Optimal Income Taxation under Unemployment: A Sufficient Statistics Approach*

PRELIMINARY DRAFT - PLEASE DO NOT CITE

Kory KROFT†
University of Toronto, NBER

Kavan KUCKO‡
Boston University

Etienne LEHMANN§
CRED (TEPP) University Panthéon-Assas Paris II

Johannes SCHMIEDER¶
Boston University, NBER and IZA

February 13, 2015

Abstract

This paper is the first to derive an empirically implementable formula for the optimal income tax, in the presence of unemployment. This formula nests a broad variety of structures of the labor market, such as the standard competitive model with fixed or flexible wages and models with matching frictions. As such, we are able to show that several previously derived optimal tax formulas are nested by this formula. Our theoretical results show that several reduced-form parameters are welfare-relevant: the macro employment elasticity with respect to taxes and the micro and macro participation elasticities with respect to taxes. We estimate all three of these reduced-form parameters using policy variation in tax liabilities stemming from the U.S. tax and transfer system for over 20 years. Since our tax formula is stated in terms of ?sufficient statistics?, we numerically solve for the optimal tax schedule using our empirical estimates and discuss how the results compare to those in the literature. Finally we also provide estimates of how employment and participation elasticities vary over the business cycle that suggest ? together with our theoretical results ? that in recessions the optimal tax schedule looks more like a Negative Income Tax (NIT) and less like an Earned Income Tax Credit (EITC).

*We would like to thank Raj Chetty, Kevin Lang, Austin Nichols, Claudia Olivetti, Daniele Paserman, Emmanuel Saez, Aleh Tsyvinski, as well as seminar participants at Boston University, McGill University, University of Connecticut, University of Toronto, SOLE Meeting 2013, and the NTA Meeting 2014, for many helpful comments on this project. All errors are our own.

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

‡Email: kavan.kucko@gmail.com. Adress: 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, 75 005 Paris, France. Etienne Lehmann is also research fellow at CREST, IRES-Université Catholique de Louvain, IDEP, IZA and CESifo and is junior member of IUF.

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

I Introduction

II The theoretical model

There are many theories of how tax policy affects the labor market. In this paper, rather than selecting one specific representation of the labor market, we derive an optimal tax formula in a general model that is consistent with a rich set of labor market responses to taxation. Following Chetty (2009), we use this "benchmark" model to identify the sufficient statistics that are necessary to compute the optimal tax policy. Then, we consider several special cases to illustrate how different models considered in the literature appear as particular cases of our general framework.

II.1 The general model

Labor markets

We start by generalizing the "formal model" in the appendix of Saez (2002) by introducing unemployment and wage responses to taxation. The size of the population is normalized to 1. There are $I + 1$ occupations, indexed by $i \in \{0, 1, ..., I\}$. Occupation 0 corresponds to non-employment. All other occupations correspond to a specific labor market where the gross wage is w_i, the net wage (or consumption) is c_i and the tax liability is $T_i = w_i - c_i$. Whatever the tax policy, gross wages are increasing in i and correspond to occupations that are increasing in skills, so $0 < w_1 < ... < w_I$.[1] The assumption of a finite number of occupations is made for tractability. It is not restrictive as the case of a continuous wage distribution can be approximated by increasing the number I of occupations to infinity. The timing of our static model is:

1. The government chooses the tax policy.

2. Each individual m chooses the occupation $i \in \{0, ..., I\}$ to participate in.

3. For each labor market $i \in \{1, ..., I\}$, only a fraction $p_i \in (0, 1]$ of participants are employed, receive gross wage w_i, pay tax T_i and consume the after-tax wage $c_i = w_i - T_i$. The remaining fraction $1 - p_i$ of participants are unemployed.

Unlike Saez (2002), we make a distinction among the *non-employed* individuals between the *unemployed* who search for a job in a specific labor market and fail to find one and the *non-participants* who choose not to search for a job. For each labor market $i \in \{1, ..., I\}$, k_i denotes the number of participants, $p_i \in (0, 1]$ denotes the fraction of them who find a job and are working, hereafter the *conditional employment probability*, and $h_i = k_i\, p_i$ denotes the number of employed workers. The

[1] This assumption reproduces the implication of the single crossing assumption in Mirrlees (1971), that more skilled workers earn a higher income according to the second order incentive constraint.

number of unemployed individuals in labor market i is $k_i - h_i = k_i(1 - p_i)$ and the unemployment rate is $1 - p_i$. The number of non-participants is k_0. The number of non-employed verifies $h_0 = k_0 + \sum_{i=1}^{I} k_i(1 - p_i)$.

All the non-employed, whether non-participants or unemployed, receive the same welfare benefit denoted b. This is because the informational structure of our static model prevents benefits from being history-dependent. Moreover, as the government only observes income, it cannot distinguish non-participants from unemployed individuals. This latter assumption seems more realistic than the polar one where the government can perfectly monitor job search. The policy choice of the government is therefore represented by the vector $\mathbf{t} = (T_1, ..., T_I, b)'$. The government faces the following budget constraint:

$$\sum_{i=1}^{I} T_i\, h_i = b\, h_0 + E \qquad \Leftrightarrow \qquad \sum_{i=1}^{I} (T_i + b)\, h_i = b + E \tag{1}$$

where $E \geq 0$ is an exogenous amount of public good to finance. One more employed worker in occupation i increases the government's revenues by the amount T_i of tax liability she pays, plus the amount of welfare benefit b she no longer receives, the sum of two defining the *employment tax*.[2] The budget constraint states that the sum of employment tax liabilities $T_i + b$ collected on all employed workers in all occupations finances the public good plus a lump-sum rebate b over all individuals.

Labor supply decisions

The structure of labor supply is as follows. We let $u(\cdot)$ be the cardinal representation of the utility individuals derive from consumption. This function is assumed increasing and weakly concave. Individual m faces an additional utility cost d_i for working in occupation i and a utility cost $\chi_i(m)$ for searching a job in labor market i.[3] Individual m enjoys a utility level equal to $u(c_i) - d_i - \chi_i(m)$ if she finds a job in labor market i, a utility level equal to $u(b) - \chi_i(m)$ if she fails to find a job in labor market i, and a utility level $u(b)$ if she chooses not to search for a job. Let $U_i = p_i\ (u(c_i) - d_i) + (1 - p_i)\ u(b)$ denote the *gross* expected utility of searching a job in occupation i absent any participation cost. Let $U_0 = u(b)$ be the utility expected out of the labor force. Individual m expects utility $U_i - \chi_i(m)$ by searching for a job in labor market i. She chooses to search if and only if $U_i - \chi_i(m) > U_j - \chi_j(m)$ for all $j \in \{0, ..., I\} \setminus \{i\}$. The set of individuals

[2] The literature uses instead the terminology *participation tax*, which we find confusing whenever unemployment is introduced. The *employment* tax $T_i + b$ captures the change in tax revenue for each additional *employed* worker. An additional *participant* being only employed with probability p_i, the change in tax revenue for each additional participant is only $(T_i + b)p_i$, which should correspond to the *participation* tax.

[3] We denote $\chi_0(m) = 0$. We furthermore assume that $\chi_i(m) = +\infty$ if individual m does not have the required skill to work in occupation i.

choosing to participate in labor market i is:

$$M_i(U_1,...,U_I,u(b)) \stackrel{\text{def}}{\equiv} \left\{ m | i = \arg\max_{j \in \{0,...,I\}} \quad U_j - \chi_j(m) \right\}$$

Assuming that the distribution of participation costs $\chi_i(m)$ across the population is smooth enough and denoting $\mu(\cdot)$ the distribution of individuals, the number k_i of participants in i is a continuously differentiable function of expected utility in each occupation through:

$$k_i = \hat{\mathcal{K}}_i(U_1,...,U_I,u(b)) \stackrel{\text{def}}{\equiv} \mu(M_i(U_1,...,U_I,u(b)))$$

Wage and labor demand responses

Rather than specify the micro-foundations of the labor market, we use reduced-forms to describe the responses of wages and conditional employment probabilities to taxation. In labor market i, the *macroeconomic* responses of gross wages to taxation are described by $w_i = \mathscr{W}_i(\mathbf{t})$, of net wages by $c_i = \mathscr{C}_i(\mathbf{t}) \stackrel{\text{def}}{\equiv} \mathscr{W}_i(\mathbf{t}) - T_i$ and of the conditional employment probability by $p_i = \mathscr{P}_i(\mathbf{t})$. At this general stage, we only assume that these functions are differentiable, that $\mathscr{P}(\cdot)$ takes values in $(0,1]$ and that $0 < b < \mathscr{W}_1(\mathbf{t}) < ... < \mathscr{W}_I(\mathbf{t})$ for all tax policies $\mathbf{t}$. The latter assumption ensures that occupations indexed with a higher i correspond to labor markets with higher skills. Gross expected utility in labor market i is given by:

$$\mathscr{U}_i(\mathbf{t}) \stackrel{\text{def}}{\equiv} \mathscr{P}_i(\mathbf{t}) \ (u(\mathscr{C}_i(\mathbf{t})) - d_i) + (1 - \mathscr{P}_i(\mathbf{t}))\, u(b) \tag{2}$$

and:

$$\frac{\partial \mathscr{U}_i}{\partial T_j} = \left[\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{1}{p_i} \frac{\partial \mathscr{P}_i}{\partial T_j} \frac{u(c_i) - d_i - u(b)}{u'(c_i)} \right] p_i\, u'(c_i) \tag{3}$$

Employment is given by:

$$h_i = \mathcal{H}_i(\mathbf{t}) \stackrel{\text{def}}{\equiv} \mathcal{K}_i(\mathbf{t}) \mathscr{P}_i(\mathbf{t}) \tag{4}$$

where participation decisions are determined through:

$$k_i \equiv \mathcal{K}_i(\mathbf{t}) \stackrel{\text{def}}{\equiv} \hat{\mathcal{K}}\left(\mathscr{U}_1(\mathbf{t}),...,\mathscr{U}_I(\mathbf{t}),u(b)\right)$$

Hence the macroeconomic participations responses are given by:

$$\frac{\partial \mathcal{K}_i}{\partial T_j} = \sum_{\ell=1}^{I} \frac{\partial \mathscr{U}_\ell}{\partial T_j} \frac{\partial \hat{\mathcal{K}}_i}{\partial U_\ell} = \sum_{\ell=1}^{I} \left[\frac{\partial \mathscr{C}_\ell}{\partial T_j} + \frac{1}{p_\ell} \frac{\partial \mathscr{P}_\ell}{\partial T_j} \frac{u(c_\ell) - d_\ell - u(b)}{u'(c_\ell)} \right] p_\ell\, u'(c_\ell) \frac{\partial \hat{\mathcal{K}}_i}{\partial U_\ell} \tag{5}$$

Social objective

We assume that the government maximizes a social welfare function $\Omega(U_1,...,U_I,u(b))$ that depends only on individuals' *expected* utilities. A specific example of such a welfare function is when the government's objective consists in a weighted sum of individuals utility:

$$\Omega(U_1,...,U_I,u(b)) = \int \gamma(m) \left[\max_i \quad p_i \left[u(c_i) - d_i\right] + (1 - p_i)\, u(b) - \chi_i(m) \right] d\mu(m)$$

where the weights $\gamma(m)$ may vary across individuals. In the particular case where the utility function $u(\cdot)$ is linear, it is the variation of weights with the characteristics of individuals through the heterogeneity in $\gamma(\cdot)$ that generates the social desire for redistribution. We generalize this case by assuming that the government applies a (weakly) concave *social valuation* function $\Phi(\cdot, m)$ on the expected utility of individuals, this valuation being allowed to vary across individuals of different characteristics (hence the second argument). The government's objective thus takes the form:

$$\Omega(U_1, ..., U_I, u(b)) \stackrel{\text{def}}{\equiv} \int \Phi\left(\max_i \quad U_i - \chi_i(m); m\right) \, d\mu(m) \tag{6}$$

The optimal policy

The government chooses the tax policy $\mathbf{t} = (T_1, ..., T_I, b)'$ to maximize (6) subject to the budget constraint (1). Let $\lambda > 0$ denote the Lagrange multiplier associated with the latter constraint. Let Φ_U denote the partial derivative of the social valuation Φ with respect to individual's utility. Following Saez (2001, 2002), we define the marginal social welfare weight of workers in occupation $i \in \{0, ..., I\}$ as:

$$g_i \stackrel{\text{def}}{\equiv} \frac{1}{k_i} \frac{\partial \Omega}{\partial U_i} \frac{u'(c_i)}{\lambda} = \frac{p_i \, u'(c_i) \, \int_{m \in M_i} \Phi_U \left(U_i - \chi_i(m); m\right) \, d\mu(m)}{\lambda \, h_i} \tag{7}$$

The social weight g_i represents the social value in monetary terms of transferring an additional dollar to an individual working in occupation i. The government is indifferent between giving one more dollar to an individual employed in labor market i and g_i more dollars of public funds. Using Equations (3) and (7), the first-order condition with respect to the tax liability T_j in labor market j is:

$$0 = \underbrace{h_j}_{\text{Mechanical effect}} + \underbrace{\sum_{i=1}^{I} \frac{\partial \mathcal{H}_i}{\partial T_j} \left(T_i + b\right)}_{\text{Behavioral effects}} + \underbrace{\sum_{i=1}^{I} \left[\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{1}{p_i} \frac{\partial \mathscr{P}_i}{\partial T_j} \, \frac{u(c_i) - d_i - u(b)}{u'(c_i)}\right] g_i \, h_i}_{\text{Social Welfare effects}} \tag{8}$$

or in matrix notation:

$$\mathbf{0} = \underbrace{\mathbf{h}}_{\text{Mechanical effect}} + \underbrace{\frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}} \cdot (\mathbf{T} + \mathbf{b})}_{\text{Behavioral effects}} + \underbrace{\mathcal{A} \cdot (\mathbf{g} \, \mathbf{h})}_{\text{Social Welfare effects}} \tag{9}$$

where for $f = \mathcal{K}, \hat{\mathcal{K}}, \mathcal{H}, \mathscr{U}, \mathscr{P}, \mathscr{W}$ and $x = T, U$, we denote $\frac{\mathbf{df}}{\mathbf{dx}}$ the square matrix of rank[4] I whose term in row j and column i is $\frac{\partial f_i}{\partial x_j}$. Moreover, $\mathbf{h} = (h_1, ..., h_I)'$ denotes the vector of employment levels, $\mathbf{g}\,\mathbf{h} = (g_1 h_1, ..., g_I h_I)'$ denotes the vector of welfare weights times employment levels, $\mathcal{A}$ denotes the square matrix of rank I whose term in row j and column i is $\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{1}{p_i} \frac{\partial \mathscr{P}_i}{\partial T_j} \, \frac{u(c_i) - d_i - u(b)}{u'(c_i)}$

[4] In particular, these matrices do not include partial derivatives with respect to b, nor do they include partial derivatives for occupation 0.

and $\cdot$ denotes the matrix product. A marginal increase in the tax liability T_j on employed worker in labor market j leads to the three following effects.

1. **Mechanical effect**: Absent any behavioral response, a unit increase in T_j increases the government's resources by a monetary amount equal to the number h_j of employed individuals in occupation j.

2. **Behavioral effects**: A unit increase in T_j induces a change $\partial\mathcal{H}_i/\partial T_j$ in the level of employment in occupation i. Each additional worker in occupation i pays tax T_i instead of receiving benefit b. For each additional worker in occupation i, the government increases its resources by the employment tax $T_i + b$. It is worth noting that $\partial\mathcal{H}_i/\partial T_j$ encapsulates labor supply and labor demand responses to changes in taxes T_j.

3. **Social welfare effects**: A unit increase in T_j affects the expected utility in occupation i by $\partial\mathscr{U}_i/\partial T_j$. Multiplying by the rate $\int_{m\in M_i} \Phi_U \left(U_i - \chi_i(m); m\right)/\lambda \; d\mu(m)$ at which each unit change in expected utility affects the social objective in monetary terms and using Equations (3) and (7), we get that the social welfare effect of tax T_j through occupation i is: $g_i h_i \left[\frac{\partial\mathscr{C}_i}{\partial T_j} + \frac{1}{p_i}\frac{\partial\mathscr{P}_i}{\partial T_j}\, \frac{u(c_i)-d_i-u(b)}{u'(c_i)}\right]$. Adding these effects for all occupations provides the social welfare effects. Note that because the social welfare function depends on expected utility U_i, the labor supply responses triggered by the change in tax liability only modifies the decisions of individuals between two occupations that are initially indifferent between these two occupations, and thus only have second-order effects on the social welfare objective, by the envelope theorem.

One difficulty in numerically implementing the optimal tax (8) is the estimation of the responses of net wages $\frac{\partial\mathscr{C}_i}{\partial T_j}$ and of the conditional employment probabilities $\frac{\partial\mathscr{P}_i}{\partial T_j}$ that appear in the social welfare effects. To overcome this problem, we show that it is possible to rewrite the optimal tax formula (8) not only in terms of the *macroeconomic* employment responses $\frac{\partial\mathcal{H}_i}{\partial T_j}$, but also in terms of the *macroeconomic* $\frac{\partial\mathcal{K}_i}{\partial T_j}$ and *microeconomic* participation responses. We define the latter as the responses of participation to a tax change in the hypothetical case where these tax changes would not affect wages $\mathbf{w} = (w_1, ..., w_I)'$ or conditional employment probabilities $\mathbf{p} = (p_1, ..., p_I)'$. This is, for instance, the case for tax reforms frequently considered in micro-econometric literature that affects only a small subset of the population, so that the general equilibrium effects of the reform on wage and conditional employment probabilities can be safely ignored. The matrices $\left.\frac{\mathbf{df}}{\mathbf{dx}}\right|_{\mathbf{w},\mathbf{p}}$ correspond to the microeconomic responses. According to Equation (3), the matrix $\mathcal{A}$ describes the corrective terms that enables one to obtain the macroeconomic responses from the microeconomic ones as described by the following Lemma:

Lemma 1. *The macroeconomic and microeconomic participation responses are related through*

$$\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}} = -\mathcal{A} \cdot \left.\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}} \qquad \textit{and}: \qquad \frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}} = -\mathcal{A} \cdot \left.\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}}$$

At the microeconomic level, a unit change in tax liability in occupation j only changes expected utility in occupation j by $p_j\, u'(c_j)$. Conversely, at the macroeconomic level, it may also affects wages and conditional employment probability in other occupations, so that, according to According to Equation (3), $\frac{\partial \mathscr{U}_i}{\partial T_j} = \left[\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{\partial \mathscr{P}_i}{\partial T_j} \frac{u(c_i)-d_i-u(b)}{p_i\; u'(c_i)}\right] p_i\; u'(c_i)$. Matrix $\mathcal{A}$ thus describes how general equilibrium effects at the macroeconomic level modify the responses of expected utility. Moreover, because individuals take their labor supply decisions by comparing expected utility in the different occupation, matrix $\mathcal{A}$ also describe macroeconomic participation responses are modified from microeconomic ones. Finally, as the social objective only depends on individuals' expected utility, the same matrix $\mathcal{A}$ also describes how the social welfare effects are magnified at the macroeconomic level, compared to the microeconomic level where there are fully described by the social welfare weights g_i times the density h_i. It is therefore empirically convenient to substitute the matrix ratio of macroeconomic over microeconomic participation responses for the matrix $\mathcal{A}$ of corrective terms. We therefore get (See Appendix A):

Proposition 1. *If* $\left.\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}}$ *is invertible, the the optimal tax system for occupations* $i = \{1, ..., I\}$ *solves the following system of equations in matrix form:*

$$\mathbf{0} = \mathbf{h} + \frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}} \cdot (\mathbf{T} + \mathbf{b}) - \frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}} \cdot \left(\left.\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}}\right)^{-1} \cdot (\mathbf{g}\; \mathbf{h}) \tag{10}$$

Equation (10) is expressed in terms of sufficient statistics. It implies that the ratio (in matrix terms) of macroeconomic to microeconomic participation responses are the sufficient statistics to estimate the matrix $\mathcal{A}$ in Equation (9), instead of the corrective terms that depends on net wages $\frac{\partial \mathscr{C}_i}{\partial T_j}$ and conditional employment probability responses $\frac{\partial \mathscr{P}_i}{\partial T_j}$. Importantly, the gap between microeconomic and macroeconomic responses does not matter for the behavioral effects, but only for the social welfare effects. This is because the matrix $\frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}}$ of macroeconomic employment responses already encapsulates the unemployment and wage responses in addition to the microeconomic participation responses.

Finally, for the sake of completeness, the first-order condition with respect to the welfare benefit b is (see Appendix B):

$$0 = -h_0 + \sum_{i=1}^{I} (T_i + b) \frac{\partial \mathcal{H}_i}{\partial b} + g_0 h_0 + \sum_{i=1}^{I} g_i h_i \left[\frac{\partial \mathscr{C}_i}{\partial b} + \frac{1}{p_i} \frac{\partial \mathscr{P}_i}{\partial b} \frac{u(c_i) - d_i - u(b)}{u'(c_i)}\right] \tag{11}$$

where the social marginal welfare weight on non-employed is:

$$g_0 \stackrel{\text{def}}{\equiv} \frac{u'(b)}{h_0} \left[\int_{m \in M_0} \frac{\Phi_U\left(u(b); m\right)}{\lambda} d\mu(m) + \sum_{i=1}^{I} \frac{g_i}{u'(c_i)} k_i (1 - p_i)\right] \tag{12}$$

In particular, if we furthermore assume there is no income effects, we get that the weighted sum of social welfare weights is 1 (See Appendix B):

$$g_0\, h_0 + \sum_{i=1}^{I} g_i\, h_i = 1$$

II.2 The case without unemployment responses

In this subsection, we consider the case where 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 $F(h_1, ..., h_I, Z)$.[5] 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. Appendix C shows that in such a case, the optimal tax formula is identical to the one obtained in the general model exposed in the appendix of Saez (2002), Equation (11) where the different types of labor are perfect substitutes.

$$0 = (1 - g_j)h_j + \sum_{i=1}^{I} (T_i + b)\ \left.\frac{\partial \mathcal{H}_i}{\partial T_j}\right|_{\mathbf{w},\mathbf{p}} \tag{13}$$

From the comparison of Equations (8) with (13), we deduce that the existence of unemployment responses to taxation (when $\frac{\mathbf{d}\mathscr{P}}{\mathbf{dT}} \neq \mathbf{0}$) modifies the optimal tax formula in the two following ways. First, the behavioral effects are modified because they depend in (8) on *macroeconomic* employment responses, inclusive of conditional employment probability responses $\frac{d\mathscr{P}_i}{dT_j}$, and not only on *microeconomic* labor supply responses $\left.\frac{\partial \mathcal{K}}{\partial \mathbf{T}}\right|_{\mathbf{w},\mathbf{p}}$, as in (13). Second, the social welfare effects are also modified because a change in tax liability in one labor market may also affect conditional employment probability and wages in another labor market and thereby the expected utility.

The production efficiency argument

Deriving Equation (13) is straightforward if on top of assuming away conditional employment probability responses ($\frac{\mathbf{d}\mathscr{P}}{\mathbf{dT}} = \mathbf{0}$), one also assumes perfect substitution across the different types of labor as does Saez (2002). Wage responses are then nil ($\frac{\mathbf{d}\mathscr{W}}{\mathbf{dT}} = 0$), so, macroeconomic responses coincide with microeconomic ones. Optimal tax formula (10) then directly simplifies to (13).

If one conversely assume that the different types of labor are imperfect substitutes, one would expect wage response effects to affect the optimal policy (Rothstein, 2010). However, Saez (2004)

[5] 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.

shows that the optimal tax formula remains intact and does not depend on the production function $F(\cdot)$ in such a case. To understand why (13) remains valid, consider as Lee and Saez (2008), a tax reform $\mathbf{dT} = (dT_1, ..., dT_I)'$ such that after-tax income in occupation j is modified by dc_j, while after-tax income in all other occupations remain unchanged. This reform induces a change in gross wages denoted $\mathbf{dw} = (dw_1, ..., dw_I)'$ where $dw_i = dT_i$ for $i \neq j$ and $dw_j = dc_j + dT_j$. Because there is no unemployment response, only the expected utility in occupation j is modified by $p_j u'(c_j) dc_j$. The social welfare effect of this reform is thus equal to $g_j\, h_j\, dc_j$. Moreover, the behavioral effects being induced only by labor supply responses and the expected utility being only modified in occupation j, the behavioral effects are given by $\sum_i (T_i + b)\ p_i \frac{\partial \hat{\mathcal{K}}_i}{\partial U_j}\ p_j u'(c_j) dc_j = -\sum_{i=1}^I (T_i + b)\ \left.\frac{\partial \mathcal{H}_i}{\partial T_j}\right|_{\mathbf{w},\mathbf{p}}\ dc_j$. Importantly, the social welfare effect and behavioral effects of this reform are the same at the microeconomic and at the macroeconomic level. Lastly, the mechanical effects are given by $\sum_{i=1}^I h_i\, dT_i = \sum_{i=1}^I h_i\, dw_i - dc_j$ and may be a priori different at the macroeconomic and at the microeconomic level because of the term $\sum_{i=1}^I h_i\, dw_i$. The latter term corresponds to minus the change in profits,[6] and is nil whenever firms are price-taker and the technology exhibits constant returns to scale. Therefore, even the mechanical effect is identical at the microeconomic and the macroeconomic level. This "production efficiency" argument (Diamond and Mirrlees, 1971) is the reason why the optimal tax formula does not depend on the degree of substitutability across labor inputs. Our contribution is to show that the "production efficiency" argument of Saez (2004) also applies when there is a positive but exogenous unemployment rate (i.e $p_i < 1$ on each labor market). This implies that if the unemployment rates are exogenous and if the production function exhibits constant returns to scale, the optimal tax schedule is given by (22) and depends only on microeconomic employment responses. Even if macroeconomic and microeconomic participation responses are different, provided these differences are only due imperfect substitutability, it sufficient to estimate only the microeconomic employment responses to taxation to compute the optimal tax policy. [7]

[6] Applying the envelope theorem to the profit function $\Pi(w_1, ..., w_I, r) \stackrel{\text{def}}{\equiv} \max_{h_1,...,h_I} F(h_1, ..., h_I) - \sum_{i=1}^I w_i h_i - rZ$ implies that $\frac{\partial \Pi}{\partial w_i} = -h_i$ and $\frac{\partial \Pi}{\partial r} = -Z$. Hence $d\Pi = \sum_{i=1}^I \frac{\partial \Pi}{\partial w_i} dw_i = -\sum_{i=1}^I h_i\, dw_i - Z\, dr$, which is equal to 0 because the profit function is constant at 0 for a price-taking firm operating under constant returns to scale and because the rate of return to capital r is assumed exogenous.

[7] Stiglitz (1982) and Naito (1999) propose 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 of Saez (2004) according to which the optimal tax formula does not depend on the elasticity of substitution across the different types of labor. 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), 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

The pure extensive case

If one now further restricts the model by assuming that labor supply responses are concentrated along the extensive margin, then one gets $\partial\mathcal{H}_i/\partial T_j = 0$ if $j \neq i$. Defining the employment extensive elasticity in occupation j through $\partial\mathcal{H}_j/\partial T_j = -\eta_j \frac{h_j}{c_j - b}$, we extend the optimal tax formula in the pure extensive model of Diamond (1980), Saez (2002, Equation (4)) and Choné and Laroque (2011) to the case where unemployment rate is positive but unresponsive to tax reforms:

$$\frac{T_j + b}{c_j - b} = \frac{1 - g_j}{\eta_j} \tag{14}$$

Equation (14) implies that if the welfare weight g_1 on the least paid employed workers is higher then 1, than the optimal tax policy verifies $-T_1 > b$, so it is optimal to transfer higher income to the working poor than to the non-employed. This is the case of an optimal negative employment tax at the bottom of the income distribution referred as an EITC type policy by Saez (2002). It is worth noting that labor supply responses being concentrated along the extensive margin is not sufficient for optimal employment tax rates to be negative for working poor. For instance, Choné and Laroque (2005) studied the pure extensive optimal tax model under a Maximin social welfare objective, so that all the welfare weights are concentrated on the non-employed, i.e. $g_0 > 0 = g_1, ..., g_I$. They obtain positive optimal employment tax rates despite all labor supply responses being concentrated on the extensive margin. This is because the social welfare effect is then nil so the optimal employment tax has to trade off a positive mechanical effect with the opposite of the behavioral effect, the sign of the latter being the sign of the employment tax.

The mixed case without unemployment

Finally, we consider the mixed case of Saez (2002) where there is no unemployment ($p_i \equiv 1$), workers are risk neutral so that there is no income response, and labor supply responses are concentrated along the intensive and the extensive margins. Each individual can choose between working in two consecutive occupations and not working. Hence, employment in occupation i can be written as a function $\widetilde{\mathcal{H}}_i(T_{i+1} - T_i, T_i - T_{i-1}, T_i + b)$ of marginal tax rates $T_{i+1} - T_i$ and $T_i - T_{i-1}$ and of the employment tax liabilities $T_i + b$. Moreover $\left.\frac{\partial\widetilde{\mathcal{H}}_i}{\partial T_i - T_{i-1}}\right|_{\mathbf{w},\mathbf{p}}$ being the mass of individual indifferent between working in occupation $i-1$ and in occupation i, we thus get $\left.\frac{\partial\widetilde{\mathcal{H}}_i}{\partial T_i - T_{i-1}}\right|_{\mathbf{w},\mathbf{p}} = -\left.\frac{\partial\widetilde{\mathcal{H}}_{i-1}}{\partial T_i - T_{i-1}}\right|_{\mathbf{w},\mathbf{p}} > 0$, which captures the intensive margin. The extensive margin is captured by

(1999) focus on the short-run hours of work and in-work effort responses along the intensive margin.

$\frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_j + b}$. The optimal tax formula (13) then simplifies to (See Appendix D):

$$\begin{aligned}
\left(T_j - T_{j-1}\right) \frac{\partial \widetilde{\mathcal{H}}_j}{\partial (T_j - T_{j-1})} &= \sum_{i=j}^{I} \left\{ (1 - g_i)\, h_i + (T_i + b)\, \frac{\partial \widetilde{\mathcal{H}}_i}{\partial (T_i + b)} \right\} \\
&= \sum_{i=j}^{I} \left(\overline{T_i + b} - (T_i + b) \right) \frac{\partial \widetilde{\mathcal{H}}_i}{\partial (T_i + b)}
\end{aligned} \tag{15}$$

where $\overline{T_i + b} \stackrel{\text{def}}{\equiv} (1 - g_i)\, h_i / \frac{\partial \widetilde{\mathcal{H}}_i}{\partial T_i + b}$ is, according to (14), the employment tax that would be optimal in the absence of intensive margin. The first part of Equation (15) is similar to optimal tax formula of Saez (2002), Equation (8) for the mixed case. The second part follows Jacquet et al. (2013). The optimal tax schedules trades off, on the one hand, to have employment tax liabilities $T_i + b$ as closed as possible to the optimal employment level $\overline{T_i + b}$ in the absence of intensive response to minimize the distortion along the extensive margin, and on the other hand, to have tax liabilities as closed as possible to constant ones, i.e. marginal tax rates as close as possible to zero, to minimize the distortions along the intensive margin. Hence, the optimal tax schedule "flattens" the optimal employment tax schedule in the absence of intensive margin. In particular, Appendix D shows the following:

Proposition 2. *If the optimal tax liability in the absence of intensive margin is increasing (i.e. if $\overline{T}_1 < ... < \overline{T}_I$) then the optimal tax schedule is also increasing (i.e. $T_1 < ... < T_I$). Moreover $T_1 > \overline{T}_1$.*

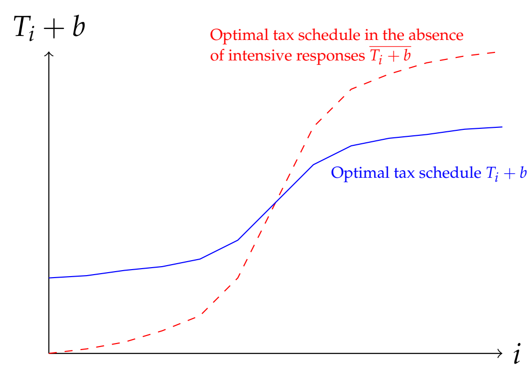

Figure 1: Intuition of Proposition 2

Proposition 2 extends to the discrete case Proposition 2 in Jacquet et al. (2013) which holds for a continuous income distribution. It is illustrated by Figure 1. The dashed curve corresponds to the optimal employment tax $\overline{T_i + b}$ in the absence of intensive responses. It describes the employment tax that equates the mechanical effect and the extensive labor supply responses. It is increasing by assumption. However, implementing this tax schedule implies high marginal tax rates, which would be too detrimental in terms of distortions along the extensive margin. The optimal tax schedule is therefore described by the solid curve which is increasing, but by a lesser extent than the dashed curve. Moreover, the solid curve is above (below) the dashed one for low (high) income levels to remain as close as possible to the dashed curve. As a consequence, if $g_1 \leq 1$ the optimal employment tax at the bottom in the absence of intensive responses is non negative, so the optimal employment tax is positive. This formalizes the intuition of Saez (2002) that optimal employment tax at the bottom is less likely to be negative the larger is the labor supply elasticity along the intensive margin.

II.3 The case with wage and unemployment responses but no-cross effects

In this section, we consider the simplest case with unemployment responses by assuming away cross effects, a simplifying assumption that will appear relevant in the empirical part. The no-cross effect assumption implies that microeconomic labor supply responses are concentrated along the extensive margin so that $\partial \hat{\mathcal{K}}_i / \partial U_j = 0$ if $i \neq j$. Moreover, it implies that for $i \neq j$, $\partial \mathscr{P}_i / \partial T_j = \partial \mathscr{W}_i / \partial T_j = 0$. Hence, we also get $\partial \mathscr{U}_i / \partial T_j = 0$ thereby $\partial \mathcal{K}_i / \partial T_j = \partial \mathcal{H}_i / \partial T_j = 0$ for $i \neq j$. This environment includes Landais et al. (2014) and Jacquet et al. (2014) as particular cases. Let $\eta_j \stackrel{\text{def}}{\equiv} -\frac{c_j - b}{h_j} \frac{\partial \mathcal{H}_j}{\partial T_j}$ denote the macroeconomic employment elasticity, $\pi_j \stackrel{\text{def}}{\equiv} -\frac{c_j - b}{k_j} \frac{\partial \mathcal{K}_j}{\partial T_j}$ denote the macroeconomic participation elasticity and $\pi_j^m \stackrel{\text{def}}{\equiv} -\frac{c_j - b}{k_j} \left. \frac{\partial \mathcal{K}_j}{\partial T_j} \right|_{w,p}$ denote the microeconomic participation elasticity. The optimal tax formula (10) then simplifies to:[8]

$$\frac{T_j + b}{c_j - b} = \frac{1 - \dfrac{\pi_j}{\pi_j^m} g_j}{\eta_j} \tag{16}$$

The no-cross effect environment is the simplest one to understand how the introduction of unemployment and wage responses modifies the optimal tax formula compared to the pure extensive margin case considered in Equation (14). First, from (4), the macroeconomic employment response η_j verifies $\eta_j = \frac{c_j - b}{p_j} \frac{\partial \mathscr{P}_j}{\partial T_j} + \pi_j$. It in particular encapsulates conditional employment responses $\frac{c_j - b}{p_j} \frac{\partial \mathscr{P}_j}{\partial T_j}$ in addition to the macroeconomic participation responses π_j. Moreover, wage and unemployment responses also modifies the macroeconomic participation responses π_j from the microeconomic ones π_j^m. Second, the response of expected utility may be different at the macroeconomic and microeconomic levels. This is because the macroeconomic responses encapsulates not only the direct of a tax change on consumption, but also the effects of a tax change on the wage $\frac{\partial \mathscr{W}_i}{\partial T_i} \neq 0$ and on the conditional employment probability $\frac{\partial \mathscr{P}_i}{\partial T_i} \neq 0$. The gap between microeconomic and macroeconomic responds also corresponds to the ratio of the macroeconomic to the microeconomic participation elasticity. So the welfare effect may be larger or lower than the social welfare weight g_i. We therefore get:

Proposition 3. *In the no cross effect case, the optimal employment tax is negative whenever* $g_1 > \frac{\pi_1^m}{\pi_1}$.

[8] In the case of II.2 where wage responses are due to imperfect substitution across the different types of labor, these wage responses implies the presence of cross effects to keep profits nil. Hence, Equation 16 is not consistent with (13).

According to (16), a negative employment tax becomes optimal whenever the social welfare weight is higher than the ratio of micro over macro participation elasticity, instead of one without unemployment and wage responses. Therefore the introduction of unemployment and wage responses increases the likelihood of the optimal employment tax for the working poor to be negative whenever the macroeconomic participation elasticity is larger than the microeconomic one, a condition that can be easily tested.

A matching version of the no-cross effect economy

To better understand how the microeconomic and macroeconomic participation elasticities may differ, we now specialize the economy by considering a Diamond (1982), Mortensen and Pissarides (1999), Pissarides (2000) matching economy. On each labor market i, the constant-returns to scale matching function gives the employment level h_i as a function $\mathcal{M}_i(\nu_i, k_i)$ of the number ν_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, unlike in II.2. Each vacancy is matched with probability $q_i = Q_i(\theta_i) \stackrel{\text{def}}{\equiv} \frac{\mathcal{M}_i(\nu_i,k_i)}{\nu_i} = \mathcal{M}_i(1, 1/\theta_i)$, which is decreasing in tightness $\theta_i \stackrel{\text{def}}{\equiv} \nu_i/k_i$. Firms creates 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 $p_i = P(\theta_i) \stackrel{\text{def}}{\equiv} \frac{\mathcal{M}_i(\nu_i,k_i)}{k_i} = \mathcal{M}_i(\theta_i, 1)$. Therefore, the conditional probability p_i is a decreasing function of the gross wage through:

$$p_i = P_i\left(Q_i^{-1}\left(\frac{\kappa_i}{y_i - w_i}\right)\right) \tag{17}$$

In the matching model, the labor demand therefore gives the conditional employment probability as decreasing function of the gross wage independently of the number of job-seekers. Therefore a policy reform that increases the labor supply without affecting the gross wage leads to a rise in employment in the same proportion as the rise in the labor supply.

To get red rid of cross effects, we assume that workers are risk neutral (hence $u(c) \equiv c$) and that an exogenous share denoted $\beta_i \in (0,1)$ of the total surplus $y_i - T_i - d_i - b$ from a match benefits the workers as in Jacquet et al. (2014).[9]. This assumption leads to:

$$w_i = \mathscr{W}_i(T_i, b) \equiv \beta_i\, y_i + (1 - \beta_i)(T_i + d_i + b) \tag{18}$$

Combining (17), (18) and the assumption that labor supply responses are concentrated along the extensive margin provides a complete search and matching micro-foundation for the no-cross effect economy. Let $\mu_i \in (0,1)$ denote the elasticity of the matching function with respect to the

[9]Recall that d_i stands for the utility cost of working in occupation i

number of job-seekers. Appendix E shows that:

$$\frac{\partial \mathscr{U}_i}{\partial T_i} = \left[-1 + \frac{\partial \mathscr{W}_i}{\partial T_i}\left(1 + \frac{w_i}{p_i}\frac{\partial \mathscr{P}_i}{\partial w_i}\frac{w_i - T_i - d_i - b}{w_i}\right)\right] p_i = \frac{\beta_i}{\mu_i}\left.\frac{\partial \mathscr{U}_i}{\partial T_i}\right|_{\mathbf{w},\mathbf{p}} \tag{19}$$

where the second equality holds only when $\mu_i > 0$ and $\beta_i < 1$. Equation (19) therefore implies:

Lemma 2. *In the search and matching economy without cross effects, the microeconomic and macroeconomic participation responses are equal either when the workers have full bargaining power so there is no wage responses, or when the Hosios (1990) condition $\beta_i = \mu_i$ for a decentralized economy without tax and transfer to be socially efficient is verified. If $\beta_i < \mu_i$ the macroeconomic response is lower then microeconomic one. If $\mu_i < \beta_i < 1$ the macroeconomic response is larger then microeconomic one.*

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. This direct effect is the only one at the microeconomic level. Second, the rise in tax liability increases the gross wage which attenuates the direct effect at the macroeconomic level. Finally, the wage increases triggers a labor demand responses that amplifies the direct effect at the macroeconomic level. If the worker get all of the surplus (i.e. if $\beta_i = 1$), wages does not respond to taxation ($\frac{\partial \mathscr{W}_i}{\partial T_i} = 0$), the wage and the conditional employment probabilities are not affected so the microeconomic and macroeconomic responses to participation are identical. Otherwise, the conditional employment probability effect dominates the wage effect whenever the labor demand elasticity is sufficiently elastic, which happens when the matching elasticity is higher than the bargaining power. Proposition 3 and Lemma 2 implies 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. This result was also established in Jacquet et al. (2014).[10]

Theoretical Appendices

The Lagrangian associated to the government's program writes:

$$\Lambda(\mathbf{t}) \overset{\text{def}}{\equiv} \sum_{i=1}^{I}(T_i + b)\mathcal{H}_i(\mathbf{t}) - b - E + \frac{1}{\lambda}\Omega\left(\mathscr{U}_1(\mathbf{t}), ..., \mathscr{U}_I(\mathbf{t}), u(b)\right) \tag{20}$$

[10]Using (19), one has $\frac{\pi_j}{\pi_j^m} = \frac{\beta_j}{\mu_j}$, so Equation (16) becomes $\frac{T_j+b}{c_j-b} = \frac{1-\frac{\beta_j}{\mu_j}g_j}{\eta_j}$ which corresponds to the optimal tax formula (19b) in Jacquet et al. (2014).

A Derivation of Lemma 1 and Equation (10)

Equation (3) implies that: $\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}} = -\mathcal{A} \cdot \left.\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}}$, so Equation (9) leads to:

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

Moreover, from $\mathcal{K}_i(\mathbf{t}) = \hat{\mathcal{K}}_i(\mathscr{U}(\mathbf{t}))$, we get that

$$\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}} = \frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}} \cdot \frac{\mathbf{d}\hat{\mathcal{K}}}{\mathbf{dU}} \qquad \text{and} \qquad \left.\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}} = \left.\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}} \cdot \frac{\mathbf{d}\hat{\mathcal{K}}}{\mathbf{dU}}$$

So we get

$$\left(\left.\frac{\mathbf{d}\mathscr{U}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}}\right)^{-1} = \frac{\mathbf{d}\hat{\mathcal{K}}}{\mathbf{dU}} \cdot \left(\left.\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}}\right)^{-1}$$

Combining these three last equations leads to (10) whenever $\left.\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\right|_{\mathbf{w},\mathbf{p}}$ is invertible.

B Derivation of Equation (11)

Differentiating (20) with respect to b gives:

$$\frac{\partial \Lambda}{\partial b} = -1 + \sum_{i=1}^{I} h_i + \sum_{i=1}^{I} (T_i + b) \frac{\partial \mathcal{H}_i}{\partial b} + \frac{u'(b)}{\lambda} \frac{\partial \Omega}{\partial b} + \sum_{i=1}^{I} \frac{\partial \mathscr{U}_i}{\partial b} \frac{\partial \Omega}{\partial U_i}$$

Differentiating (2) with respect to b gives:

$$\frac{\partial \mathscr{U}_i}{\partial b} = (1 - p_i) u'(b) + p_i u'(c_i) \left[\frac{\partial \mathscr{C}_i}{\partial b} + \frac{1}{p_i} \frac{\partial \mathscr{P}_i}{\partial b} \frac{u(c_i) - d_i - b}{u'(c_i)}\right]$$

Using $h_0 = 1 - \sum_{i=1}^{I} h_i$ and Equations (7) and (12) leads to (11).

From $\frac{\partial \mathscr{C}_i}{\partial T_i} = \frac{\partial \mathscr{W}_i}{\partial T_i} - 1$ and for $j \neq i$, $\frac{\partial \mathscr{C}_i}{\partial T_i} = \frac{\partial \mathscr{W}_i}{\partial T_i}$, the sum of (8) for all T_j minus Equation (11) leads to:

$$\begin{aligned} 0 \quad = & \quad \sum_{i=1}^{0} h_i + \sum_{i=1}^{I} (T_i + b) \left(\sum_{j=1}^{I} \frac{\partial \mathcal{H}_i}{\partial T_j} - \frac{\partial \mathcal{H}_i}{\partial b}\right) - \left(g_0 h_0 + \sum_{i=1}^{I} g_i h_i\right) \\ + & \quad \sum_{i=1}^{I} g_i h_i \left(\sum_{j=1}^{I} \frac{\partial \mathscr{W}_i}{\partial T_j} - \frac{\partial \mathscr{W}_i}{\partial b}\right) + \sum_{i=1}^{I} g_i h_i \frac{u(c_i) - d_i - u(b)}{u'(c_i)} \left(\sum_{j=1}^{I} \frac{\partial \mathscr{P}_i}{\partial T_j} - \frac{\partial \mathscr{P}_i}{\partial b}\right) \end{aligned} \tag{21}$$

In the absence of income effects, a simultaneous change in all tax liabilities and welfare benefit $\Delta T_1 = ... = \Delta T_i = -\Delta b$ induces no changes in wages, conditional employment probabilities not employment levels, so that $\sum_{i=1}^{I} \frac{\partial \mathscr{W}_i}{\partial T_i} = \frac{\partial \mathscr{W}_i}{\partial b}$, $\sum_{i=1}^{I} \frac{\partial \mathscr{P}_i}{\partial T_i} = \frac{\partial \mathscr{P}_i}{\partial b}$ and $\sum_{i=1}^{I} \frac{\partial \mathcal{H}_i}{\partial T_i} = \frac{\partial \mathcal{H}_i}{\partial b}$ plugging these equalities in (21) leads to: $g_0 h_0 + \sum_{i=1}^{I} g_i h_i = 1$.

C The case without unemployment but with imperfect factor substitutability

In the absence of unemployment responses to taxation, one has $\frac{\partial \mathscr{P}_i}{\partial T_j} = 0$. The matrix $\mathcal{A}$ of corrective terms $\frac{\partial \mathscr{C}_i}{\partial T_j} + \frac{\partial \mathscr{P}_i}{\partial T_j} \frac{u(c_i) - d_i - u(b)}{p_i \, u'(c_i)}$ thus coincides with $\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}$. Moreover, we get: $\frac{\mathbf{dK}}{\mathbf{dT}} = -\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\Big|_{\mathbf{p,w}}$ and $\frac{\mathbf{dH}}{\mathbf{dT}} = -\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}}\Big|_{\mathbf{p,w}}$. Equation (10) becomes:

$$\begin{aligned}
\mathbf{0} \quad &= \mathbf{h} - \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}}\Big|_{\mathbf{p,w}} \cdot (\mathbf{T} + \mathbf{b}) + \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\Big|_{\mathbf{p,w}} \cdot \left(\frac{\mathbf{d}\mathcal{K}}{\mathbf{dT}}\Big|_{\mathbf{p,w}} \right)^{\mathbf{-1}} \cdot (\mathbf{g}\,\mathbf{h}) \\
&= \mathbf{h} - \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot \frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}}\Big|_{\mathbf{p,w}} \cdot (\mathbf{T} + \mathbf{b}) + \frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \cdot (\mathbf{g}\,\mathbf{h}) \\
&= \left(\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}} \right)^{\mathbf{-1}} \cdot \mathbf{h} - \frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}}\Big|_{\mathbf{p,w}} \cdot (\mathbf{T} + \mathbf{b}) + \mathbf{g}\,\mathbf{h}
\end{aligned} \tag{22}$$

Moreover, the firm's profit-maximization program defined the profit function through:

$$\Pi(w_1, ..., w_I, r) \stackrel{\text{def}}{\equiv} \max_{h_1,...,h_I,Z} \quad F(h_1, ..., h_I, Z) - \sum_{i=1}^{I} w_i\, h_i - r\, Z$$

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 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. It therefore also an eigenvector of Matrix $\left(\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}\right)^{-1}$ associated to eigenvalue -1, i.e $\left(\frac{\mathbf{d}\mathscr{C}}{\mathbf{dT}}\right)^{\mathbf{-1}} \cdot \mathbf{h} = -\mathbf{h}$. Therefore Equation (22) simplifies to:

$$\mathbf{0} = \mathbf{1} - \mathbf{g}\,\mathbf{h} + \frac{\mathbf{d}\mathcal{H}}{\mathbf{dT}}\Big|_{\mathbf{p,w}} \cdot (\mathbf{T} + \mathbf{b})$$

which corresponds to (13).

D The competitive model under mixed responses

Under full employment and labor supply responses being concentrated along the intensive and extensive margins, we get that Equation (13) simplifies to:

$$0 = (1 - g_j) h_j + (T_{j-1} + b) \frac{\partial \mathcal{H}_{j-1}}{\partial T_j} + (T_j + b) \frac{\partial \mathcal{H}_j}{\partial T_j} + (T_{j+1} + b) \frac{\partial \mathcal{H}_{j+1}}{\partial T_j}$$

Using $\mathcal{H}_i(T_{i+1} - T_i, T_i + b, T_i - T_{i-1}) \equiv \widetilde{\mathcal{H}}_i\,(T_{i+1} - T_i, T_i + b, T_i - T_{i-1})$ leads to:

$$\begin{aligned}
0 \quad = \quad & (1 - g_j)\, h_j + (T_{j-1} + b) \frac{\partial \widetilde{\mathcal{H}}_{j-1}}{\partial T_j - T_{j-1}} + (T_j + b) \left(\frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_j - T_{j-1}} + \frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_j + b} - \frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_{j+1} - T_j} \right) \\
& - (T_{j+1} + b) \frac{\partial \widetilde{\mathcal{H}}_{j+1}}{\partial T_{j+1} - T_j}
\end{aligned} \tag{23}$$

Absent behavioral response along the intensive margin, one would have the optimal employment tax would be: $\overline{T_j+b} = \frac{(1-g_j)h_j}{\frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_j+b}}$. Let us then denote

$$x_j \stackrel{\text{def}}{\equiv} \left(1-g_j\right) h_j + (T_j+b)\frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_j+b} = \left(\overline{T_j+b} - (T_j+b)\right)\frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_j+b} \tag{24}$$

From $\frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_j - T_{j-1}} = \frac{\partial \widetilde{\mathcal{H}}_{j-1}}{\partial T_j - T_{j-1}}$, the first-order condition (23) simplifies to:

$$0 = x_j - \left(T_{j-1} - T_j\right)\frac{\partial \widetilde{\mathcal{H}}_{j-1}}{\partial T_j - T_{j-1}} + \left(T_{j+1} - T_j\right)\frac{\partial \widetilde{\mathcal{H}}_j}{\partial T_{j+1} - T_j} \tag{25}$$

Summing (25) for all occupations leads to $\sum_{i=1}^{I} x_i = 0$, as $\frac{\partial \widetilde{\mathcal{H}}_I}{\partial T_{I+1} - T_I} = \frac{\partial \widetilde{\mathcal{H}}_0}{\partial T_1 - T_0} = 0$, the latter equality because a change in $T_1 - T_0 = T_1 + b$ corresponds to an extensive response and is already encapsulated in $\frac{\partial \widetilde{\mathcal{H}}_1}{\partial T_1+b} = -\frac{\partial \widetilde{\mathcal{H}}_0}{\partial T_1+b}$. Summing (25) for all occupations $i \in \{j, ..., I+1\}$, leads to:

$$\left(T_{j-1} - T_j\right)\frac{\partial \widetilde{\mathcal{H}}_{j-1}}{\partial T_j - T_{j-1}} = \sum_{i=j}^{I} x_i = -\sum_{i=1}^{j-1} x_i \tag{26}$$

which leads to (15)

We now show that if $\overline{T_j+b} < \overline{T_{j-1}+b}$ for all $j \in \{2, ..., I\}$, then $\sum_{i=j}^{I} x_i > 0$ for all $j \in \{2, ..., I\}$.

1. Assume by contradiction that for some j, one has $\sum_{i=j}^{I} x_i \leq 0$ and $x_{j-1} \geq 0$. Then, according to (26), one has $T_j \leq T_{j-1}$. Combining this inequality with (24) and the assumption that $\overline{T}_j > \overline{T}_{j-1}$, one has $T_j \leq T_{j-1} \leq \overline{T}_{j-1} < \overline{T}_j$. Hence $x_j > 0$, in which case $\sum_{i=j+1}^{I} x_i < 0$.

   We therefore get that $x_{j-1} \geq 0$ and $\sum_{i=j}^{I} x_i \leq 0$ leads to $x_j > 0$ and $\sum_{i=j+1}^{I} x_i < 0$. Iterating this reasoning for increasing j leads for $j = I$ to $x_I > 0$ and $\sum_{i=I}^{I} x_i = x_I < 0$, a contradiction.

2. Assume by contradiction that for some j, one has $\sum_{i=j}^{I} x_i \leq 0$ and $x_{j-1} < 0$. Then $\sum_{i=j-1}^{I} x_i \leq 0$. Moreover, from (26), we get that $T_{j-2} > T_{j-1}$. Combining this inequality with (24) and the assumption that $\overline{T}_{j-1} > \overline{T}_{j-2}$, we get that $T_{j-2} > T_{j-1} > \overline{T}_{j-1} > \overline{T}_{j-2}$ and hence x_{j-2}.

   We therefore get that $x_{j-1} < 0$ and $\sum_{i=j}^{I} x_i \leq 0$ leads to $x_{j-2} < 0$ and $\sum_{i=j-1}^{I} x_i < 0$. Iterating this reasoning for decreasing j leads for $j = 1$ to $\sum_{i=1}^{I} x_i < 0$, a contradiction.

We therefore get that $T_j > T_{j-1}$ for all $j \in \{2, ..., I\}$. Moreover, as $0 = \sum_{i=1}^{I} x_i < \sum_{i=2}^{I} x_i$, we must get that $x_1 < 0$, thereby from (24), $T_1 + b > \overline{T_1+b}$.

E Proof of Lemma 2 and Equation (19)

In the no-cross effect economy, one has $\frac{\partial \mathscr{U}_i}{\partial T_j} = 0$ whenever $j \neq i$. Moreover, from $p_i = \mathscr{L}_i(w_i)$, we get from (3) the first equality in (19). 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: $\frac{w_i}{p_i}\frac{\partial \mathscr{P}_i}{\partial w_i} = -\frac{1-\mu_i}{\mu_i}\frac{w_i}{y_i-w_i}$. Hence,

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

Equation (18) implying that $\frac{w_i - T_i - d_i - b}{y_i - w_i} = \frac{\beta_i}{1-\beta_i}$ and $\frac{\partial \mathscr{W}_i}{\partial T_i} = 1 - \beta_i$, we get:

$$\frac{\partial \mathscr{U}_i}{\partial T_i} = \left[-1 + (1-\beta_i)\left(1 - \frac{1-\mu_i}{\mu_i}\frac{\beta_i}{1-\beta_i}\right)\right] p_i$$

which leads to (19) as $\left.\frac{\partial \mathscr{U}_i}{\partial T_i}\right|_{\mathbf{w},\mathbf{p}} = -p_i$ under risk neutrality.

References

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics 1*(1), 451–488.

Choné, P. and G. Laroque (2005). Optimal incentives for labor force participation. *Journal of Public Economics 89*(2-3), 395–425.

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

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

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

Diamond, P. and J. Mirrlees (1971). Optimal taxation and public production. *American Economic Review 61*, 8–27 also 261–278.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies 57*(2), 279–98.

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

Jacquet, L., E. Lehmann, and B. Van der Linden (2014). Optimal income taxation with kalai wage bargaining and endogenous participation. *Social Choice and Welfare 42*(2), 381–402.

Landais, C., P. Michaillat, and E. Saez (2014). Optimal unemployment insurance over the business cycle. Working Paper 16526, National Bureau of Economic Research.

Lee, D. and E. Saez (2008). Optimal minimum wage policy in competitive labor markets. *NBER Working Papers 14 320*.

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

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

Mortensen, D. T. and C. A. Pissarides (1999). New developments in models of search in the labor market. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 39, pp. 2567–2627. Elsevier.

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

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory, 2nd Edition*, Volume 1 of *MIT Press Books*. The MIT Press.

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

Rothstein, J. (2010). Is the eitc as good as an nit? conditionnal tax transfers and tax incidence. *American Economic Journal: Economic Policy 2*(1), 177–208.

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

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

Saez, E. (2004). Direct or indirect tax instruments for redistribution: short-run versus long-run. *Journal of Public Economics 88*, 503–518.

Stiglitz, J. E. (1982). Self-selection and pareto efficient taxation. *Journal of Public Economics 17*(2), 213–240.