NBER WORKING PAPER SERIES

SHOULD ROBOTS BE TAXED?

Joao Guerreiro
Sergio Rebelo
Pedro Teles

Working Paper 23806
http://www.nber.org/papers/w23806

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2017

We thank V.V. Chari for very useful discussions. Teles thanks the support of FCT as well as the ADEMU project, "A Dynamic Economic and Monetary Union", funded by the European Union's Horizon 2020 Program under grant agreement N° 649396. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Joao Guerreiro, Sergio Rebelo, and Pedro Teles. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Should Robots be Taxed?
Joao Guerreiro, Sergio Rebelo, and Pedro Teles
NBER Working Paper No. 23806
September 2017
JEL No. H21,O33

ABSTRACT

We use a model of automation to show that with the current U.S. tax system, a fall in automation costs could lead to a massive rise in income inequality. This inequality can be reduced by raising marginal income tax rates and taxing robots. But this solution involves a substantial efficiency loss for the reduced level of inequality. A Mirrleesian optimal income tax can reduce inequality at a smaller efficiency cost, but is difficult to implement. An alternative approach is to amend the current tax system to include a lump-sum rebate. In our model, with the rebate in place, it is optimal to tax robots only when there is partial automation.

Joao Guerreiro
Department of Economics
Northwestern University
Evanston, IL 60208
jguerreiro@u.northwestern.edu

Sergio Rebelo
Northwestern University
Kellogg School of Management
Department of Finance
Leverone Hall
Evanston, IL 60208-2001 and
CEPR
and also NBER
s-rebelo@northwestern.edu

Pedro Teles
Banco de Portugal
R. Francisco Ribeiro 2
1150 Lisboa
Portugal
and Univ Catolica Portuguesa and CEPR
pteles@ucp.pt

1 Introduction

The American writer Kurt Vonnegut began his career in the public relations division of General Electric. One day, he saw a new milling machine operated by a punch-card computer outperform the company's best machinists. This experience inspired him to write a novel called "Player's Piano." It describes a world in which school children take a test at an early age that determines their fate. Those who pass, become engineers and design robots used in production. Those who fail, have no jobs and live from government transfers. Are we converging to this dystopian world? How should public policy respond to the impact of automation on the demand for labor?

These questions have been debated ever since 19th-century textile workers in the U.K. smashed the machines that eliminated their jobs. As the pace of automation quickens, Bill Gates re-ignited this debate by proposing that robots should be taxed.

In this paper, we use a simple model of automation to compare the equilibrium that emerges under the current U.S. tax system (which we call the status quo), the first-best solution to a planner's problem without information constraints, and the second-best solutions associated with different configurations of the tax system.

Our model has two types of workers which we call routine and non-routine. Routine workers perform tasks that can be automated. We find that robot taxes are optimal only when there is partial automation. These taxes help increase the wages of routine workers, giving the government an additional instrument to reduce income inequality. Once there is full automation, it is not optimal to tax robots. Routine workers do not work, so taxing robots distorts production decisions without reducing income inequality.[1]

Under the current U.S. tax system, modeled using the after-tax income function

[1]These results show that the reason why it can be optimal to tax robots in our model differs from the rationale used by Bill Gates to motivate robot taxation. Gates argued that robots should be taxed to replace the tax revenue that the government collected from routine workers before their jobs were automated. In our model, when there is full automation the government collects no tax revenue from routine workers yet it is optimal not to tax robots.

estimated by Heathcote, Storesletten and Violante (2014), full automation never occurs. As the cost of automation falls, the wages of non-routine workers rise while the wages of routine workers fall to make them competitive with robot use. The result is a large rise in income inequality and a substantial decline in the welfare of routine workers.

The level of social welfare obtained in the status quo is much worse than that achieved in the first-best solution to an utilitarian social planner problem without information constraints. But this first-best solution cannot be implemented when the government does not observe the worker type. The reason is that the two types of agents receive the same consumption but non-routine workers supply more labor than routine workers. As a result, non-routine workers have an incentive to act as routine workers and receive their bundle of consumption and hours worked.

To circumvent this problem, we solve for the optimal tax system imposing, as in Mirrlees (1971), the constraint that the government does not observe the worker type or the workers' labor input. The government can observe total income and consumption of the two types of workers, as well as the use of robots by firms. We assume that taxes on robots are linear for the reasons emphasized in Guesnerie (1995): non-linear taxes on intermediate inputs are difficult to implement in practice because they create arbitrage opportunities. This assumption, which is standard in a Ramsey (1927) setting, restricts the outcomes that can be achieved when robot taxation is optimal.

A Mirrleesian optimal tax system can improve welfare relative to the status quo. In fact, it can yield a level of welfare that is close to that of the first-best solution. Unfortunately, Mirrleesian tax systems are known to be complex and potentially difficult to implement in practice.

For this reason, we study the optimal policy when the tax schedule is constrained to take a simple, exogenous form. Specifically, we consider the income tax schedule proposed by Heathcote, Storesletten and Violante (2014) and linear robot taxes. We compute the parameters of the income tax function and the robot tax rate that maximize social welfare. We find that income inequality can be reduced by raising marginal tax rates and taxing robots. Tax rates on robot use can be as high as 30 percent and full automation never occurs, so routine workers keep their jobs. But this solution yields poor outcomes in terms of efficiency and distribution.

We consider a modification of the Heathcote, Storesletten and Violante (2014) tax schedule that allows for lump-sum rebates that ensure that all workers receive a minimum income. We find that this modification improves both efficiency and distribution relative to a tax system without rebates.

In the three best systems in terms of welfare, the first-best, Mirrleesian optimal taxes and simple income taxes with lump-sum rebates there is full automation once the costs of automation are sufficiently low. These solutions resemble the world of "Player's Piano." Only non-routine workers have jobs. Routine workers live off government transfers and, despite losing their jobs, are better off than in the status quo.

One might expect that optimal robot taxation would follow from well-known principles of optimal taxation in the public finance literature. We know from the intermediate-goods theorem of Diamond and Mirrlees (1971) that it is not optimal to distort production decisions by taxing intermediate goods. Since robots are in essence an intermediate good, taxing them should not be optimal.

The intermediate-good theorem relies on the assumption that "net trades" of different goods can be taxed at different rates. In our context, this assumption implies that the government can use different tax schedules for routine and non-routine workers. We study two environments where there are limits to the government's ability to tax different workers at different rates, Mirrlees (1971)-type information constraints and a simple exogenous tax system common to both types of workers. We find that it is optimal to tax robots in both environments when there is partial automation.

We know from the work of Atkinson and Stiglitz (1976) that when the income tax system is non-linear it is not optimal to distort production decisions by taxing intermediate goods. But, as stressed by Jacobs (2015), Atkinson and Stiglitz (1976)'s result depends critically on the assumption that workers with different productivities are perfect substitutes in production. This assumption does not hold in our model. Taxing robots can be optimal because it loosens the incentive compatibility constraint of non-routine workers.

The paper is organized as follows. In Section 2 we describe our model of automation. Section 3 describes the status-quo equilibrium, i.e. the equilibrium under the current U.S. income tax system and no robot taxes. Section 4 describes the first-best solution to the problem of an utilitarian planner. In Section 5, we analyze a Mirrleesian second-best solution to the planner's problem. In Section 6, we study numerically the optimal tax system that emerges when income taxes are constrained to take the functional form proposed by Heathcote, Storesletten and Violante (2014) both with and without lump-sum rebates. In Section 7, we compare the different policies we consider both in terms of social welfare and of the utility of different agents. Section 8 relates our results to classical results on production efficiency in the public finance literature. Section 9 concludes.

2 Model

In this section, we discuss a simple model of automation that allows us to address the optimal tax policy questions posed in the introduction. The model has two representative households who draw utility from consumption of private and public goods and disutility from labor. One household supplies routine labor and the other non-routine labor. The consumption good is produced with non-routine labor, routine labor and robots. Robots and routine labor are used in a continuum of tasks. They are both perfect substitutes in performing these tasks.

Households There are two representative households, one composed of non-routine workers and the other of routine workers. The index $j = n, r$, denotes non-routine and routine labor, respectively.

Household j derives utility from consumption, C_j, and from the provision of a public good, G. Each household has one unit of time and derives disutility from labor (N_j). The household's utility function is given by

$$U_j = u(C_j) - v(N_j) + g(G). \tag{1}$$

The functions $u(C_j)$ and $g(G)$ are differentiable, strictly increasing and concave. The function $v(N_j)$ is differentiable, strictly increasing and convex. In order to guarantee that the optimal choices of consumption and leisure are interior solutions, we assume that the following conditions hold: $\lim_{C_j \to 0} u'(C_j) = \infty$ and $\lim_{N_j \to 1} v'(N_j) = \infty$.

Household j chooses C_j and N_j to maximize utility (1), subject to the budget constraint

$$C_j \leq w_j N_j - T(w_j N_j),$$

where w_j denotes the wage rate received by the household type j and $T(\cdot)$ denotes the tax schedule.

We now describe the problem of the firms, starting with the firms that produce robots.

Robot producers Robots can be used in $i \in [0,1]$ tasks. The cost of producing a robot is the same across tasks and is equal to ϕ units of output. A representative firm producing robots to automate task i maximizes profits

$$\pi_i = \max_{x_i} p_i x_i - \phi x_i.$$

It follows that

$$p_i = \phi,$$

and that $\pi_i = 0$.

Final good producers The representative producer of final-goods hires non-routine labor (N_n), routine labor (n_i) for each task i, and buys intermediate goods (x_i) which we refer to as robots, also for each task i. There is a continuum of tasks that can be performed by either routine labor or robots. We denote by m the fraction of these tasks that are automated, i.e. performed by robots. For convenience, we assume that when m tasks from the continuum $[0, 1]$ are automated, the automated tasks are in the interval $[0, m]$.[2]

The production function is given by

$$Y = A \left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di \right]^{\frac{1-\alpha}{\rho}} N_n^\alpha, \tag{2}$$

with $\alpha \in (0, 1)$ and $\rho \in (0, 1)$.

The problem of the firm is to maximize profits,

$$\pi = Y - w_n N_n - w_r \int_m^1 n_i di - (1 + \tau_x) \int_0^m \phi x_i di,$$

where Y is given by (2), and τ_x is an ad-valorem tax rate on intermediate goods.

The optimal choices of N_n, x_i for $i \in [0, m]$, n_i for $i \in (m, 1]$ require that

$$w_n = \alpha A \left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di \right]^{\frac{1-\alpha}{\rho}} N_n^{\alpha-1}, \tag{3}$$

$$(1 + \tau_x)\phi = (1 - \alpha) A \left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di \right]^{\frac{1-\alpha}{\rho}-1} N_n^\alpha x_i^{\rho-1}, \text{ for } i \in [0, m] \tag{4}$$

$$w_r = (1 - \alpha) A \left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di \right]^{\frac{1-\alpha}{\rho}-1} N_n^\alpha n_i^{\rho-1}, \text{ for } i \in (m, 1] \tag{5}$$

are satisfied. It follows that it is optimal to use the same level of routine labor, n_i in the $1 - m$ tasks that have not been automated and that the optimal use of robots is also the same in the m automated tasks.

´The optimal level of automation is $m = 0$ if $w_r < (1 + \tau_x)\phi$. The firm chooses to fully automate ($m = 1$) and to employ no routine workers ($n_i = 0$) if $w_r > (1 + \tau_x)\phi$.

[2]Since tasks are symmetric, there is no loss of generality associated with this assumption.

If $w_r = (1+\tau_x)\phi$, the firm is indifferent between any level of automation $m \in [0,1]$. In this case, equations (4) and (5) imply that the levels of routine labor and robots are the same,

$$x_i = n_l, \text{ for } i \in [0,m] \text{ and } l \in (m,1].$$

Since the technology has constant returns to scale, profits are zero, $\pi = 0$.

Government The government chooses taxes and the optimal level of government spending, subject to the budget constraint

$$G \leq T(w_r N_r) + T(w_n N_n) + \int_0^m \tau_x \phi x_i di. \tag{6}$$

Equilibrium An equilibrium is a set of allocations $\{C_r, N_r, C_n, N_n, G, n_i, x_i, m\}$, prices $\{w_r, w_n, p_i\}$ and a tax system $\{T(\cdot), \tau_x\}$ such that: (i) given prices and taxes, allocations solve each households' problem; (ii) given prices and taxes, allocations solve each firm's problem; (iii) the government budget constraint is satisfied; and (iv) markets clear.

The market clearing condition for routine labor is

$$\int_m^1 n_i di = N_r. \tag{7}$$

The market-clearing condition for the output market is

$$C_r + C_n + G \leq A\left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di\right]^{\frac{1-\alpha}{\rho}} N_n^\alpha - \int_0^m \phi x_i di. \tag{8}$$

The market clearing condition (7), and firms first-order condition (5) imply

$$n_i = \frac{N_r}{1-m}, \text{ for } i \in (m,1].$$

In an equilibrium with automation ($m > 0$) in which $w_r = (1+\tau_x)\phi$, we also have

$$x_i = \frac{N_r}{1-m}, \text{ for } i \in [0,m]$$

In such an equilibrium, using equation (4) together with the two previous conditions, we obtain

$$m = 1 - \left[\frac{(1+\tau_x)\phi}{(1-\alpha)A}\right]^{1/\alpha} \frac{N_n}{N_r}. \tag{9}$$

The wage rates of both non-routine and routine labor are independent of preferences,

$$w_n = \alpha \frac{A^{1/\alpha}(1-\alpha)^{\frac{1-\alpha}{\alpha}}}{[(1+\tau_x)\phi]^{\frac{1-\alpha}{\alpha}}}, \tag{10}$$

$$w_r = (1+\tau_x)\phi. \tag{11}$$

The reason for this result is twofold. In an equilibrium with automation, the real wage of routine workers must equal the cost of using robots, $(1+\tau_x)\phi$. Since the production function is constant returns to scale, optimizing the use of robots and routine workers yields a production that is linear in non-routine labor. So, the marginal product of non-routine labor is constant and independent of the number of non-routine labor hours employed by the firm.

It is also useful to note that in any equilibrium

$$\frac{w_r N_r}{w_n N_n} = \frac{(1-\alpha)(1-m)}{\alpha}. \tag{12}$$

An increase in automation reduces the income share of routine workers while keeping constant the share of non-routine workers. In this sense, an increase in automation leads to an increase in pre-tax income inequality.

3 The status-quo equilibrium

In this section, we describe the status-quo equilibrium, i.e. the equilibrium under the current U.S. income tax system and no taxes on robot use ($\tau_x = 0$). We model the U.S. income tax system using the empirically-plausible functional form for after-tax income proposed by Heathcote, Storesletten and Violante (2014). In their specification, the after-tax income of household j is given by

$$y(w_j N_j) = \lambda (w_j N_j)^{1-\gamma}, \tag{13}$$

where $\gamma < 1$. Using PSID data, Heathcote, Storesletten and Violante (2014) estimate that $\gamma = 0.185$, which means that income taxes are close to linear. They find that their specification yields a good fit to the data with an R^2 of 0.92.

This formulation implies that total and average taxes paid by agent j are given by

$$T(w_jN_j) = w_jN_j - \lambda(w_jN_j)^{1-\gamma},$$
$$\frac{T(w_jN_j)}{w_jN_j} = 1 - \lambda(w_jN_j)^{-\gamma}.$$

The parameter λ controls the level of taxation, higher values of λ imply lower average taxes. The parameter γ controls the progressivity of the tax code. When γ is positive (negative), the average tax rate rises (falls) with income, so the tax system is progressive (regressive).

We assume in all our numerical work that the utility function takes the form:

$$u(C_j) - v(N_j) + g(G) = \log(C_j) + \theta \log(1 - N_j) + \chi \log(G). \tag{14}$$

We set $\theta = 1.63$, so that in the status-quo equilibrium agents choose to work one third of their time endowment. We choose $\chi = 0.25$ and assume that the government sets government spending equal to 20 percent of output and adjusts λ to satisfy its budget constraint. Given our choice for χ, this policy is optimal in the sense that it maximizes the average utility of the two workers in the economy. On the production side, we normalize A to one and choose $\alpha = 0.5$, so that the wages of routine and non-routine workers are identical in the absence of automation. These parameter choices are used in all of our numerical experiments.

We vary ϕ, the cost of producing robots on the interval $\Phi = (0; (1-\alpha)A]$. The upper bound of the interval, $(1-\alpha)A$, is the lowest value of ϕ consistent with no automation in the status quo (see equation (9)).

Figure 1 describes the effect of changes in the cost of automation. As ϕ falls, the wage of routine workers fall and the wage of non-routine workers rises. Since the utility function is logarithmic and wages are the only source of income, hours worked remain constant for both routine and non-routine workers. This property reflects the offsetting nature of income and substitution effects. Given that as ϕ falls, wages of routine workers fall and their hours worked remain constant, their income, consumption and utility fall. In contrast, non-routine workers benefit from rising income, consumption and utility. As ϕ falls, the parameter that controls the level of taxation, λ, rises, which implies a decline in the overall level of taxation. This decline is due to the fact that non-routine workers pay an increasing share of the tax revenue and their income rises faster than output as ϕ falls.

In sum, our analysis suggests that the current U.S. tax system will lead to massive income and welfare inequality in response to a fall in the costs of automation.

4 The first-best allocation

A first-best allocation solves a problem where the social planner maximizes a social-welfare function subject to the economy's resource constraints. In this problem, there are no restrictions on the ability of the planner to transfer income across agents.

We assume that the planner has an utilitarian social welfare function which assigns equal weights to the utilities of the individual agents. The planner chooses $C_r, N_r, C_n, N_n, G, m, \{x_i, n_i\}$ to maximize average utility,

$$V = \frac{1}{2}\left[u(C_r) - v(N_r) + g(G)\right] + \frac{1}{2}\left[u(C_n) - v(N_n) + g(G)\right].$$

One interpretation of the social welfare function is as follows. Workers are identical ex-ante because they do not know whether their skills can be automated or not, i.e. whether they will be routine or non-routine workers. The planner maximizes the worker's ex-ante expected utility.

The planner's resource constraints are

$$C_r + C_n + G \leq Y - \phi \int_0^m x_i di, \tag{15}$$

$$Y = A \left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di \right]^{\frac{1-\alpha}{\rho}} N_n^\alpha, \tag{16}$$

$$\int_m^1 n_i di = N_n. \tag{17}$$

We derive in the appendix the solution to this problem. The first-best solution shares three properties with the competitive equilibrium. First, routine labor is evenly allocated to the tasks that have not been automated, $n_i = N_r/(1-m)$, for $i \in (m, 1]$ and is zero otherwise. Second, intermediate goods are evenly allocated to the activities that have been automated. Third, for $m \in (0,1)$ the level of intermediate goods used in each automated activity is the same as the amount of labor used in non-automated activities, $x_i = N_r/(1-m)$ for $i \in [0, m]$.

The expression for the level of automation that occurs in the first best is the same as in the status-quo equilibrium,

$$m = \max \left\{ 1 - \left[\frac{\phi}{(1-\alpha)A} \right]^{1/\alpha} \frac{N_r}{N_n}, 0 \right\}, \tag{18}$$

because in both cases robot use is not taxed.

Labor supply decisions in the first best are determined by the following conditions

$$\frac{v'(N_n)}{u'(C_n)} = \frac{\alpha Y}{N_n}, \tag{19}$$

$$\frac{v'(N_r)}{u'(C_r)} \geq \frac{(1-m)(1-\alpha)Y}{N_r}, \tag{20}$$

where $\alpha Y/N_n$ is the marginal product of non-routine workers and $(1-m)(1-\alpha)Y/N_r$ is the marginal product of routine workers. These conditions hold in a competitive equilibrium with zero marginal income taxes.

It is optimal to equate the marginal utility of consumption of the two types of workers. Since utility is separable in consumption, achieving that goal requires equalizing the consumption of the two households

$$C_r = C_n.$$

Finally, the optimal provision of public goods equates their marginal utility to their opportunity cost in terms of private consumption,

$$g'(G) = \frac{1}{2}u'(C_n).$$

Equations (19), (20), together with the fact that $C_r = C_n$, imply that the worker with the highest marginal product works the most. This property implies that non-routine workers have lower utility than routine workers. As a result, this equilibrium cannot be implemented when the government observes total income but does not observe agent types. The reason is that non-routine workers would choose to act as routine workers in order to receive higher utility. In the next section, we analyze the optimal tax system taking into account the information constraints that the worker type and the labor input are not directly observed by the government.

Figure 2 illustrates the properties of the first-best. In panel A, we see that full automation occurs once ϕ falls below 0.31. The real wage rate for both types of workers are the same as in the status-quo equilibrium.[3] The consumption and utility of both agents rise as ϕ falls. Figure 2 also shows that implementing the first-best solution requires large transfers from non-routine to routine workers.

[3]The reason for this property is as follows. Equations (10) and (11) imply that wages depend on technological parameters (α and A), the cost of automation, and the value of τ_x. Since $\tau_x = 0$ in the status quo and there is production efficiency in the first-best allocation, the wages are the same in both allocations.

5 Mirrleesian optimal taxation

In this section, we characterize the optimal non-linear income tax in the presence of automation, when the government observes an agent's total income but does not observe the agent's type or labor supply, as in the canonical Mirrlees (1971) problem. As discussed in the introduction, we focus on the case where robot taxes are linear.

In the Mirrlees (1971) model, the productivities of the different agents are exogenous. In our model, these productivities are endogenous and depend on τ_x. This property is central to the question we are interested in studying: is it optimal to distort production decisions by taxing the use of robots to redistribute income from non-routine to routine workers to increase social welfare?

We assume that $\phi < \alpha^\alpha(1-\alpha)^{1-\alpha}A$, so that if $\tau_x \leq 0$ non-routine workers earn a higher wage, $w_n > w_r$ in an equilibrium with automation (see equations (10) and (11)). Note that an increase in τ_x raises the wage of routine workers and lowers the wage rate of non-routine agents.

When working with Mirrleesian-style optimal taxation problems, it is useful to express the problem in terms of the income instead of hours worked. The income earned by agent j is

$$Y_j = w_j N_j. \tag{21}$$

The planner's problem is to choose the allocations $\{Y_j, C_j\}$, G, and τ_x to maximize social welfare

$$V = \frac{1}{2}\left[u(C_r) - v\left(Y_r/w_r\right) + g(G)\right] + \frac{1}{2}\left[u(C_n) - v\left(Y_n/w_n\right) + g(G)\right], \tag{22}$$

subject to the resource constraint

$$C_r + C_n + G \leq Y_n \frac{\tau_x + \alpha}{\alpha(1+\tau_x)} + \frac{Y_r}{1+\tau_x}. \tag{23}$$

and two incentive compatibility (IC) constraints

$$u(C_n) - v\left(Y_n/w_n\right) \geq u(C_r) - v\left(Y_r/w_n\right), \tag{24}$$

$$u(C_r) - v\left(Y_r/w_r\right) \geq u(C_n) - v\left(Y_n/w_r\right), \tag{25}$$

The wages of the two types of workers are dictated by production and are given by equations (10) and (11).

Any competitive equilibrium satisfies equations (23), (24), and (25). In addition, any allocation that satisfies these three equations can be decentralized as a competitive equilibrium.

Household optimality implies that the utility associated with the bundle of consumption and income assigned to agent j, $\{C_j, Y_j\}$, must be at least as high as the utility associated with any other bundle $\{C, Y\}$ that satisfies the budget constraint $C \leq Y - T(Y)$:

$$u(C_j) - v(Y_j/w_j) \geq u(C) + v(Y/w_j), \tag{26}$$

In particular, routine workers must prefer their bundle, $\{C_r, Y_r\}$, to the bundle of non-routine workers, $\{C_n, Y_n\}$. Similarly, non-routine workers must prefer their bundle, $\{C_n, Y_n\}$, to the bundle of routine workers, $\{C_r, Y_r\}$. These requirements correspond to the two IC constraints, (24), and (25), so these conditions are necessary.

We show in the Appendix that equation (23) is necessary by combining the first-order conditions to the firms' problems with the resource constraint, (19). In addition, we show that conditions (24), (25), and (23) are also sufficient. To see that (24) and (25) summarize the household problem note that it is possible to choose a tax function such that agents prefer the bundle C_j, Y_j no any other bundle. For example, the government could choose a tax function that sets the agent's after-tax income to zero for any choice of Y different from Y_j, $j = r, n$. These results are summarized in the following proposition.

Proposition 1. *Equations* (23), (25) *and* (24) *characterize the set of implementable allocations. These conditions are necessary and sufficient for a competitive equilibrium.*

Since the government can choose an arbitrary tax function, it is only bound by the incentive compatibility constraints which characterize the informational problem. This property means that the income tax function that is assumed here to implement the optimal allocation is without loss of generality. Any other implementation would at least have to satisfy the same two incentive constraints.

The tax on intermediate goods provides the government with an additional instrument relative to the Mirrlees (1971) setting. The planner can use this instrument to affect the income of the two types of workers but its use creates a divergence between an agents' productivity and his wage rate. We now describe some useful results for this economy, which are proved in the appendix.

Proposition 2. *The agent with a higher pre-tax wage has a higher pre-tax income and consumption, i.e. if $w_i > w_j$ then $Y_i \geq Y_j$ and $C_i \geq C_j$. This result implies that the incentive compatibility constraint for agent i binds.*

Proposition 3. *If $w_i > w_j$, the incentive compatibility of agent i binds with equality and the condition $Y_i \geq Y_j$ is equivalent to the incentive compatibility constraints of the two agents.*

To bring the analysis closer to a canonical Mirrleesian approach, we maximize the planner's objective in two steps. First, we set τ_x to a given level and solve for the optimal allocations. Second, we find the optimal level of τ_x. We define the level of social welfare conditional on a value of τ_x, as

$$W(\tau_x) = \max_{C_n, C_r, Y_n, Y_r} \frac{1}{2}\left[u(C_r) - v\left(Y_r/w_r\right) + g(G)\right] + \frac{1}{2}\left[u(C_n) - v\left(Y_n/w_n\right) + g(G)\right]$$

subject to

$$\begin{aligned} & u(C_i) - v\left(Y_i/w_i\right) = u(C_j) - v\left(Y_j/w_i\right), \\ & Y_i \geq Y_j, \\ & C_r + C_n + G \leq Y_r \frac{\tau_x + \alpha}{\alpha(1+\tau_x)} + \frac{Y_n}{1+\tau_x}. \end{aligned}$$

where the index i denotes the agent with the higher pre-tax wage. An optimal choice of τ_x requires that $W'(\tau_x) = 0$.

Proposition 4. *In the optimal plan, the tax on robots is always low enough that* $w_r \leq w_n$.

The intuition for this result is as follows. Suppose that τ_x is such that the wage on routine workers is higher than that of non-routine workers ($w_r > w_n$). This property would require a large, positive value for τ_x. A marginal reduction in τ_x has two benefits. First, it increases production efficiency by reducing the distortions associated with robot use. Second, it reduces the informational rent that the planner must give the routine agent so that he does not mimic the non-routine agent.

Proposition 5. *In the optimal plan, when automation is incomplete (*$m < 1$*), so routine workers are used in productive activities (*$Y_r > 0$*), robot taxes are strictly positive (*$\tau_x > 0$*).*

The first part of this proposition is similar to our previous result. Suppose that $\tau_x < 0$, then $w_n > w_r$. A marginal increase in τ_x has two benefits. First, it strictly increases output and hence the amount of goods available for consumption. Second, it reduces the relative wage w_n/w_r and makes the non-routine worker less inclined to mimic the routine workers. This property can be easily seen by rewriting the IC for non-routine workers in terms of hours worked,

$$u(C_r) - v(N_r) \geq u(C_n) - v\left(w_n N_n / w_r\right).$$

The second part of this result establishes that production efficiency is not optimal, so it is optimal to tax robot use. The intuition for this result is as follows. Since a zero tax on robots maximizes output, a marginal increase in that tax produces only second-order output losses. On the other hand, increasing τ_x generates a first-order gain from loosening the informational restriction. Therefore, a planner that chooses $\tau_x = 0$ can always improve its objective with a marginal increase in τ_x.

Robot taxes are optimal only when automation is incomplete ($m < 1$), so non-routine workers are employed in production ($Y_r > 0$). When full automation is optimal ($m = 1$, $Y_n = 0$) there are no informational gains from taxing robots. Since the robot tax distorts production and does not help loosen the IC of the non-routine agent, the optimal value of τ_x is zero. We prove this result in the appendix.

Now, we turn to the study of the optimal wedges. In what follows, we assume that $Y_n \geq Y_r$ is a non-binding constraint. The optimality conditions imply the following marginal rates of substitution

$$\frac{v'(Y_n/w_n)}{u'(C_n)} = w_n \frac{\tau_x + \alpha}{\alpha(1+\tau_x)}$$
$$\frac{v'(Y_r/w_r)}{u'(C_r)} \geq \frac{0.5 - \eta_n}{0.5 - \eta_n \frac{v'(Y_r/w_n)/w_n}{v'(Y_r/w_n)/w_r}} \frac{w_r}{1+\tau_x}, \quad (= \text{if } Y_r > 0).$$

In Mirrlees-type problems it is common to be optimal to not distort the choices of high-ability agents. In contrast, in our model non-routine workers are subsidized at the margin when automation is incomplete. This subsidy corrects the distortion created by the fact that in the presence of robot taxes, the wages of non-routine workers are lower than their marginal productivity.

Routine workers are taxed at the margin when automation is incomplete for two reasons. First, this tax corrects the distortion created by the fact that in the presence of robot taxes, the wages of routine workers are higher than their marginal productivity. Second, taxing routine workers makes it less appealing for non-routine workers to mimic routine workers and loosens the IC of non-routine workers.

Figure 3 illustrates the properties of the equilibrium associated with Mirrleesian optimal taxation. We see that full automation occurs for values of ϕ lower than 0.26. The robot tax is positive, reaching a maximum value of 10 percent. This tax is used only when automation is incomplete. With complete automation, routine workers supply zero hours. At that point, taxing robots to raise the wages of routine workers does not help reduce income inequality. Since taxing robots distorts production and doesn't help redistribute income, it is optimal to set these taxes to zero.

Consumption of non-routine workers is higher than that of routine workers. Since non-routine workers work harder than routine workers, the former need to receive higher consumption to satisfy their IC constraint. Both types of workers see their consumption rise as ϕ goes towards zero. This outcome is achieved through large transfers to the routine workers.

Once there is full automation, the utility of routine and non-routine workers is equalized. This property follows from the IC constraint for non-routine workers (equation (24)). This constraint is binding and when automation is full, the pre-tax income of routine workers is zero, and the right-hand side of the equation is equal to the utility of the routine worker.

6 Optimal policy with simple income taxes

The previous section describes the optimal income tax in the presence of Mirrlees-style information constraints. Despite its natural appeal, this optimal tax schedule can be complex and difficult to implement. For this reason, in this section we characterize the optimal tax policy when the after-tax income schedule is constrained to have the simple form proposed by Heathcote, Storesletten and Violante (2014) (see equation (13)).

When the government is restricted to setting income taxes consistent with the functional form (13), the competitive equilibrium in the economy can be summarized by the following conditions:

$$C_n = \lambda(\alpha Y)^{1-\gamma}, \tag{27}$$

$$N_n = \frac{1-\gamma}{1-\gamma+\theta}, \tag{28}$$

$$C_r = \lambda[(1-\alpha)(1-m)Y]^{1-\gamma}, \tag{29}$$

$$N_r = \frac{1-\gamma}{1-\gamma+\theta}, \tag{30}$$

$$m = \max\left\{1 - \left(\frac{\phi(1+\tau_x)}{(1-\alpha)A}\right)^{1/\alpha} \frac{N_r}{N_n}, 0\right\}, \tag{31}$$

$$Y = A\left(\frac{N_r}{1-m}\right)^{1-\alpha} N_r^{\alpha}, \tag{32}$$

$$C_n + C_r + G \leq Y - \phi\frac{m}{1-m}N_r, \tag{33}$$

$$x_i = \frac{N_n}{1-m}, \text{for } i \in [0, m], \tag{34}$$

$$n_i = \frac{N_n}{1-m}, \text{for } i \in (m, 1], \tag{35}$$

$$p_i = \phi. \tag{36}$$

Taking the ratio between equations (27) and (29), we can see that a necessary condition is

$$\frac{C_n}{C_r} = \left[\frac{\alpha}{(1-\alpha)(1-m)}\right]^{1-\gamma}. \tag{37}$$

This equation shows that there are two ways to make the ratio C_n/C_r closer to one. The first is to raise τ_x which leads to a fall in the level of automation, m. The second, is to make γ closer to one, i.e. make the tax system more progressive. Both approaches have drawbacks. Taxing robot use reduces production efficiency and making the tax system more progressive reduces incentives to work. To see the latter effect, note that hours worked are given by equations (28) and (30). As γ approaches one, hours worked approach zero.

We can think of the planner as choosing allocations $\{C_r, C_n, G, m\}$ and progressivity γ, subject to (37) and

$$C_r + C_n + G \leq A\left(\frac{1}{1-m}\right)^{1-\alpha} \frac{1-\gamma}{1-\gamma+\theta} - \phi\frac{m}{1-m}\frac{1-\gamma}{1-\gamma+\theta}, \tag{38}$$

which represents the resource condition in the equilibrium definition, where the variables Y, N_n and N_r have been replaced by their equilibrium expressions. These two conditions are necessary and sufficient to describe the set of implementable allocations in terms of $\{C_r, C_n, G, m\}$ and γ. They are necessary because they follow from straightforward manipulations of equilibrium conditions. Furthermore, given any allocations $\{C_r, C_n, G, m\}$ and γ that satisfy these two constraints, we can use equation (27) to find a value for λ; such a value for λ also satisfies equation (29), since equation (37) must be satisfied. In addition, equations (28) and (30) can be used to find solutions for N_n and N_r, respectively. Given an optimal allocation for m, equation (31) can be solved by a choice of τ_x, and equation (32) yields a value for Y. These solutions, together with equation (38), imply that equation (33) is also satisfied. Finally, equation (34) can be used to solve for a value for each x_i, equation (35) for each n_i, and equation (36) for p_i.

Optimality implies the following condition

$$\mu \frac{A(1-\alpha)}{(1-m)(1-\gamma+\theta)}\left[(1-m)^{\alpha} - \frac{\phi}{A(1-\alpha)}\right] = 0.5 - \frac{C_r}{C_r + C_n}, \tag{39}$$

where μ is the Lagrange multiplier associated with the resource constraint, (38). This expression implies that when automation is incomplete, robot use is always taxed. To see this result, it is useful to note that when $\tau_x = 0$ the equilibrium level of automation is such that[4]

$$1 - m = \left[\frac{\phi}{(1-\alpha)A}\right]^{1/\alpha}. \tag{40}$$

Equation (40) implies that the left-hand side of equation (39) is equal to zero which is possible only when the $C_r = C_n$. Since the instruments available to the government are distortionary, it is in general not optimal to equalize the consumption of the two agents. Equation (39) implies that when $C_n > C_r$, the level of m is lower that that implied by the competitive equilibrium with $\tau_x = 0$. So, in order for equation (39) to hold, τ_x must be positive.

Figure 4 shows that the form of the tax function constrains heavily the outcomes that can be achieved. Full automation never occurs and robot taxes are used for all values of ϕ reaching values as high as $\tau_x = 0.33$. As the costs of automation decline, the

[4]This result reflects the fact that in equilibrium $N_n = N_r$.

progressivity of the income tax rises. But there is still a large divergence in wage rates, consumption and utility across the two types of workers. The reason is the transfers from non-routine to routine workers are relatively modest.

Optimal policy with lump-sum rebates In both the first-best allocation and the Mirrlees-style allocation, routine workers drop out of the labor force once automation costs are sufficiently low. That property is absent in an equilibrium where taxes take the form proposed by Heathcote, Storesletten and Violante (2014). The reason is simple. Equation (13) implies that when before-tax income is zero, after-tax income is also zero. A worker who drops out of the labor force has zero consumption and $-\infty$ utility. In order to make the outcomes that can be achieved with a simple tax system closer to the Mirrlees-style allocation we now consider the case where the government can use a lump-sum rebate, T.

In this specification, the after-tax income of household j is

$$y(w_j N_j) = \lambda(w_j N_j)^{1-\gamma} + T. \tag{41}$$

To simplify the algebra, it is useful to think of the planner as choosing $\tilde{T}$ such that $\tilde{T} = T/\left(\lambda Y^{1-\gamma}\right)$. Consumption levels and working hours for each agent are given by

$$C_r = \lambda Y^{1-\gamma}[[(1-\alpha)(1-m)]^{1-\gamma} + \tilde{T}] \tag{42}$$

$$N_r = \frac{1-\gamma}{1-\gamma+\theta+\frac{\theta\tilde{T}}{[(1-\alpha)(1-m)]^{1-\gamma}}} \tag{43}$$

$$C_n = \lambda Y^{1-\gamma}[\alpha^{1-\gamma} + \tilde{T}] \tag{44}$$

$$N_n = \frac{1-\gamma}{1-\gamma+\theta+\theta\tilde{T}/\alpha^{1-\gamma}} \tag{45}$$

The planner chooses allocations $\{C_n, C_r, m\}$ and the instruments $\{\gamma, \tilde{T}\}$, subject to

$$C_n = \left\{ \frac{\alpha^{1-\gamma} + \tilde{T}}{[(1-\alpha)(1-m)]^{1-\gamma} + \tilde{T}} \right\} C_r, \tag{46}$$

$$C_r + C_n + G = A \left[\frac{N_r}{(1-m)} \right]^{1-\alpha} N_n^{\alpha} - \phi \frac{m}{1-m} N_r, \tag{47}$$

$$\tilde{T} \geq 0, \tag{48}$$

where N_r and N_n are given by equations (43) and (45). Equations (46), (47) and (48) are necessary and sufficient for an equilibrium.

In this economy, the lump-sum transfer plays a similar role to tax progressivity. The consumption ratio is given by,

$$\frac{C_r}{C_n} = \frac{[(1-\alpha)(1-m)]^{1-\gamma} + \tilde{T}}{\alpha^{1-\gamma} + \tilde{T}},$$

It is easy to see that when $C_r/C_n < 1$, an increase in $\tilde{T}$ increases C_r/C_n, reducing consumption inequality. However, lump-sum rebates have to be financed with distortionary income taxes.

Figure 5 illustrates the properties of this allocation. In this equilibrium, income is redistributed through a large lump-sum transfer, in other words, the government guarantees a minimum income to all agents in the economy. Workers have two sources of income: wages and transfers. For this reason, income and substitution effects of changes in wages are no longer offsetting. As a consequence, the two types of workers supply a different number of hours and their hours vary with ϕ. When automation is incomplete, robot taxes are used as an additional source of redistribution and τ_x can go as high as 35 percent. Complete automation occurs for values of ϕ lower than 0.2. When automation is incomplete, the income tax is regressive ($\gamma < 0$) to reduce the distortions on the labor supply of the non-routine agents.

7 Comparing different policies

In this section, we compare the first-best allocation with the allocations associated with different policies in terms of social welfare and the utility of routine and non-routine workers. In the figures discuss below we use the labels FB, SQ, OT, ST and STL to refer to the first-best, status quo, Mirrleesian optimal taxes, simple taxes, and simple taxes with lump-sum rebates, respectively.

Figure 6 shows the utility of the social planner for values of ϕ in the interval $(0, (1-\alpha)A]$. Recall that $(1-\alpha)A$ is the lowest value of ϕ for which there is no automation in the status quo. Social welfare rises as the costs of automation fall both for the first best and for all the policies we consider. We see that the Mirrlees allocation is relatively close in terms of welfare to the first-best allocation. The solution with simple taxation and rebates ranks next in terms of welfare, followed by the solution with simple taxes without rebates. The status quo is by far the worst allocation.

A fall in the cost of automation can have very different consequences for routine and non-routine workers. To illustrate this property, we measure the utility of the two types of workers relative to the status-quo equilibrium with $\phi = (1-\alpha)A$. We call this allocation the no-automation benchmark. Panel A (B) of Figure 7 shows how much routine (non-routine) workers would pay as percentage of consumption to go back to the no-automation benchmark for different values of ϕ.

Panel A of Figure 7 shows that the utility of routine workers in the first-best allocation improves as ϕ falls. In contrast, in the status quo, routine workers become increasingly worse off as ϕ falls. With Mirrleesian optimal taxation, routine workers are made better off once ϕ becomes sufficiently low (lower than 0.35). For simple taxes with and without rebates, routine workers are better off than in the no-automation benchmark for values of ϕ lower than 0.21 and 0.09, respectively.

Figure B of Figure 7 shows that non-routine workers prefer the no-automation benchmark to the first best for high levels of ϕ (higher than 0.23). This preference reflects the large transfers that non-routine workers make to routine workers in the first best. For values of ϕ lower than 0.23, non-routine workers prefer the first best to the no-automation benchmark. The reason is that the wage of non-routine workers is high enough to compensate the transfers they make to routine workers. For values of ϕ lower than 0.39 non-routine workers prefer the status quo to all other allocations. This preference results from a combination of high wages and relatively low taxes.

For any level of ϕ lower that $(1-\alpha)A$, routine workers rank the first-best allocation first, Mirrleesian optimal taxation second, simple taxes with rebates third, and simple taxes without rebates fourth and the status quo last. In contrast, non-routine workers rank the status quo first and the first best last. Mirrleesian optimal taxation and simple taxes with and without rebates rank in between the two extremes.

8 Relation to the public finance literature

Our results stand in sharp contrast to the celebrated Diamond and Mirrlees (1971) result that an optimal tax system should ensure efficiency in production and therefore leave intermediate goods untaxed. In our framework, this property would imply that the tax on robots should be zero. Another important reference is Atkinson and Stiglitz (1976). These authors argue that in an economy with Mirrleesian income taxes distorting the use of commodities is not optimal, as long as these commodities are separable from leisure in the utility function. Since uniform taxation can be interpreted as production efficiency, those results may also appear to contradict ours. In this section, we discuss the relation between these different results.

Relating our results to Diamond and Mirrlees (1971) It is central to the Diamond-Mirrlees intermediate good theorem that all net trades can be taxed at potentially different (linear) rates. In our model, this property would mean that the labor of the two types of workers can be taxed at different rates. At the heart of the failure of the Diamond and Mirrlees (1971) theorem in our model is the fact that the government cannot discriminate between the two types of workers.

The result in Diamond and Mirrlees holds if we assume that the planner can use different linear taxes for routine and non-routine workers, λ_r and λ_n. In this case, household optimality implies that

$$\frac{v'(N_j)}{u'(C_j)} = \lambda_j w_j, \quad \text{and} \quad C_j = \lambda_j w_j N_j.$$

With the ability to affect each marginal rate of substitution independently, the only constraints faced by the planner are the resource constraint

$$C_r + C_n + G \leq A \left(\frac{N_r}{1-m} \right)^{1-\alpha} N_n^{\alpha} - \phi \frac{m}{1-m} N_r,$$

and the implementability conditions for household optimality

$$u'(C_j)C_j - v'(N_j)N_j = 0, \quad \text{for } j = r, n.$$

These three conditions are necessary and sufficient for an equilibrium.

When the government can use different tax rates for each type of worker, the level of intermediate goods appears only in the resource constraint and not in the implementability condition. Since intermediate goods do not interfere with incentives, they are chosen to maximize output for given levels of hours worked. This objective is achieved by not distorting production, setting $\tau_x = 0$.

When the tax system requires both types of worker to pay the same tax rate ($\lambda_r = \lambda_n$), the planner has to distort the labor supply decisions of the two types of workers in the same way, which gives rise to the following additional implementability restriction

$$\frac{v'(N_n)/u'(C_n)}{v'(N_r)/u'(C_r)} = \frac{w_n}{w_r}. \tag{49}$$

The value of τ_x no longer appears only in the resource constraint, it appears in equation (49) because the wage ratio is a function of τ_x. As a result, to relax restriction

(49), it might be optimal for the planner to choose values of τ_x that are different from zero.

This result depends crucially on the fact that different labor types interact differently with the intermediate good, which means that distorting the use of intermediate goods affects in different ways the wage rates of different workers. If the production function was weakly separable in labor types and intermediate inputs, the wage ratio would be independent of the usage of intermediate inputs and production efficiency would be optimal.

In our model, robots are substitutes of routine workers and complements of non-routine. A tax on robots decreases the wage rate of non-routine workers and increase the wage rate of routine workers. This property means that it can be optimal to use robot taxes.

Relating our results to Atkinson and Stiglitz (1976) In Mirrleesian optimal taxation the planner can choose the entire income tax schedule. One might expect production efficiency to be optimal given the high degree of flexibility associated with income taxes. Indeed, Atkinson and Stiglitz (1976) show that, under weak conditions, uniform commodity taxation (production efficiency) is optimal. Jacobs (2015) shows that the Atkinson and Stiglitz (1976) result relies on a form of separability such that commodities do not interact differently with different agent types. This form of separability is not present in our model which is why it can be optimal to tax robots.

The intuition for this result is the same we used in discussing proposition 5. Because the government does not know the type of the agent and only observes income, it is restricted to use incentive compatible tax systems. Since different types interact differently with the intermediate good, the distorting production decisions may help in the screening process. To see this property, it is useful to write the incentive compatibility constraint as follows:

$$u(C_i) - v(N_i) \geq u(C_j) - v\left(w_j N_j / w_i\right).$$

Crucially, this incentive compatibility constraint involves the wage ratio. Whenever the taxation of intermediate goods affects this ratio, production efficiency may no longer be optimal. When intermediate goods are not separable in production from the two labor types, taxing intermediate goods affects the wage ratio and it might be optimal to distort production

In sum, the classical results on production efficiency in the public finance literature depend on one of two key assumptions: (i) the government can tax differently every consumption good and labor type; or (ii) the environment is such that production distortions do not help in shaping incentives. Both assumptions fail in our model. On the one hand, the government cannot design income tax systems that independently target each type of worker. On the other hand, robots are substitutes for routine workers and complements to non-routine workers, so a tax on robots affects the ratio of the wages of these two types of workers.

9 Conclusions

Our analysis suggests that without changes to the current U.S. tax system, a sizable fall in the costs of automation would lead to a massive rise in income inequality. Even though routine workers keep their jobs, their wages fall to make them competitive with the possibility of automating production.

Income inequality can be reduced by raising the marginal tax rates paid by high-income individuals and by taxing robots to raise the wages of routine workers. But this solution involves a substantial efficiency loss for the reduced level of inequality.

A Mirrleesian optimal income tax can reduce inequality at a smaller efficiency cost than the variants of the U.S. tax system discussed above, coming close to the levels of social welfare obtained in the first-best allocation. Unfortunately, this tax system can be complex and difficult to implement.

An alternative approach is to amend the tax system to include a rebate that is independent of income. In our model, with this rebate in place, it is optimal to tax robots for values of the automation cost that lead to partial automation. For values of the automation cost that lead to full automation, it is not optimal to tax robots. Routine workers lose their jobs and live off government transfers, just like in Kurt Vonnegut's "Player's piano."

10 References

Atkinson, Anthony Barnes, and Joseph E. Stiglitz. "The design of tax structure: direct versus indirect taxation." *Journal of Public Economics* 6.1-2 (1976): 55-75.

Diamond, Peter A., and James A. Mirrlees. "Optimal taxation and public production I: Production efficiency." *The American Economic Review* 61.1 (1971): 8-27.

Guesnerie, Roger *A Contribution to the Pure Theory of Taxation*. Cambridge-UK: Cambridge University Press, 1995.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. "Optimal tax progressivity: An analytical framework." No. w19899. National Bureau of Economic Research, 2014.

Jacobs, Bas "Optimal Inefficient Production," manuscript, Erasmus University, Rotterdam, 2015.

Jacobs, Bas and Robin Boadway "Optimal linear commodity taxation under optimal non-linear income taxation," CESifo Working Paper: Public Finance, No. 4142, 2013.

Mirrlees, James A. "An exploration in the theory of optimum income taxation." *The Review of Economic Studies* 38.2 (1971): 175-208.

Ramsey, Frank P. "A Contribution to the Theory of Taxation." *The Economic Journal* 37.145 (1927): 47-61.

A Appendix

A.1 The first-best allocation

We define the first-best allocation in this economy as the solution to an utilitarian welfare function, absent informational constraints. This absence implies that the planner can perfectly discriminate among agents and enforce any allocation. The optimal plan solves the following problem

$$V = \max_{\{C_r,N_r,C_n,N_n\},G,m,\{x_i,n_i\}} \frac{1}{2}\left[u(C_r) - v(N_r) + g(G)\right] + \frac{1}{2}\left[u(C_n) - v(N_n) + g(G)\right].$$

$$C_r + C_n + G \leq A\left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di\right]^{\frac{1-\alpha}{\rho}} N_n^\alpha - \int_0^m \phi x_i di, \quad [\mu],$$

$$\int_m^1 n_i di = N_r, \quad [\eta].$$

The first-order conditions with respect to n_i and x_i are

$$\mu(1-\alpha)A\left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di\right]^{\frac{1-\alpha}{\rho}-1} N_n^\alpha n_i^{\rho-1} = \eta, \quad \forall i \in (m,1]$$

$$(1-\alpha)A\left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di\right]^{\frac{1-\alpha}{\rho}-1} N_n^\alpha x_i^{\rho-1} = \phi, \quad \forall i \in [0,m].$$

The first equation implies that the marginal productivity of routine labor should be constant across the activities that use routine labor. This property means that $n_i = N_r/(1-m)$ for $i \in (m,1]$ and $n_i = 0$, otherwise. The same property applies to robots used in the activities where they are used, $x_i = x$ for $i \in [0,m]$ and $x_i = 0$, otherwise.

To characterize the optimal allocations we replace n_i and x_i in the planner's problem, which can be rewritten as

$$V = \max_{\{C_r,N_r,C_n,N_n\},G,m,x} \sum_{j=r,n} \frac{1}{2}\left[u(C_j) - v(N_j) + g(G)\right].$$

$$C_r + C_n + G \leq A\left[mx^\rho + (1-m)\left(\frac{N_r}{1-m}\right)^\rho\right]^{\frac{1-\alpha}{\rho}} N_n^\alpha - \phi mx, \quad [\mu].$$

The first-order conditions with respect to x and m are, respectively,

$$(1-\alpha)A\left[mx^{\rho}+(1-m)\Big(\frac{N_r}{1-m}\Big)^{\rho}\right]^{\frac{1-\alpha}{\rho}-1}N_n^{\alpha}x^{\rho-1}=\phi,$$

$$\frac{1-\alpha}{\rho}A\left[mx^{\rho}+(1-m)\Big(\frac{N_r}{1-m}\Big)^{\rho}\right]^{\frac{1-\alpha}{\rho}-1}N_n^{\alpha}\left[x^{\rho}-(1-\rho)\Big(\frac{N_r}{1-m}\Big)^{\rho}\right]=\phi x.$$

The ratio of these two equations implies that if automation is positive, $m > 0$, then $x = N_r/(1-m)$. Using this condition, we obtain

$$V=\max_{\{C_r,N_r,C_n,N_n\},G,m,x}\sum_{j=r,n}\frac{1}{2}\left[u(C_j)-v(N_j)+g(G)\right].$$

$$C_r+C_n+G\leq A\Big(\frac{N_r}{1-m}\Big)^{1-\alpha}N_n^{\alpha}-\phi m\frac{N_r}{1-m},\quad [\mu].$$

The first-order condition with respect to the level of automation implies that

$$(1-\alpha)A\frac{1}{(1-m)^{2-\alpha}}N_r^{1-\alpha}N_n^{\alpha}-\phi\frac{N_r}{(1-m)^2}=0\Leftrightarrow m=1-\left[\frac{\phi}{A(1-\alpha)}\right]^{1/\alpha}\frac{N_r}{N_n},$$

provided that m is interior. Then,

$$m=\max\left\{1-\left[\frac{\phi}{A(1-\alpha)}\right]^{1/\alpha}\frac{N_r}{N_n},0\right\}.$$

Furthermore, the first-order conditions with respect to C_r, C_n, N_r, N_n, and G are

$$\frac{1}{2}u'(C_r)=\mu,$$

$$\frac{1}{2}u'(C_n)=\mu,$$

$$\frac{1}{2}v'(N_r)\geq\frac{\mu}{N_r}(1-\alpha)(1-m)Y,$$

$$\frac{1}{2}v'(N_n)=\mu\alpha Y,$$

$$g'(G)=\mu.$$

The first-order condition with respect to N_r is presented with inequality, because the constraint $N_r \geq 0$ may bind when automation costs are low. The combination of the first two equations implies that

$$u'(C_r) = u'(C_n) \Leftrightarrow C_r = C_n.$$

The optimal marginal rates of substitution are given by the combination of the marginal utility of consumption and leisure for each individual

$$\frac{v'(N_r)}{u'(C_r)} \geq (1-\alpha)(1-m)\frac{Y}{N_r},$$
$$\frac{v'(N_n)}{u'(C_n)} = \alpha\frac{Y}{N_n}.$$

Finally, from the first-order conditions for G and C_j it follows that

$$g'(G) = \frac{1}{2}u'(C_j). \tag{50}$$

A.2 Proof of Proposition 1

In an equilibrium, robot producers set the price of robots equal to their marginal cost

$$p_i = \phi. \tag{51}$$

Optimality for final goods producers implies that

$$x_i = \begin{cases} \frac{N_r}{1-m}, & m \in [0, m], \\ 0, & \text{otherwise} \end{cases} \tag{52}$$

$$n_i = \begin{cases} \frac{N_r}{1-m}, & m \in (m, 1], \\ 0, & \text{otherwise} \end{cases} \tag{53}$$

$$m = \max\left\{1 - \left[\frac{(1+\tau_x)\phi}{(1-\alpha)A}\right]^{1/\alpha}\frac{N_r}{N_n}, 0\right\}, \tag{54}$$

$$Y = A\left[\int_0^m x_i^\rho di + \int_m^1 n_i^\rho di\right]^{\frac{1-\alpha}{\rho}} N_n^\alpha, \tag{55}$$

$$w_r = (1-\alpha)(1-m)\frac{Y}{N_r}, \tag{56}$$

$$w_n = \alpha\frac{Y}{N_n}. \tag{57}$$

The resource constraint is

$$C_r + C_n + G = Y - \int_0^m \phi x_i, \tag{58}$$

We can let equation (51) define the price of robots, equation (52) define x_i, equations (53), (54) and (55) determine n_i, m, and Y, respectively. Assuming that m is interior, the wage equations (56) and (57) can be written as (10) and (11). These equations can be used to solve for the equilibrium wage rates. Combining the results above, we can write the resource constraint as

$$C_r + C_n + G = \alpha\frac{A^{1/\alpha}(1-\alpha)^{\frac{1-\alpha}{\alpha}}}{[(1+\tau_x)\phi]^{\frac{1-\alpha}{\alpha}}}\frac{\tau_x+\alpha}{\alpha(1+\tau_x)}N_n + \phi N_r.$$

Replacing $N_j = Y_j/w_j$, the resource constraint can be written as

$$C_r + C_n + G = Y_n\frac{\tau_x+\alpha}{\alpha(1+\tau_x)} + \frac{Y_r}{1+\tau_x}. \tag{59}$$

This derivation makes it clear that resource constraint (59) is necessary and sufficient for optimality in the production side of the economy.

Household optimality requires that

$$u(C_j) - v\Big(\frac{Y_j}{w_j}\Big) \geq u(C) - v\Big(\frac{Y}{w_j}\Big), \quad \forall (C,Y) : C \leq Y - T(Y).$$

The following incentive compatibility are necessary constraints

$$u(C_n) - v\Big(\frac{Y_n}{w_n}\Big) \geq u(C_r) - v\Big(\frac{Y_r}{w_n}\Big),$$
$$u(C_r) - v\Big(\frac{Y_r}{w_r}\Big) \geq u(C_n) - v\Big(\frac{Y_n}{w_r}\Big).$$

These are also sufficient conditions, because the planner can set the tax schedule $T(\cdot)$ such that for all $Y \notin \{Y_n, Y_r\}$ the allocation is worse for both agents than their respective allocation. $T(Y) = Y$ is an example of such a mechanism.

A.3 Proof of Proposition 2 and 3

By adding the two incentive compatibility constraints, we obtain

$$v\Big(\frac{Y_i}{w_j}\Big) - v\Big(\frac{Y_j}{w_j}\Big) \geq v\Big(\frac{Y_i}{w_i}\Big) - v\Big(\frac{Y_j}{w_i}\Big).$$

Since $v(\cdot)$ is a convex function and $w_i > w_j$, a necessary and sufficient condition is $Y_i \geq Y_j$.

Then, from the IC for i we can see that

$$u(C_i) - u(C_j) \geq v\Big(\frac{Y_i}{w_i}\Big) - v\Big(\frac{Y_j}{w_i}\Big) \geq 0,$$

which implies that $C_i \geq C_j$.

We now show by contradiction that the incentive compatibility of agent i binds. Suppose we have found a solution where that IC is not binding, then

$$u(C_i) - v\Big(\frac{Y_i}{w_i}\Big) > u(C_j) - v\Big(\frac{Y_i}{w_j}\Big),$$

which implies that $C_i > C_j$. Suppose then a different allocation where Y_i, Y_j, G and τ_x are kept at the same level but $C_i' = C_i - \epsilon$ and $C_j' = C_j + \epsilon$, for some small $\epsilon > 0$, which preserves the inequality

$$u(C_i') - v\Big(\frac{Y_i}{w_i}\Big) \geq u(C_j') - v\Big(\frac{Y_i}{w_j}\Big).$$

This new allocation is still resource feasible and loosens the incentive compatibility constraint for agent j.

Concavity of $u(\cdot)$ guarantees that welfare evaluated at this new allocation is strictly greater than the initial candidate solution, which contradicts the premise that the initial allocation was a solution to the problem.

We have already shown that the condition $Y_i \geq Y_j$ is necessary. It remains to be shown that it can replace the incentive compatibility of agent j. We now show that since the IC of i binds with equality, $Y_i \geq Y_j$ implies the IC of j

$$u(C_i) - u(C_j) = v\Big(\frac{Y_i}{w_i}\Big) - v\Big(\frac{Y_i}{w_j}\Big) \leq v\Big(\frac{Y_i}{w_j}\Big) - v\Big(\frac{Y_j}{w_j}\Big)$$
$$\Rightarrow u(C_i) - v\Big(\frac{Y_i}{w_j}\Big) \leq u(C_j) - v\Big(\frac{Y_j}{w_j}\Big).$$

A.4 Proof of Proposition 4

We show this result by arguing that such a level of τ_x cannot be optimal. We show that $w_r \leq w_n$ by conditioning the problem on a level of τ_x such that $w_r > w_n$ and showing that it cannot be optimal.

Suppose that non-routine workers earn a higher wage rate than routine workers. This property implies that the allocations solve

$$\begin{aligned} W(\tau_x) = \max_{C_r, C_n, Y_n, Y_r} & [u(C_r) - v\left(Y_r/w_r\right) + g(G)] + [u(C_n) - v\left(Y_n/w_n\right) + g(G)]\,, \\ & u(C_r) - v\left(Y_r/w_r\right) = u(C_n) - v\left(Y_n/w_r\right), \quad [\eta_r] \\ & Y_r \geq Y_n, \quad [\psi] \\ & C_r + C_n + G \leq Y_n \frac{\tau_x + \alpha}{\alpha(1+\tau_x)} + \frac{Y_r}{1+\tau_x}, \quad [\mu]. \end{aligned}$$

Then envelope condition is given by

$$\begin{aligned} W'(\tau_x) &= v'(Y_r/w_r)(1+\eta_r)\frac{Y_r}{w_r^2}\frac{\partial w_r}{\partial \tau_x} + v'(Y_n/w_n)\frac{Y_n}{w_n^2}\frac{\partial w_n}{\partial \tau_x} \\ &- \eta_r v'(Y_n/w_r)\frac{Y_n}{w_r^2}\frac{\partial w_r}{\partial \tau_x} + \mu\Big[Y_n\frac{1-\alpha}{\alpha(1+\tau_x)^2} - \frac{Y_r}{(1+\tau_x)^2}\Big]. \end{aligned}$$

It is useful to note that

$$\begin{aligned} \frac{\partial w_r}{\partial \tau_x} &= \phi, \quad \Rightarrow \frac{\frac{\partial w_r}{\partial \tau_x}}{w_r} = \frac{1}{(1+\tau_x)}, \\ \frac{\partial w_n}{\partial \tau_x} &= -\frac{A^{1/\alpha}(1-\alpha)^{1/\alpha}}{\phi^{\frac{1-\alpha}{\alpha}}(1+\tau_x)^{1/\alpha}}, \quad \Rightarrow \frac{\frac{\partial w_n}{\partial \tau_x}}{w_n} = -\frac{1-\alpha}{\alpha(1+\tau_x)}. \end{aligned}$$

We can use these conditions to write

$$W'(\tau_x) = v'(Y_r/w_r)(1+\eta_r)\frac{Y_r}{w_r}\frac{1}{1+\tau_x} - v'(Y_n/w_n)\frac{Y_n}{w_n}\frac{1-\alpha}{\alpha(1+\tau_x)}$$
$$- \eta_r v'(Y_n/w_r)\frac{Y_n}{w_r}\frac{1}{1+\tau_x} + \mu\Big[Y_n\frac{1-\alpha}{\alpha(1+\tau_x)^2} - \frac{Y_r}{(1+\tau_x)^2}\Big].$$

The first-order conditions for optimality of the problem imply that

$$v'(Y_r/w_r)(1+\eta_r)\frac{Y_r}{w_r}\frac{1}{1+\tau_x} = \frac{\mu Y_r}{(1+\tau_x)^2} - \psi\frac{Y_r}{1+\tau_x},$$
$$v'(Y_n/w_n)\frac{Y_n}{w_n}\frac{1-\alpha}{\alpha(1+\tau_x)} = \mu Y_n\frac{\tau_x+\alpha}{\alpha(1+\tau_x)}\frac{1-\alpha}{\alpha(1+\tau_x)} + \psi Y_n\frac{1-\alpha}{\alpha(1+\tau_x)} + \eta_r v'(Y_n/w_r)\frac{Y_n}{w_r}\frac{1-\alpha}{\alpha(1+\tau_x)},$$

which replaced in the envelope condition implies

$$W'(\tau_x) = \frac{\mu Y_r}{(1+\tau_x)^2} - \psi\frac{Y_r}{1+\tau_x} - \mu Y_n\frac{\tau_x+\alpha}{\alpha(1+\tau_x)}\frac{1-\alpha}{\alpha(1+\tau_x)} - \psi Y_n\frac{1-\alpha}{\alpha(1+\tau_x)}$$
$$- \eta_r v'(Y_n/w_r)\frac{Y_n}{w_r}\frac{1-\alpha}{\alpha(1+\tau_x)} - \eta_r v'(Y_n/w_r)\frac{Y_n}{w_r}\frac{1}{1+\tau_x} + \mu\Big[Y_n\frac{1-\alpha}{\alpha(1+\tau_x)^2} - \frac{Y_r}{(1+\tau_x)^2}\Big].$$

$$W'(\tau_x) = -\psi\frac{Y_r}{1+\tau_x} - \mu Y_n\frac{\tau_x(1-\alpha)}{\alpha^2(1+\tau_x)^2} - \psi Y_n\frac{1-\alpha}{\alpha(1+\tau_x)} - \eta_r v'(Y_n/w_r)\frac{Y_n}{w_r}\frac{1}{\alpha(1+\tau_x)} < 0.$$

A.5 Proof of Proposition 5

We follow the same strategy used to prove Proposition 4. To produce a contradiction suppose not, suppose we find an allocation $\{C_n, Y_n, C_n, Y_n, G, \tau_x\}$ such that $\tau_x \leq 0$. Using the previous results, we know that $w_n > w_r$. This result implies that $Y_n \geq Y_r$, $C_n \geq C_r$, and the IC of the non-routine worker binds. These allocations solve the original optimization problem, or equivalently they solve

$$\begin{aligned}
W(\tau_x) = \max_{C_r, C_n, Y_n, Y_n, G} \sum_j \Big[u(C_j) - v\Big(\frac{Y_j}{w_j} + g(G)\Big)\Big], \quad \text{s.to.} \\
u(C_n) - v\Big(\frac{Y_n}{w_n}\Big) = u(C_r) - v\Big(\frac{Y_r}{w_n}\Big), \quad [\eta_n], \\
Y_n \geq Y_r, \quad [\psi], \\
C_r + C_n + G \leq Y_n \frac{\tau_x + \alpha}{\alpha(1+\tau_x)} + \frac{Y_r}{1+\tau_x}, \quad [\mu].
\end{aligned}$$

The envelope condition is

$$\begin{aligned}
W'(\tau_x) = & - v'\Big(\frac{Y_n}{w_n}\Big)(1+\eta_n)\frac{Y_n}{w_n}\frac{1-\alpha}{\alpha(1+\tau_x)} + v'\Big(\frac{Y_r}{w_r}\Big)\frac{Y_r}{w_r}\frac{1}{1+\tau_x}, \\
& + \eta_n v'\Big(\frac{Y_r}{w_n}\Big)\frac{Y_r}{w_n}\frac{1-\alpha}{\alpha(1+\tau_x)} + \frac{\mu}{1+\tau_x}\Big[Y_n \frac{1-\alpha}{\alpha(1+\tau_x)} - \frac{Y_r}{1+\tau_x}\Big].
\end{aligned}$$

The first-order conditions for optimality imply that

$$\begin{aligned}
v'\Big(\frac{Y_n}{w_n}\Big)(1+\eta_n)\frac{1}{w_n} &= \mu \frac{\tau_x + \alpha}{\alpha(1+\tau_x)} - \psi. \\
v'\Big(\frac{Y_r}{w_r}\Big)\frac{1}{w_r} &= \mu \frac{1}{1+\tau_x} + \psi + \eta_n v'\Big(\frac{Y_r}{w_n}\Big)\frac{1}{w_n}.
\end{aligned}$$

Replacing these expressions in the envelope condition implies

$$W'(\tau_x) = -\mu\tau_x \frac{1-\alpha}{\alpha^2(1+\tau_x)^2} + \psi \frac{Y_n(1-\alpha)}{\alpha(1+\tau_x)} + \psi \frac{Y_r}{1+\tau_x} + \eta_n v'\Big(\frac{Y_r}{w_n}\Big)\frac{Y_r}{w_n}\frac{1}{\alpha(1+\tau_x)},$$

which shows that, as long as $\tau_x \leq 0$, then $W'(\tau_x) \geq 0$.

A.5.1 The full automation case ($m = 1$, $Y_r = 0$)

If the optimal plan has $Y_r = 0$ then $Y_n > 0$, which follows from the Inada conditions on utility. This result implies that $\psi = 0$. From the envelope condition we can see that

$$W'(\tau_x) = -\mu\tau_x \frac{1-\alpha}{\alpha^2(1+\tau_x)^2} = 0 \Leftrightarrow \tau_x = 0. \tag{60}$$

B Figures

Figure 1: Status-Quo Equilibrium

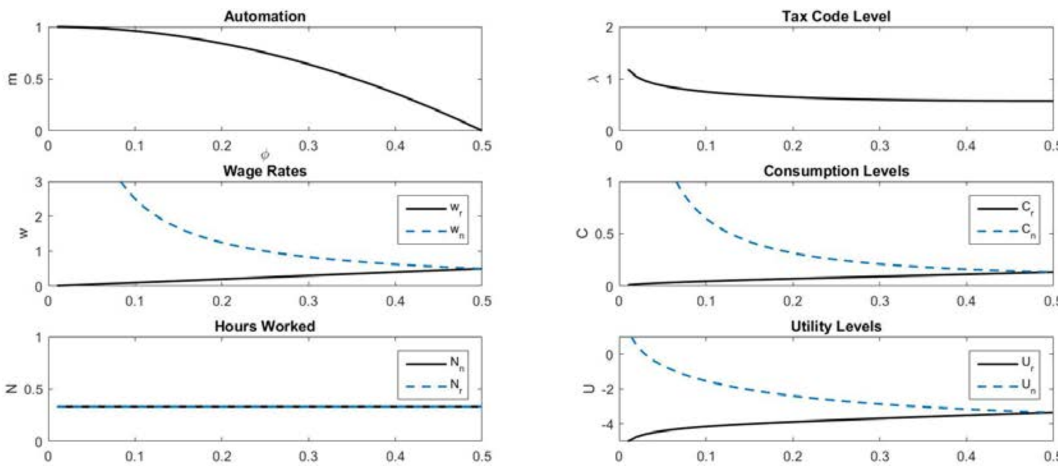

Figure 2: First Best

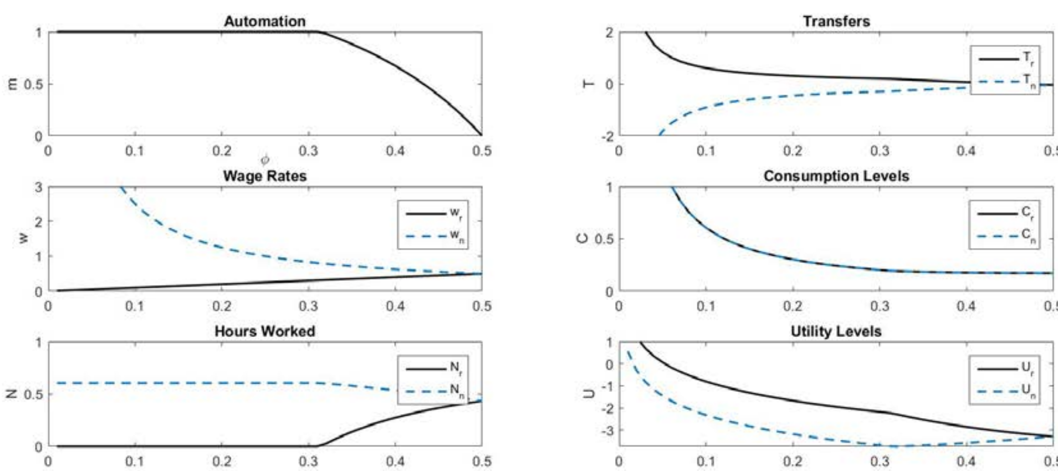

Figure 3: Mirrleesian Optimal Taxation

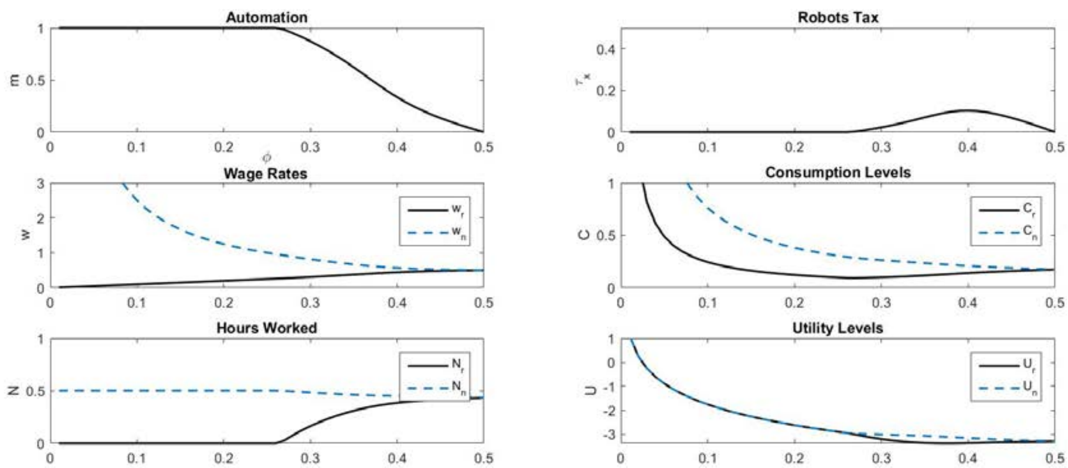

Figure 4: Simple Taxes - Panel A

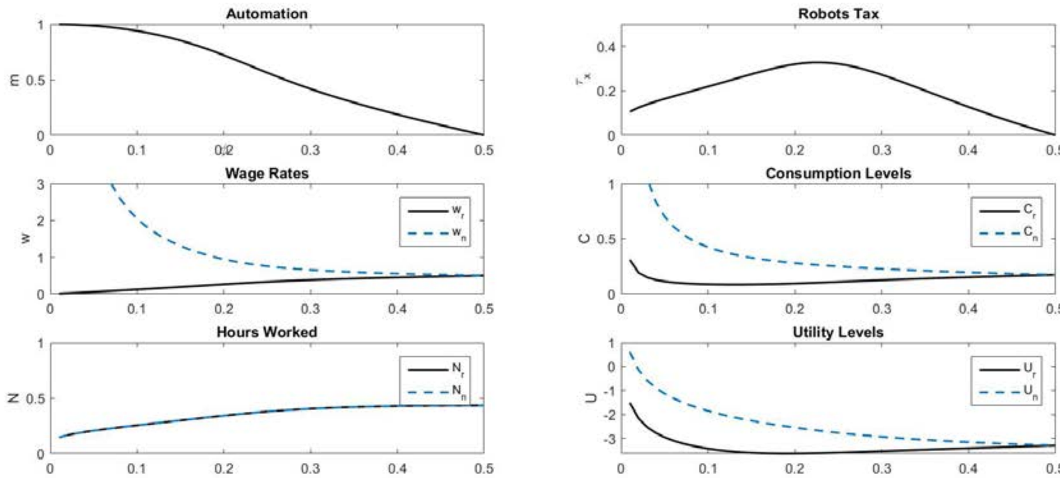

Figure 4: Simple Taxes - Panel B

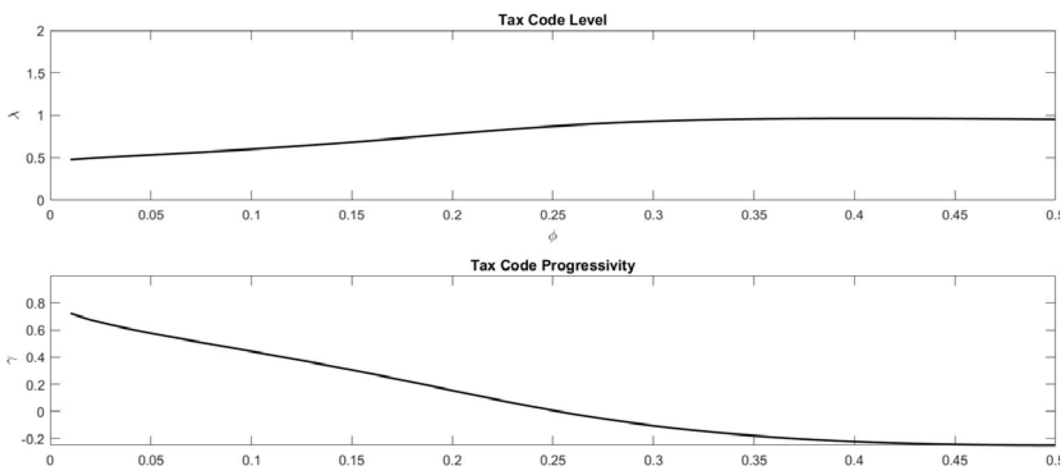

Figure 5: Simple Taxes & Lump Sum Rebate - Panel A

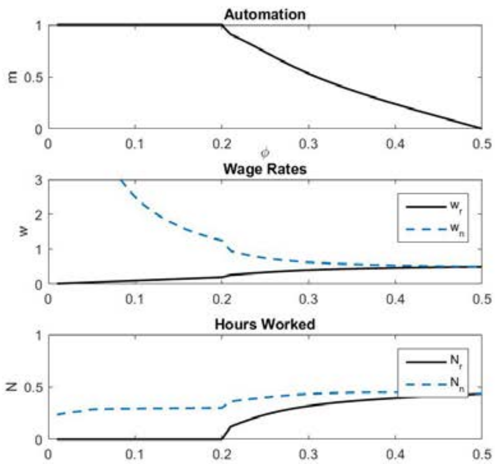

Figure 5: Simple Taxes & Lump Sum Rebate - Panel B

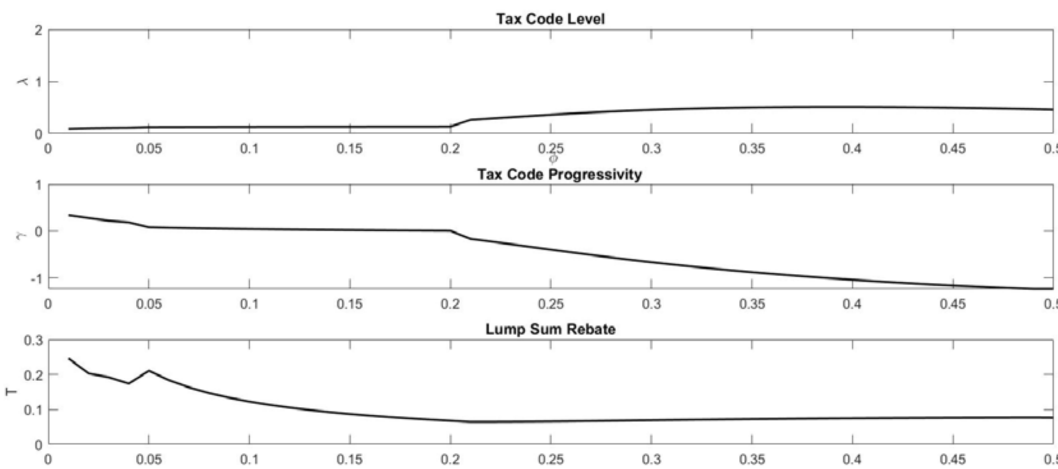

Figure 6: Welfare

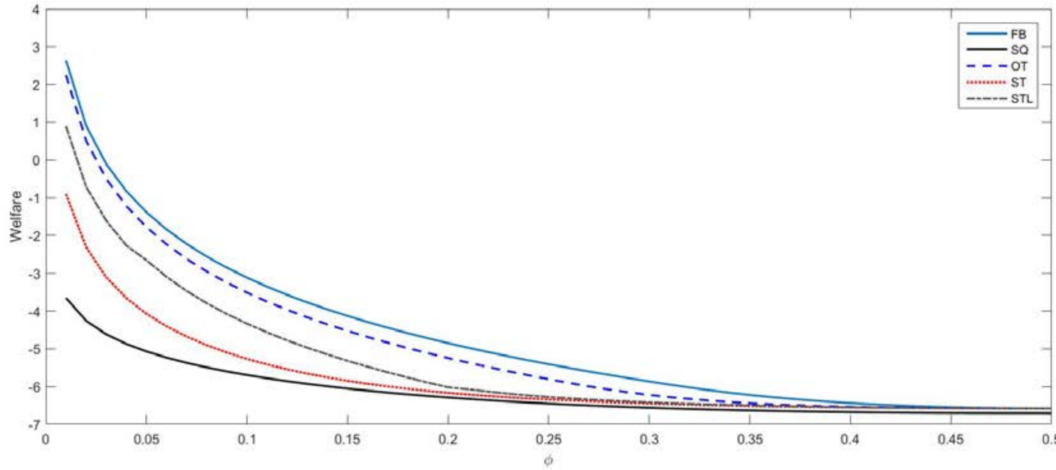

Figure 7: Consumption Equivalent - Panel A

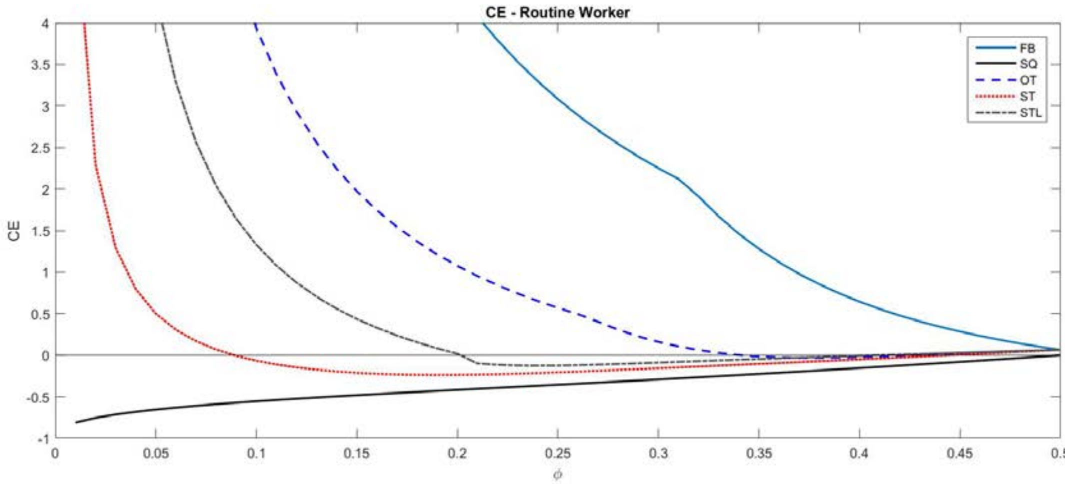

Figure 7: Consumption Equivalent - Panel B