = 0$ and the redistributive taste parameter $\nu = 4$. Note that for high values of μ (around 0.3 and higher), the optimizer does not converge to a solution for large values of ν , which is why the yellow line (LLMV) in figure (b) and (d) is so choppy.

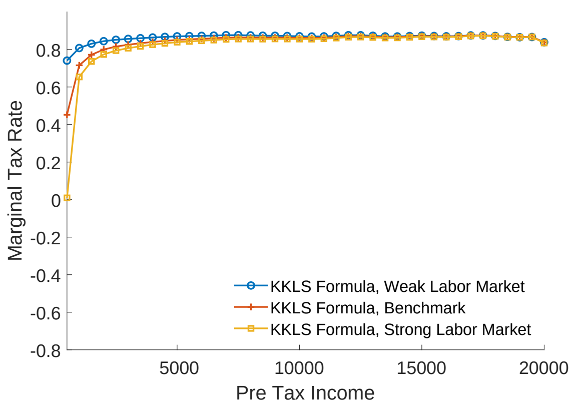
Figure A-8: Optimal Tax and Transfer Schedule in Weak vs. Strong Labor Markets - High Redistributive Preferences $\nu = 4$



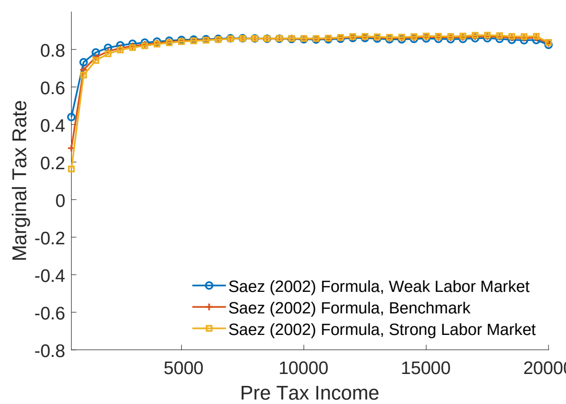
(a) KKLS formula: Post vs. Pre-tax income



(b) Saez (2002) formula: Post vs. Pre-tax income



(c) KKLS formula: Marginal tax rates



(d) Saez (2002) formula: Marginal tax rates

Notes: Simulations of the optimal tax and transfer schedule using eqn (15) in the paper. Panels a) and c) contrast the optimal tax and transfer schedule during weak and strong labor markets using the empirical estimates in the paper. Panels a) and c) use the optimal tax formula in the paper where both the employment elasticity η_{low} varies between boom and recessions as well as the Macro/Micro participation response. Panels b) and d) set the Macro/Micro participation response ratio to 1 and let's only the employment elasticity vary over the cycle and thus corresponds to the Saez (2002) formula . The solid line shows the tax schedule using the weak labor market estimates from Table 4 based on the 6 month change in the unemployment rate. The line with plus signs shows the tax schedule for the corresponding strong labor market estimates from Table 4. All simulations assume $\varepsilon = 0.25$, $\eta_{high} = 0$, $\mu = 0$ and the redistributive taste parameter $\nu = 4$.