Government and human capital in a model of development through modernization and specialization

Yuki, Kazuhiro

Kyoto University

December 2009

Government and human capital in a model of development through modernization and specialization

Kazuhiro Yuki*

December, 2009

Abstract

Economic development is associated with the shift of production from the traditional sector (e.g. traditional agriculture and the urban informal sector) to the modern sector (e.g. modern manufacturing and commercial agriculture). Human capital accumulation, particularly, education and job training of skilled workers, is a crucial factor in the modernization of an economy. Several institutions such as the protection of property rights and the strength of the rule of law also are considered essential. Thus, the government has an important role as the main provider of 'institution-maintaining' services, although it often faces a difficulty in providing adequate amounts of the services due to costly hiring of educated officers and tax avoidance.

This paper analyzes interactions among taxation, the provision of the public services, human capital accumulation, and modernization, based on a dynamic dual economy model, which draws on the Becker and Murphy (1992) model of skill and task specialization, and examines conditions for successful development. Distributions of political power and wealth as well as sectoral productivities and the cost of education affect the outcome qualitatively. In particular, the socially desirable distribution of political power is such that educated (uneducated) individuals should have dominant power at an early (late) stage of development. Further, it is shown that several novel or overlooked inefficiencies arise naturally from realistic features of the model and appropriate redistribution can correct the inefficiencies except at a fairly early stage of development.

Keywords: dual economy; government; human capital; inequality; overeducation; redistribution; specialization

JEL Classification Number: H23, O11, O15, O17.

*Faculty of Economics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan. Phone: +81-75-753-3532. E-mail address: yuki@econ.kyoto-u.ac.jp.

1 Introduction

Economic development is associated with the shift of production and employment from the traditional sector (e.g. traditional agriculture and the urban informal sector) to the modern sector (e.g. modern manufacturing and commercial agriculture). Since the modern sector employs advanced technologies and thus requires a greater proportion of skilled workers, human capital accumulation is a crucial factor in development. Job training of skilled workers as well as education are important in raising the sector's productivity, as suggested, for example, by the analysis of firm-level productivity in five developing economies by Tan and Barta (1996).

Recently, several institutions too have come to be recognized as fundamental determinants of development. For example, Rodrik et al. (2004) estimate relative contributions of institutions, geography, and trade in determining income levels of nations, and find that by far the most important is the quality of institutions, which is measured by a composite index (the rule of law index) developed by Kaufmann et al. (2002) to capture the protection of property rights, the strength of the rule of law, and the incidence of crime. Because these institutional measures reflect 'institution-maintaining' public services greatly, the finding suggests that the government has an important role as the main provider of such services.

The government in a developing economy, however, often faces a difficulty in providing adequate amounts of the services, because it needs to hire educated officers who are highly costly due to skill scarcity, and, if it imposes a high tax rate to raise enough revenue, economic activities escape to the traditional sector for tax avoidance.[1] Inadequate supplies of the services, by contrast, would result in low productivity, particularly, of the modern sector that relies much more on the services, a small size of the modern sector, and low returns to human capital investment. Thus, a choice of the tax rate, which would be affected by the distribution of political power over the population, are likely to be critical for the economy's fate.

This paper analyzes interactions among the above-mentioned factors – taxation, the provision of the governmental services, human capital accumulation, and modernization – employing a dynamic dual economy model, which draws on the Becker and Murphy (1992) model of skill and task specialization, and examines conditions for successful development. Distributions of political power and wealth as well as sectoral productivities and the cost of education affect the outcome qualitatively. In particular, the socially desirable distribution of political power is such that educated (uneducated) individuals should have dominant power at an early (late) stage of development. Further, it is shown that several novel or overlooked inefficiencies arise naturally from realistic features of the model and appropriate redistribution can correct the inefficiencies except at a fairly early stage of development.

The model is concerned with a small open economy that comprises up to two sectors producing the final good: the traditional sector (*sector T*) employing unskilled workers, and the modern sector

[1] Schneider (2005) finds that the tax and social security payment burden is among the most important factors affecting the size of the unofficial (informal) economy in both developing and developed countries.

(*sector M*) employing skilled as well as unskilled workers. More specifically, in sector M, the final good is produced using unskilled workers and the intermediate product, which in turn is produced by combining constant varieties of 'tasks' performed by skilled workers. Each task requires a task-specific skill developed through time-consuming training. As is stressed by Becker (1981) and Rosen (1983), since the development of task-specific skills exhibits increasing returns in nature, an increase in the degree of skill and task specialization among skilled workers raises the sector's productivity. However, with a higher degree of specialization, a greater number of skilled workers of distinct specialization must be involved in the production, which raises the cost of coordinating their activities. Hence, the degree of specialization is limited by the coordination cost, as in Becker and Murphy (1992).

Unlike Becker and Murphy, however, the government plays a role in reducing the coordination cost. The government imposes a value-added tax on sector M (sector T avoids taxation) and employs skilled workers to provide the cost-reducing service.[2] Real-life examples of the service include the maintenance of law and order, the establishment and enforcement of property rights, the regulation of economic activities in areas with non-negligible market failures, and, when market incompleteness is severe, the provision of credit and information stimulating market transactions. Qualitative results are not affected by including unproductive public services into the model.

The dynamic structure of the model is of an OLG variety. An individual, who is born identical to others in terms of abilities and preferences, lives for two periods. In childhood, she receives a transfer from the parent to invest in assets and education. Education is required to become a skilled worker, but its direct cost must be financed by the received transfer due to a lack of loan markets for the investment. In adulthood, she becomes a skilled or unskilled worker depending on the educational choice. (When she chooses a skilled job, she devotes a portion of time to develop task-specific skills.) Then, she receives labor and capital incomes and spends them on the consumption of the final good and a transfer to a single child, from which she derives utility (impure altruism). Generations go by in this fashion. Individuals of the same generation are heterogeneous in terms of education and wealth due to differences in received transfers and the credit constraint. The distribution of wealth determines the proportion of individuals *accessible* to education, and the proportion, the amount of the governmental service, and the tax rate determine the return to education and thus the proportion of individuals *taking* education (skilled workers) and the sectoral composition of production and employment.

In order to illuminate the dynamics of the economic structure, the simplest case is examined first, in which the tax rate is fixed and all lineages can access education eventually (through wealth accumulation) irrespective of the initial distribution of wealth. The tax rate affects the dynamics critically. If the rate is too high or too low, *nobody* takes education, sector M is not in operation, and output is lowest. When the tax rate is too low, the educational investment is not profitable because a lack of the governmental service results in limited specialization among skilled workers and thus low productivity in sector M. When the rate is too high, what makes education unrewarding is the

[2] The government does not provide services that directly affect consumers' utilities.

tax burden that more than offsets the resultant high degree of specialization and induces unskilled workers, who are complementary to skilled workers in sector M, to choose sector T for tax avoidance. In contrast, if the rate is in the intermediate range, the proportion of educated and thus skilled workers increases and production and employment shift from sector T to sector M over time. The growth of sector M raises tax revenue and the government too expands. Unskilled workers also shift to sector M, because higher numbers of skilled workers in sector M and in the government have, ceteris paribus, positive effects on the sector's productivity. After sector T ceases operation, the wage inequality between skilled and unskilled workers falls over time, and, in the long run, the skilled wage net of the education cost is equalized to the unskilled wage and the structural change ends.

In this economy, unless the tax rate is extreme, *overeducation is inevitable in the long run*: the proportion of educated individuals is higher than the socially optimal level (in terms of efficiency).[3] The reason is that only the educated benefit directly from employment opportunities at the government and thus the private return to education is higher than the social return. Overeducation occurs only at a late stage of development, since the social return is positive while not many can access education. The fact that only the educated benefit directly from governmental positions, together with the absence of taxation in sector T, also leads to *the oversized traditional sector* at an early stage.

Next, the tax rate is endogenized to examine effects of the distribution of political power (as before, all lineages can access education eventually). The rate is chosen by a politically influential group so as to maximize their incomes in each period.[4] Since educated and uneducated individuals have different stakes in the tax policy, the outcome when the educated choose the policy is contrasted with the other case. When only sector M is active (at an equilibrium tax rate), the rate selected by the educated (the uneducated) is higher (lower) than the socially optimal rate, because the educated overevaluate the contribution of the governmental service (thus taxation) on output and the uneducated underevaluate it. By contrast, when sector T too is in operation, the tax rate chosen by *the educated* is best in the competitive economy but is *lower* than the optimal rate, while the uneducated are indifferent among any rates (thus they may choose an extreme rate that forbids production in sector M). Considering this result and the fact that, under the dominance of the uneducated, inequality is lower and structural change when sector T is not active is faster (since the unskilled wage is higher), the socially desirable distribution of political power is such that the educated should have decisive power while sector T is active, i.e. at a relatively early stage of development, whereas the uneducated too should have influence on the policy at the later stage.

The preceding results show that both groups cannot choose the tax rate and thus the size of public service optimally, nor can they avoid overeducation at a late stage and oversized sector T at an early

[3]To be exact, the optimal level is defined to be the one maximizing aggregate labor income net of the education cost, which is a measure of aggregate (private) consumption for given aggregate assets.

[4] As for the timing of the policy decision, two cases are considered. In the *commitment case*, the tax rate that is optimal to them before education is completed is implemented in adulthood. By contrast, in the *non-commitment case*, the tax rate can be chosen after education and thus an ex-ante optimal rate may not be implemented. This paper mainly focuses on the commitment case because one of results of the other case is not robust under a more realistic setting.

stage. Things are different if redistribution is feasible. Suppose that the government can impose a proportional tax on wages generated in non-traditional sectors and use the revenue to provide a lump-sum transfer to the sectors' workers with political power.[5] Then, when only sector M is active without redistribution, *both groups choose the optimal value-added tax rate.* Intuitively, in order to take as much as possible from the other group through redistribution, the powerful group take into account effects of taxation not only on their wage but also on the opponent's wage. Further, *when the uneducated determine policies, overeducation can be avoided, and efficient and equitable allocation is realized in the long run.*[6] This is because redistribution from the educated to the uneducated corrects the excessive private return to education. In contrast, when sector T is active without redistribution, if the educated have power, redistribution is not implemented and the outcome is same as before, whereas the result changes greatly if the uneducated have power: if the educated are not very scarce, efficient allocation is attained with *sector T shut down* (redistribution corrects the insufficient return to choosing sector M) and the value-added tax set optimally (*higher* than the rate chosen by the educated), while otherwise, *sector M becomes oversized* and/or inefficiency may *worsen* compared to the economy without redistribution. Hence, the uneducated should have *decisive* power from an earlier stage than the previous economy, i.e. even when sector T is active without redistribution.

Finally, the case in which the initial distribution of wealth affects the long-run outcome, which is when the productivity of sector T is low relative to the education cost, is examined. Now choices of policies are even more critical because they determine, through effects on disposable incomes, whether descendants of each type of workers can access education or not. In the exogenous tax case, if the tax rate is extreme, *irrespective of* the initial distribution, nobody can access education and only sector T is in operation in the long run. Otherwise, when the initial distribution is such that only a small portion of individuals can afford education at the beginning and/or the tax rate is high or low, skilled workers are limited in number, the government is small, and the wage inequality persists in the long run. When the initial distribution and the tax rate are appropriate, by contrast, the economy succeeds in complete modernization and the inequality disappears eventually.[7] Depending on the tax rate, the long-run outcome can be very different even when the initial distribution of wealth is *identical.* Because income levels of unskilled workers are critical for successful structural change, in the endogenous tax case, the uneducated should control the policy from a lower stage of development than the economy in which the initial distribution does not matter. For example, it is possible that the

[5]The transfer to the educated may be interpreted as benefits and privileges associated with pollical power. When only non-targeted transfer is available, results when the educated (the uneducated) have power are same as the original economy without (with) redistribution.

[6]While the social return to education is non-negative, efficient allocation is realized when the educated choose policies too. However, overeducation arises eventually: in the commitment case (footnote 4), redistribution is not implemented and the outcome is same as before at the last stage, and overeducation worsens in the non-commitment case.

[7]Consistent with the model's implications, Deininger and Olinto (2000) find that an economy's growth rate is affected negatively by initial land inequality (a proxy for initial wealth inequality) and positively by its mean years of schooling per working person, which in turn is negatively affected by the initial inequality. Easterly (2007) finds that higher structural inequality, which he claims reflects such historical events as conquest, colonization, slavery, and land distribution by the state or colonial power, leads to a lower level of development, worse institutions, and less education.

economy remains stagnant under the political dominance of the educated but it succeeds in structural change under the dominance of the uneducated. In particular, when the educated have power and can employ a redistributive policy, the economy *inevitably* stagnates.

Main contributions of the paper are as follows. First, as mentioned earlier, it analyzes interactions among taxation, the provision of the public service, human capital accumulation, and modernization based on a dynamic dual economy model, and examines conditions for successful development. Further, effects of distributions of political power and wealth on the outcome are examined. There exist papers that examine related issues, but this paper studies aspects unexplored by them, as detailed below.

Similar to this work, Acemoglu (2005) examines how economic performance is related to the government's ability to tax and provide public services and finds that both strong and weak governments result in low output. Based on the model populated by citizens and a selfish ruler who determines policies to maximize rents (tax revenue minus expenditure on public services),[8] effects of two kinds of exogenous governmental abilities, economic (the ability to keep citizens from evading taxation) and political (the ability to avoid replacement by citizens), are analyzed. He is particularly interested in the political ability and thus dynamic interplay between rulers and citizens, hence, for simplicity, many elements of the model are reduced-form and analyses are limited to steady states. This paper, by contrast, is interested in interactions with human capital accumulation and modernization, thus it constructs a more structural model and analyzes dynamics, but does not model transitions of government and thus does not examine effects of the political ability.[9]

Acemoglu (2008) constructs a model where individuals with high or low entrepreneurial skill become entrepreneurs or workers and the government chooses redistributive taxation and the fixed cost to start a business incurred by new entrepreneurs. He examines trade-off between oligarchy (rule by incumbent entrepreneurs) and democracy (rule by workers) and is particularly interested in dynamic inefficiency of the entry barrier due to time-varying entrepreneurial skill. The present paper compares rule by the educated and rule by the uneducated in a model where individuals with homogeneous innate ability but heterogeneous wealth decide on educational investment and the government chooses the amount of productive service as well as redistributive taxation. Both papers examine effects of the distribution of political power on investment and sectoral allocations of individuals, but the Acemoglu's focus is on the misallocation of ability, while this paper is interested in how the effects change with development and how they interact with the distribution of wealth.

Besley and Persson (2009) consider a two-period economy in which the government invests in legal capacity that contributes to the private sector's output and fiscal capacity that raises the ability to tax, provides public goods, may redistribute incomes, and determines income tax rates. The government is controlled by one of two groups of individuals and can treat the groups differentially in policies, and

[8] He finds that the tax rate preferred by the ruler (citizens) is higher (lower) than the optimal rate, which is similar to the this paper's result that the rate preferred by the educated (the uneducated) is higher (lower) than the optimal rate.

[9] The economic ability of the government is endogenous in the sense that the propensity of unskilled workers to escape for sector T with higher tax is related to the proportion of individuals accessible to education.

the transition of power can occur exogenously between the periods. They examine what lead to good provision of the state capacities and show that it is related to the presence of public goods highly valued by both groups, as well as political stability and more representative political institutions. They model government in greater detail than the present paper and Acemogul's papers and focus on interactions among different governmental functions, while this paper is interested in interactions of government with human capital accumulation and modernization. Further, this paper is interested in sectoral allocations of individuals and examines issues related to hiring of educated officials.

The second contribution is that the paper shows that the oversized traditional sector at an early stage and overeducation at a late stage of development and the inefficient level of the public service (when the tax rate is endogenous) all arise naturally from two realistic features of the model – the exclusive access of skilled workers to governmental positions and, except for overeducation, the absence of taxation in sector T – and that appropriate redistribution can correct these inefficiencies except at a fairly early stage. Bourguignon and Verdier (2000) examine the dynamics of education, inequality, and development in a model where education has positive externality and the educated have political power to choose redistribution to the poor. (They may implement costly redistribution because of the externality.) Redistribution raises efficiency since, as in this paper, it enables credit constrained individuals to access education, but it does not have the above-mentioned roles.

Third, it shows that the desirable distribution of political power changes with development and is affected by factors such as the availability of a redistributive policy and the productivity of sector T relative to the education cost. The educated should have decisive power at an early stage of development mainly because the size of the public service is more efficient, while the uneducated should have power at a late stage primarily because inequality is lower, structural change is promoted through education of the poor, and, when redistribution is possible, overeducation can be prevented. If rule by the uneducated is interpreted as democracy, one of the implications is consistent with the finding by Baum and Lake (2003) that, in non-poor countries, democracy raises economic growth through increased enrollment rates of secondary education, and the one by Doucouliagos and Ulubaşoğlu (2008), who perform meta-regression analysis on existing 84 studies on the democracy-growth relationship, that democracy increases growth through human capital accumulation.

Aside from papers cited earlier, this paper is related to the literature that examines the relationship between skill and task specialization and growth. The seminal work by Becker and Murphy (1992) emphasizes the dependence of specialization on the coordination cost and knowledge, and investigates its implications for economic growth and industrial organization. In particular, they examine interactions among knowledge accumulation, specialization, and growth using a one-sector growth model, and explore the division of labor between the consumption good sector and the human capital production sector in a two-sector growth model. This paper is indebted to their work in modeling specialization among skilled workers engaged in the production of the intermediate product. However, distinct from theirs, the paper distinguishes the sector that has productivity gains from specialization (sector M) from the sector without gains (sector T), models development as the shift from the latter sector to the former, and examines the role of the cost-reducing public service in development.

Other works in the literature that draw on Becker and Murphy include Tamura (1996), Davis (2003), and Yuki (2006), of which closely related are Davis (2003) and Yuki (2006). In order to explain the increasing importance of governmental activities in advanced nations over the past century, Davis (2003) extends the one-sector growth model of Becker and Murphy so that the coordination cost depends on governmental expenditures, similarly to the present paper. However, it does not consider roles of the sectoral shift, the credit constraint, and the distribution of political power in development. Yuki (2006) examines the interplay among the extent of market (the size of population), task specialization, and development, employing a model similar to the present paper, but considers neither the role of government nor the effect of the distribution of political power.

The modeling of the educational decision and intergenerational transmission of wealth draws on the literature that examines the interplay between income distribution and growth through human capital accumulation, including Galor and Zeira (1993), Ljungqvist (1993), Benabou (1996), and Yuki (2007, 2008). Closely related are Galor and Zeira (1993) and Yuki (2007, 2008), in which, as in this paper, the educational investment is constrained by intergenerational transfers motivated by impure altruism. Neither papers do not consider issues analyzed here.

The paper is organized as follows. Section 2 presents the static part of the model, and Section 3 integrates the static part into the dynamic part and derives critical conditions for examining the dynamics. Section 4 analyzes the model and derives results, and Section 5 concludes. Proofs of lemmas and propositions except Lemma 9 and Proposition 6, which are straightforward, are in Appendix.

2 Static model

This section presents the static part of the model, that is, production decisions of firms and workers and the determination of output and wages *given* the numbers of skilled and unskilled workers (all variables are presented without time subscript). The full-fledged model is presented in the next section.

2.1 Final good production

There exist up to two private sectors, T (*traditional*) and M (*modern*), both producing the same final good. Sector T hires unskilled workers and sector M hires both skilled and unskilled workers. In real economy, sector T corresponds to sectors such as traditional or subsistence agriculture and the urban informal sector, and sector M corresponds to sectors such as modern manufacturing and agriculture.

The production function of a representative firm of sector T is given by

$$Y_T = A_T N_{LT}, \tag{1}$$

where Y_T is the output, A_T is the productivity, and N_{LT} is the number of unskilled workers of the sector. A representative final good producer of sector M has the following production function:

$$Y_M = A_M {N_{LM}}^{\alpha} {I_H}^{1-\alpha}, \tag{2}$$

where definitions of Y_M, A_M, and N_{LM} are same as corresponding symbols of sector T, and I_H is the input of the intermediate product produced by skilled workers, the determination of which is explained below. Final good producers of both sectors behave competitively in goods and factors markets.

2.2 Intermediate product

The production technology of the intermediate product is the one employed by Becker and Murphy (1992) for their final good. The intermediate product is produced with inputs of a continuum (measure 1) of different kinds of *tasks* using the Leontief technology:

$$Y_H = min_{\,0 \leq s \leq 1} \{ Y_H(s) \}, \tag{3}$$

where Y_H is the *gross* output of the intermediate product and $Y_H(s)$ is the outcome of task s $(0 \leq s \leq 1)$. The specification implies that every task is equally essential in the production of Y_H, which tries to capture the fact that, in modern sectors, many different tasks, each of which is difficult to be substituted for others, must be combined to yield final output.

The outcome of task s is, in turn, a linear function of the effective labor of skilled workers engaged in the task:

$$Y_H(s) = \int h^i(s) l_y^i(s) di, \tag{4}$$

where $l_y^i(s)$ is the time spent on the production activity of the task by skilled worker i and $h^i(s)$ is the amount of her *task-specific human capital*, which is developed with her time input, $l_h^i(s)$:

$$h^i(s) = \gamma \big[l_h^i(s) \big]^{\theta}, \quad \theta > 0. \tag{5}$$

The specification reflects the fact that tasks performed by skilled workers require highly specialized skills and thus substantial time needs to be spent to build up such skills. Task-specific skill production is not modeled for unskilled workers, because, in real economy, unskilled jobs require much less investment in specific skills to perform their tasks satisfactorily.[10]

Let $l^i(s)$ be the total time spent on the task by the worker. She allocates the time between the development of the specific skill and the production activity to maximize the outcome:

$$max_{\{l_y^i(s), l_h^i(s)\}} \; h^i(s) l_y^i(s) = \gamma \big[l_h^i(s) \big]^{\theta} l_y^i(s), \quad s.t. \quad l_y^i(s) + l_h^i(s) = l^i(s). \tag{6}$$

[10] Tan and Batra (1996) find that the incidence of formal training is positively associated with mean education and the proportion of skilled workers in an enterprise's workforce in Columbia, Indonesia, Malaysia, Mexico, and Taiwan. Further, they find that formal training of skilled workers has a positive and significant impact on firm-level productivity, while the effect of training of unskilled workers is statistically insignificant. Asplund (2005) notes that, in European countries, the incidence of company-provided training is considerably lower in low-skill/low-pay industries, even after controlling for a large set of personal and employer characteristics.

From the maximization problem, the outcome equals

$$h^i(s)l_y^i(s) = \Gamma[l^i(s)]^{1+\theta}, \quad \text{where} \quad \Gamma \equiv \frac{\gamma\theta^\theta}{(1+\theta)^{1+\theta}}. \tag{7}$$

Notice that the outcome exhibits increasing returns to the total time spent on the task. This is because the return to invest in task-specific skill increases with the time spent on the production activity, as is identified by Becker (1981) and Rosen (1983).

2.3 Coordination cost

The total output of the intermediate product is maximized when each task is performed by a single worker. Because skilled workers are identical in terms of ability before receiving job trainings, the maximum output would be attained when each skilled worker performs an equal measure of distinct tasks. However, such outcome is *not* realized due to various costs of coordinating activities of workers with distinct skills, such as the cost of processing information among specialized workers and the cost of enforcing commands (if they belong to the same firm) or contracts (if they belong to different firms) under information asymmetry and incompleteness.

Assume that these coordination problems result in the loss of the intermediate product and, following Becker and Murphy (1992), the total coordination cost associated with the problems depends on the size of a *production team*, defined as a group of skilled workers with *distinct* specialization that is just enough to produce the intermediate product. Unlike Becker and Murphy, however, the government plays a role in reducing the coordination cost. The government imposes a value-added tax on sector M and employs skilled workers to provide the cost-reducing service.[11] Real-life examples of the service include the maintenance of law and order, the establishment and enforcement of property rights, contract law, and proper market regulations, and, when market incompleteness is severe, the provision of credit and information stimulating market transactions. Note that *unproductive public services* can be easily included into the model without affecting results qualitatively (see footnote 14 in the next subsection). Sector T is assumed to avoid taxation, reflecting the fact that small and often unregistered enterprises and farmers in urban informal and rural sectors are difficult to be taxed in developing economies (Burgess and Stern, 1993).

The coordination cost of a production team is given by

$$\widehat{C} = \lambda S^{1+\delta}(\widehat{N}_G)^{-\rho}, \tag{8}$$

where S is the size of the team, that is, the minimum number of skilled workers each of whom is engaged in different tasks and, as a group, can produce the intermediate product, and $\widehat{N}_G$ is the number of governmental skilled workers *per team*. Note that the coordination cost increases more than proportionally with the team's size, and the cost-reducing service is provided exclusively to the team and does not spill over to other teams. Let N_{HM} and N_G be the total number of skilled workers

[11] It does not provide services that directly affect consumers' utilities.

in sector M and in the government, respectively. Because the number of production teams is $\frac{N_{HM}}{S}$, $\widehat{N}_G = \frac{N_G}{N_{HM}/S}$ is satisfied and the cost function can be expressed as

$$\widehat{C} = \lambda S^{1+\delta-\rho}\left(\tfrac{N_G}{N_{HM}}\right)^{-\rho}. \tag{9}$$

2.4 Degree of task specialization and output of sector M

Now the determination of S is considered. Each skilled worker has total time of 1 to allocate among various tasks, i.e. $\int l^i(s)ds = 1$. Because the total variety of tasks is measure 1, when the size of a production team is S and workers are allocated to maximize the team's output, each worker performs $1/S$ of the tasks and thus the total time spent on each task is S. Hence, from (4), (7), and (9), the *net* output of the intermediate product produced by a team of size S is

$$\Gamma S^{1+\theta} - \lambda S^{1+\delta-\rho}\left(\tfrac{N_G}{N_{HM}}\right)^{-\rho}. \tag{10}$$

Skilled workers in sector M form teams (choose S) so that the net output per worker is maximized:[12]

$$S = \left[\tfrac{\theta\Gamma}{\lambda(\delta-\rho)}\right]^{\frac{1}{\delta-\rho-\theta}}\left(\tfrac{N_G}{N_{HM}}\right)^{\frac{\rho}{\delta-\rho-\theta}}. \tag{11}$$

It is assumed, for simplicity, that the market framework of the intermediate product subsector is such that this allocation of skilled workers across tasks is realized in a decentralized manner.[13]

Since the number of teams is $\frac{N_{HM}}{S}$, the aggregate net output of the intermediate product is

$$I_H = N_{HM}\left[\Gamma S^{\theta} - \lambda S^{\delta-\rho}\left(\tfrac{N_G}{N_{HM}}\right)^{-\rho}\right]. \tag{12}$$

By substituting (12) into (2), the total output of the final good in sector M equals

$$Y_M = A_M\left[\Gamma S^{\theta} - \lambda S^{\delta-\rho}\left(\tfrac{N_G}{N_{HM}}\right)^{-\rho}\right]^{1-\alpha} {N_{LM}}^{\alpha} {N_{HM}}^{1-\alpha}. \tag{13}$$

From (11), (13), and the CRS production function of sector T, the per capita output of the final good depends on the sectoral distribution of workers, but *not* on the size of labor force, which is henceforce normalized to be 1. Denote the *proportions* of sector i skilled workers, sector j unskilled workers, and total skilled workers in the labor force by H_i $(i = M, G)$, L_j $(j = M, T)$, and H, respectively, where $H_M + H_G = H$ and $L_M + L_T = 1 - H$.

The size of a production team is rewritten as

[12] To be more accurate, $S = N_{HM}$, if the right-hand side of (11) is greater than N_{HM}. From equations such as (23) below, it can be shown that this happens if the number of skilled workers is very small or the tax rate is very high. Because the situation where every skilled worker is engaged in distinct tasks is empirically unlikely, it is assumed that the population size is large enough that this situation is negligible. Further, as will be shown in Lemma 2 of Section 3.2, when the tax rate is very high, nobody becomes a skilled worker and thus this situation does not arise in an equilibrium (under appropriate assumptions on values of parameters and exogenous variables). See footnote 15 also.

[13] For example, suppose that each production team is a firm that produces the intermediate product and sells it to final good producers in sector M. It hires skilled workers and allocates them across tasks so as to maximize profits. Then, from profit-maximizing and free entry conditions, the number of hired skilled workers, S, is given by (11).

$$S = \left[\frac{\theta\Gamma}{\lambda(\delta-\rho)}\right]^{\frac{1}{\delta-\rho-\theta}} \left(\frac{H_G}{H_M}\right)^{\frac{\rho}{\delta-\rho-\theta}}. \tag{14}$$

Remember that S is also the total time that a skilled worker spends on each of her tasks. Thus, hereafter S is called *the degree of (task) specialization*, which increases with $\frac{H_G}{H_M}$. By substituting (14), $N_{HM} = H_M$, and $N_G = H_G$ into (12) and dividing the resulting expression by H, the net output of the intermediate product *per skilled worker* is[14]

$$\Omega\left(\tfrac{H_G}{H}\right) \equiv \Omega_0 \left(\tfrac{H_G}{H}\right)^{\frac{\rho\theta}{\delta-\rho-\theta}} \left(1 - \tfrac{H_G}{H}\right)^{1-\frac{\rho\theta}{\delta-\rho-\theta}}, \tag{15}$$

$$\text{where} \quad \Omega_0 \equiv \left\{\Gamma^{\delta-\rho}\left[\tfrac{\theta}{\lambda(\delta-\rho)}\right]^{\theta}\right\}^{\frac{1}{\delta-\rho-\theta}} \left(1 - \tfrac{\theta}{\delta-\rho}\right). \tag{16}$$

Then, I_H and Y_M are conveniently expressed as:

$$I_H = \Omega\left(\tfrac{H_G}{H}\right) H, \tag{17}$$

$$Y_M = A_M {L_M}^{\alpha} \left[\Omega\left(\tfrac{H_G}{H}\right) H\right]^{1-\alpha}. \tag{18}$$

2.5 Wages

The government imposes a proportional tax of rate $\tau \in [0, 1]$ on the value added generated in sector M. When the sector is in operation, from (18), the unskilled wage equals

$$w_l = (1-\tau)\alpha A_M \left[\Omega\left(\tfrac{H_G}{H}\right) \tfrac{H}{L_M}\right]^{1-\alpha}. \tag{19}$$

The remaining after-tax income goes to skilled workers, hence the skilled wage equals

$$w_h = \tfrac{(1-\tau)(1-\alpha)}{H-H_G} A_M {L_M}^{\alpha} \left[\Omega\left(\tfrac{H_G}{H}\right) H\right]^{1-\alpha}. \tag{20}$$

From (1), $w_l = A_T$ when $L_T > 0$, which is true iff the RHS of (19) at $L_M = 1 - H$ is less than A_T, i.e.

$$H < \left\{1 + \left[\tfrac{(1-\tau)\alpha A_M}{A_T}\right]^{\frac{1}{1-\alpha}} \Omega\left(\tfrac{H_G}{H}\right)\right\}^{-1}. \tag{21}$$

When $w_l = A_T$, the skilled wage (if sector M is in operation) becomes

$$w_h = (1-\alpha) A_M \left(\tfrac{\alpha A_M}{A_T}\right)^{\frac{\alpha}{1-\alpha}} (1-\tau)^{\frac{1}{1-\alpha}} \Omega\left(\tfrac{H_G}{H}\right) \tfrac{H}{H-H_G}. \tag{22}$$

The wages can be expressed as functions of the tax rate. Since the governmental budget constraint is $w_h H_G = \tau Y_M$, from (18) and (20),

$$\frac{H_G}{H} = \frac{\tau}{1-\alpha(1-\tau)}. \tag{23}$$

From the above equation, (15), and (22), the skilled wage when $L_T > 0$ is

$$w_h = (1-\alpha) A_M \Omega_0 \left(\tfrac{\alpha A_M}{A_T}\right)^{\frac{\alpha}{1-\alpha}} (1-\tau)^{\frac{1}{1-\alpha}} \left[\tfrac{\tau}{(1-\alpha)(1-\tau)}\right]^{\frac{\rho\theta}{\delta-\rho-\theta}}. \tag{24}$$

[14] Unproductive public services can be included into the model without affecting qualitative results. Suppose that only a fraction $\phi \in (0, 1)$ of governmental officers are engaged in the productive service. Then, the model is same as before except that $\Omega(\frac{H_G}{H})$ is replaced by $\phi^{\frac{\rho\theta}{\delta-\rho-\theta}} \Omega(\frac{H_G}{H})$. (The first $\frac{H_G}{H}$ of eq. 15 is now $\frac{\phi H_G}{H}$ from eqs. 12 and 14.)

Similarly, the skilled and unskilled wages when $L_T = 0$ (and thus $L_M = 1 - H$) are

$$w_h = A_M \Omega_0{}^{1-\alpha}[1-\alpha(1-\tau)]\left(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right)^{1-\alpha}\left(\frac{1-H}{H}\right)^{\alpha}. \tag{25}$$

$$w_l = A_M \Omega_0{}^{1-\alpha}\alpha(1-\tau)\left(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right)^{1-\alpha}\left(\frac{H}{1-H}\right)^{1-\alpha}, \tag{26}$$

where $\frac{\rho\theta}{\delta-\rho-\theta} < 1$ is assumed.[15] The next lemma summarizes relations between the wages and τ.

Lemma 1 *(i) When $L_T = 0$, there exists a single τ satisfying $\frac{\partial w_l}{\partial \tau} = 0$, τ_l, and $\frac{\partial w_l}{\partial \tau} \gtreqless 0$ for $\tau \lesseqgtr \tau_l$. Similar statements hold for w_h when $L_T > 0$ and w_h when $L_T = 0$ as well.*

(ii) Let τ satisfying $\frac{\partial w_h}{\partial \tau} = 0$ when $L_T > 0$ be $\overline{\tau}_h$ and when $L_T = 0$ be τ_h, respectively. Then, $\tau_h > \overline{\tau}_h = \frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta} > \tau_l$.

3 Dynamics

This section integrates the production decisions presented in the previous section into the dynamic part of the model. Consider a discrete-time small open OLG economy. In the economy, there exists a continuum of individuals who are homogeneous in terms of innate abilities and preferences and live for two periods. There is no uncertainty in the model.

3.1 Lifetime of an individual

Childhood: In childhood, an individual receives a transfer from her parent and spends it on two investment options, assets (which yields interest rate r) and education (which costs e but enables her to become a skilled worker in adulthood), in order to maximize future income. The educational investment must be self-financed because loan markets for such investment are not available. Consider an individual born into lineage i in period $t-1$ (*generation t*) who receives b_t^i units of transfer and can allocate it between asset a_t^i and education e_t^i. If the return from education is strictly higher than the one from assets, the allocation is determined by b_t^i:

$$a_t^i = b_t^i, \qquad e_t^i = 0, \quad \text{if} \quad b_t^i < e, \tag{27}$$

$$a_t^i = b_t^i - e, \; e_t^i = e, \quad \text{if} \quad b_t^i \geq e_t. \tag{28}$$

Adulthood: At the beginning of adulthood, an individual makes an occupational choice based on the educational investment. Then, she obtains income from assets and labor supply and spends it on consumption c_t^i and a transfer to her single child b_{t+1}^i. Her utility maximization problem is:

$$max_{\{c_t^i, b_{t+1}^i\}}\, u_t^i = \left(c_t^i\right)^{1-\gamma_b}\left(b_{t+1}^i\right)^{\gamma_b}, \quad s.t. \quad c_t^i + b_{t+1}^i = w_t^i + (1+r)a_t^i, \tag{29}$$

[15] For S to be always given by (14), $\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta} < 1$ must be assumed: if $\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta} \geq 1$, w_h when $L_T > 0$ increases with τ as long as S is given by (14) (see equation 24), and thus it is maximized when τ is high enough that $S = N_{HM}$ holds (see footnote 12), which means that, when τ is very high, $S = N_{HM}$ can be an equilibrium. The stronger assumption $\frac{\rho\theta}{\delta-\rho-\theta} < 1$ is imposed so that w_h when $L_T = 0$ does not increase with τ monotonically.

where w_t^i is her wage and $\gamma_b \in (0,1)$. By solving the problem, her consumption and transfer equal

$$c_t^i = (1-\gamma_b)\{w_t^i + (1+r)a_t^i\}, \tag{30}$$

$$b_{t+1}^i = \gamma_b\{w_t^i + (1+r)a_t^i\}. \tag{31}$$

Generational change: At the beginning of period $t+1$, current adults pass away, current children become adults, and new children are born into the economy. Since each adult has one child, the population of each generation is time-invariant and normalized to be one.

3.2 Determination of sectoral and skill distributions of workers

Since individuals must self-finance the education cost, only those who received transfers greater than e can access education. Let the fraction of such individuals in *generation t* (born in period $t-1$) be Fr_t. Further, for them to actually take education, education must be *profitable*, i.e. $w_{h,t} - (1+r)e \geq w_{l,t}$ must hold. The following assumption is imposed to ensure that it is profitable when the unskilled wage is lowest (equals A_T) and the skilled wage is highest (at $\tau = \overline{\tau}_h$ from Lemma 1).

Assumption 1 $(1-\alpha)^{1-\frac{\rho\theta}{\delta-\rho-\theta}} A_M \big(\frac{\alpha A_M}{A_T}\big)^{\frac{\alpha}{1-\alpha}} \Omega_0 (1-\overline{\tau}_h)^{\frac{1}{1-\alpha}} \big[\frac{\overline{\tau}_h}{1-\overline{\tau}_h}\big]^{\frac{\rho\theta}{\delta-\rho-\theta}} > A_T + (1+r)e.$

The next lemma shows that, when the tax rate is very high or very low, *nobody* takes education and becomes a skilled worker and thus sector M is not in operation. Since all results in this subsection are quasi-static (variables of different generations do not coexist in equations), variables are presented without time subscript for simplicity.

Lemma 2 *There exist τ_0^s and τ_0^b satisfying $\tau_0^s < \overline{\tau}_h < \tau_0^b$,[16] such that, for $\tau < \tau_0^s$ and $\tau > \tau_0^b$, $H = 0$ and thus sector M is not in operation.*

A change in the tax rate has three effects on the skilled wage, which is expressed as $w_h = (1-\tau)(1-\alpha)A_M\big(\frac{L_M}{H_M}\big)^{\alpha}\big[\Omega\big(\frac{H_G}{H}\big)\frac{H}{H_M}\big]^{1-\alpha}$ from (20): the *direct taxation effect*, the *productivity effect*, and the *worker ratio effect*. Through the direct taxation effect, higher τ lowers w_h. The productivity effect works through a change in the productivity of sector M, $\Omega\big(\frac{H_G}{H}\big)\frac{H}{H_M}$. Higher τ raises $\frac{H_G}{H}$ (equation 23) and thus the amount of available governmental service, which reduces the coordination cost and raises the degree of task specialization S (eq. 14). Higher S in turn raises the sector's productivity and thus w_h. In aggregate, through the two effects, higher τ has a positive (negative) influence on w_h when $\tau < (>)\, \overline{\tau}_h$. Finally, the worker ratio effect operates through a change in $\frac{L_M}{H_M}$: if higher τ raises (lowers) $\frac{L_M}{H_M}$, it acts on w_h positively (negatively).

When $L_T > 0$, the sign of the worker ratio effect is same as the total impact of the first two effects,[17] hence, a tax increase raises (lowers) the skilled wage for $\tau < (>)\, \overline{\tau}_h$ (Lemma 1). Lemma 2

[16] Superscripts s and b are for 'small' and 'big', respectively. The superscripts will be used for other variables for the same meanings.

[17] The worker ratio effect depends on the total impact of the first two effects on the *un*skilled wage, which is qualitatively same as the impact on w_h: when $\tau < (>)\, \overline{\tau}_h$, higher τ has a positive (negative) effect on the sector M's unskilled wage through the two effects, thus $\frac{L_M}{H_M}$ must rise (fall) to keep the wage constant at $w_l = A_T$ (sector T cannot be taxed).

shows that, when $\tau < \tau_0^s$ or $\tau > \tau_0^b$, the wage is lowered to the point that education is unprofitable and all individuals work in sector T.[18]

The rest of the subsection examines sectoral and skill distributions of workers when the tax rate is in the intermediate range, i.e. $\tau \in [\tau_0^s, \tau_0^b]$. As long as $L_T > 0$ is satisfied, from (24) and Lemma 2, $w_h - (1+r)e \geq w_l = A_T$ holds for any H, so all individuals who can afford education take it and become skilled workers, i.e. $H = Fr$. Since $L_T > 0$ is satisfied iff $w_l < A_T$ holds with $L_T = 0$, from (21), (15), (23), and $H = Fr$, the dividing line between the case $L_T > 0$ and the case $L_T = 0$ is:

$$Fr = \overline{H}(\tau) \equiv \left\{1 + \left[\tfrac{(1-\tau)\alpha A_M}{A_T}\right]^{\frac{1}{1-\alpha}} \Omega_0 \frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}} [(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right\}^{-1}. \quad (32)$$

When $Fr < \overline{H}(\tau)$, $L_T > 0$, where $w_l = A_T$ and thus w_h is independent of $H = Fr$ (see eq. 24), and when $Fr \geq \overline{H}(\tau)$, $L_T = 0$. The next lemma examines the relation between Fr and τ satisfying (32), that is, the shape of the dividing line on the (Fr, τ) plane (see Figure 1 below).

Lemma 3 *On the (Fr, τ) plane, the dividing line between $L_T > 0$ and $L_T = 0$, $Fr = \overline{H}(\tau)$, is defined for $\tau \in [\tau_0^s, \tau_0^b]$, positively sloped for $\tau > max\{\tau_l, \tau_0^s\}$, and, when $\tau_l > \tau_0^s$, negatively sloped for $\tau < \tau_l$.*

From its definition, the shape of the line reflects the three effects of taxation explained above on the unskilled wage when $L_T = 0$, which is expressed as $w_l = (1-\tau)\alpha A_M \left[\Omega\left(\frac{H_G}{H}\right)\frac{H}{H_M}\frac{H_M}{L_M}\right]^{1-\alpha}$ from (19). As before, higher τ acts on w_l negatively through the direct taxation effect and positively through the productivity effect. By contrast, w_l is always negatively affected through the worker ratio effect: higher τ raises $\frac{H_G}{H}$ and thus $\frac{L_M}{H_M} = \frac{1-H}{H-H_G}$. When $\tau > \tau_l$, the productivity effect is dominated and w_l decreases with τ, while the opposite happens when $\tau < \tau_l$ (Lemma 1 (i)) and thus the association between τ and $\overline{H}(\tau)$ is as stated in the lemma.[19]

When $L_T = 0$, i.e. $Fr \geq \overline{H}(\tau)$, the net return to education, $w_h - (1+r)e - w_l$, decreases with Fr and becomes negative when Fr is high enough. This is because, from (25) and (26), w_h decreases and w_l increases with $H = Fr$, $w_h = +\infty$ and $w_l = 0$ at $Fr = 0$, and $w_h = 0$ and $w_l = +\infty$ at $Fr = 1$. Thus, if $w_h - (1+r)e < w_l$ is satisfied with $H = Fr$, in an equilibrium, only some of those who can access education take it and $w_h - (1+r)e = w_l$ holds. The case in which $w_h - (1+r)e > w_l$ holds with $H = Fr$ is called the *unequal opportunity case*, while the case in which $w_h - (1+r)e \leq w_l$ is satisfied with $H = Fr$ (and thus $H \leq Fr$ in an equilibrium) is called the *equal opportunity case*.[20] From (25) and (26), the dividing line between the two cases is

[18] The argument is concerned only with the case $L_T > 0$, because $L_T = 0$ is not possible for $\tau < \tau_0^s$ and $\tau > \tau_0^b$: given τ, whenever $w_h - (1+r)e < w_l$ holds with $L_T > 0$, it does with $L_T = 0$.

[19] Qualitatively, effects of taxation on the size of L_T (when $L_T > 0$) are same as those on $\overline{H}(\tau)$ (and opposite to effects on w_l when $L_T = 0$), which is clear from the fact that L_T is obtained by equating (26) (with $1 - H$ replaced by $1 - H - L_T$) with A_T. Thus, the size of sector T is affected positively by the direct taxation effect and negatively by the productivity effect. Consistent with this, Friedman et al. (2000) find more over-regulation and greater corruption, both of which lower the productivity effect (see footnote 14), are associated with a greater unofficial economy.

[20] The name of the former case is from the fact that the rate of return from education exceeds the interest rate and access to such profitable investment opportunity is constrained by received transfers.

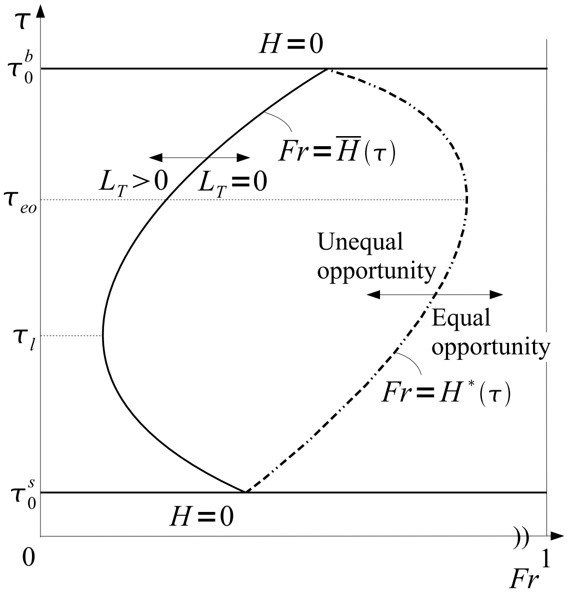

Figure 1: Positions of the critical loci when $\tau_l > \tau_0^s$ and $\tau_{eo} < \tau_0^b$

$$A_M \Omega_0{}^{1-\alpha}\left(\tfrac{1-Fr}{Fr}\right)^{\alpha}\left(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right)^{1-\alpha}\left(1-\tfrac{\alpha(1-\tau)}{1-Fr}\right) = (1+r)e. \tag{33}$$

When $\tau \in (0,1)$, the LHS of the equation decreases with Fr, equals $+\infty$ at $Fr = 0$, and equals $-\infty$ at $Fr = 1$, thus, for any $\tau \in (0,1)$, there exists a single Fr satisfying (33). Denote the dividing line by $Fr = H^*(\tau)$. The next lemma presents the relation between Fr and τ satisfying the equation.

Lemma 4 *On the $(Fr,\ \tau)$ plane, the dividing line between the unequal and equal opportunity cases, $Fr = H^*(\tau)$, is defined for $\tau \in [\tau_0^s, \tau_0^b]$, positively sloped for $\tau < min\{\tau_{eo}, \tau_0^b\}$, where $\tau_{eo} > \tau_h$, and, when $\tau_{eo} < \tau_0^b$, negatively sloped for $\tau > \tau_{eo}$. Further, it intersects with $Fr = \overline{H}(\tau)$ at $\tau = \tau_0^s, \tau_0^b$.*

The shape of the dividing line reflects effects of taxation on $w_h - w_l$ when $L_T = 0$, which is expressed as $w_h - w_l = (1-\tau)A_M\left[\Omega\left(\frac{H_G}{H}\right)\frac{H}{H_M}\right]^{1-\alpha}\left[(1-\alpha)\left(\frac{L_M}{H_M}\right)^{\alpha} - \alpha\left(\frac{H_M}{L_M}\right)^{1-\alpha}\right]$. A change in the tax rate acts on both wages equi-proportionately through the negative direct taxation effect and the positive productivity effect, while, through the worker ratio effect, it affects w_h positively and w_l negatively. Thus, higher τ has a positive impact on $w_h - w_l$ through the productivity and worker ratio effects and a negative impact through the direct taxation effect. Because the former two effects dominate when τ is low, together with the fact the wage differential decreases with Fr, the shape of the dividing line is as stated in the lemma.

Figure 1 illustrates positions of the critical loci, $\tau = \tau_0^s$, $\tau = \tau_0^b$, $Fr = \overline{H}(\tau)$, and $Fr = H^*(\tau)$ on the (Fr, τ) plane, when $\tau_l > \tau_0^s$ and $\tau_{eo} < \tau_0^b$ are satisfied. Remember that $H = 0$ for $\tau < \tau_0^s$ and $\tau > \tau_0^b$; $L_T > (=)0$ for $Fr < (\geq)\overline{H}(\tau)$; and $H = Fr$ for $Fr < H^*(\tau)$, while $H = H^*(\tau)$ for $Fr \geq H^*(\tau)$.

The lemmas show how sectoral and skill distributions of workers are determined for each combination of Fr and τ. With the dynamics of Fr_t and τ_t, the lemmas allow one to explore how the structure of the economy changes over time. The next subsection examines the dynamics of transfers of each lineage and thereby derives the dynamics of Fr_t for given τ.

3.3 Dynamics of individual transfers and Fr_t

3.3.1 Dynamics of individual transfers

The dynamic equation linking the received transfer b_t^i to the transfer given to the next generation b_{t+1}^i is derived from the transfer rule (31). For a current unskilled worker, i.e. one who has received $b_t^i < e$, it is obtained by substituting $w_t^i = w_{l,t}$ and $a_t^i = b_t^i$ into (31):

$$b_{t+1}^i = b_l(b_t^i; w_{l,t}) \equiv \gamma_b\{w_{l,t} + (1+r)b_t^i\}. \tag{34}$$

The assumption $\gamma_b(1+r) < 1$ is made so that the fixed point of the equation for *given* $w_{l,t}$, $b_l^*(w_{l,t}) \equiv \frac{\gamma_b}{1-\gamma_b(1+r)} w_{l,t}$, exists. The fixed point becomes crucial in later analyses.

For a present skilled worker, i.e. one who has received $b_t^i \geq e$, the dynamic equation is

$$b_{t+1}^i = \ b_h(b_t^i; w_{h,t}) \equiv \gamma_b\{w_{h,t} + (1+r)(b_t^i - e)\}, \tag{35}$$

which is obtained by substituting $w_t^i = w_{h,t}$ and $a_t^i = b_t^i - e$ into (31). The fixed point for given $w_{h,t}$ is $b_h^*(w_{h,t}) \equiv \frac{\gamma_b}{1-\gamma_b(1+r)}[w_{h,t} - (1+r)e]$. In the equal opportunity case, i.e. $Fr_t \geq H^*(\tau)$, $w_{h,t} - (1+r)e = w_{l,t}$ holds, so the two equations coincide, and when the educational investment is not profitable, i.e. $\tau < \tau_0^s$ or $\tau > \tau_0^b$, all individual dynamics follow (34) with $w_{l,t} = A_T$.

The two dynamic equations show that, when $\tau \in [\tau_0^s, \tau_0^b]$, the dynamics of transfers within a lineage depend on the time evolution of wage levels and thus the proportion of skilled workers H_t (and τ). Since $H_t = Fr_t$ for $Fr_t < H^*(\tau)$ and $H_t = H^*(\tau)$ for $Fr_t \geq H^*(\tau)$, the individual dynamics are ultimately dependent on the dynamics of Fr_t.

3.3.2 Dynamics of Fr_t

The dynamics of Fr_t (the proportion of individuals who can afford education) are in turn determined by the dynamics of individual transfers. Thus, the individual and aggregate dynamics are interrelated.

In the unequal opportunity case, i.e. $Fr_t < H^*(\tau)$, if children of some of unskilled workers become accessible to education through wealth accumulation, Fr_t would increase over time, while, if a portion of skilled workers cannot leave transfers to cover the cost of education, Fr_t would decrease. The former occurs iff there exist lineages satisfying $b_t^i < e$ and $b_{t+1}^i \geq e$. From (34), the following condition must be satisfied for such lineages to exist:

$$b_l^*(w_{l,t}) \equiv \frac{\gamma_b}{1-\gamma_b(1+r)} w_{l,t} > e. \tag{36}$$

By contrast, the latter case occurs iff lineages satisfying $b_t^i \geq e$ and $b_{t+1}^i < e$ exist. From (35), the necessary condition is

$$b_h^*(w_{h,t}) \equiv \frac{\gamma_b}{1-\gamma_b(1+r)}\{w_{h,t} - (1+r)e\} < e. \tag{37}$$

Since $b_h^*(w_{h,t}) \geq b_l^*(w_{l,t})$ is satisfied, the above equations do not hold simultaneously. If (36) is satisfied,

$Fr_{t+1} \geq Fr_t$, while if (37) holds, $Fr_{t+1} \leq Fr_t$. Note that $Fr_{t+1} = Fr_t$ is possible depending on the distribution of transfers. However, if the condition continues to hold, Fr_t does change at some point. When neither equations are satisfied, $Fr_{t+1} = Fr_t$.

In the equal opportunity case ($Fr_t \geq H^*(\tau)$), the dynamics of Fr_t are determined by the relative value of $b_l^*(w_{l,t}) = b_h^*(w_{h,t})$ to e, and when the educational investment is unprofitable ($\tau < \tau_0^s$ or $\tau > \tau_0^b$), the dynamics depend on the relative value of $b_l^*(w_{l,t}) = \frac{\gamma_b}{1-\gamma_b(1+r)} A_T$ to e.

4 Analysis

This section examines how the structure of the economy changes over time from a given initial distribution of wealth (transfers). Since the dynamics of Fr_t differ greatly depending on the value of $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T$ ($b_l^*(w_{l,t})$ when $w_{l,t} = A_T$) relative to e, the section is divided into two subsections based on the relative value. In each subsection, initially the dynamics are examined for given τ, then τ is endogenized, and finally the case in which income redistribution too is a policy option is examined.

4.1 When $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T > e$

This is the case in which the productivity of sector T is high enough (relative to the education cost) that, even when the unskilled wage is lowest, descendants of unskilled workers gain access to education eventually. Hence, from any initial distribution of wealth, Fr_t increases over time.

4.1.1 Exogenous tax rate

The next proposition presents the dynamics of the economic structure when the tax rate is fixed.

Proposition 1 *Suppose* $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T > e$ *and* τ *is fixed.*

(*i*) (*Dynamics of* Fr_t) Fr_t *increases over time and* $Fr_t = 1$ *is satisfied in the long run.*

(*ii*) (*Dynamics of sectoral and skill distributions of workers*) *Given* $\tau \in [\tau_0^s, \tau_0^b]$, (*a*) *when* $L_{T,t} > 0$, $H_t = Fr_t$, $H_{M,t}$, $H_{G,t}$, *and* $L_{M,t}$ *all increase proportionally, while* $L_{T,t}$ *decreases over time;* (*b*) *when* $L_{T,t} = 0$ *and* $Fr_t < H^*(\tau)$, $H_t = Fr_t$, $H_{M,t}$, *and* $H_{G,t}$ *increase and* $L_{M,t} = L_t = 1 - Fr_t$ *decreases over time; and* (*c*) *when* $Fr_t \geq H^*(\tau)$, $H_t = H^*(\tau)$ *and the sectoral allocation of workers is time-invariant. When* $\tau < \tau_0^s$ *or* $\tau > \tau_0^b$, $H_t = 0$ *and* $L_{T,t} = 1$.

(*iii*) (*Wage dynamics*) *Given* $\tau \in (\tau_0^s, \tau_0^b)$, (*a*) *when* $L_{T,t} > 0$, *both wages are constant;* (*b*) *when* $L_{T,t} = 0$ *and* $Fr_t < H^*(\tau)$, $w_{l,t}$ *increases and* $w_{h,t}$ *decreases over time; and* (*c*) *when* $Fr_t \geq H^*(\tau)$, *the wages are constant and* $w_{h,t} - (1+r)e = w_{l,t}$. *When* $\tau \leq \tau_0^s$ *or* $\tau \geq \tau_0^b$, $w_{l,t} = A_T$ ($= w_{h,t} - (1+r)e$ *when* $\tau = \tau_0^s, \tau_0^b$).

The dynamics of Fr_t and other aspects of the economic structure can be grasped readily by employing a phase diagram. Figure 2 presents the dynamics when $\tau_l > \tau_0^s$ and $\tau_{eo} < \tau_0^b$,[21] which is obtained

[21]Remember that τ_l is defined as the tax rate maximizing w_l when $L_T = 0$ and τ_{eo} as the tax rate at the inflection point of $Fr = H^*(\tau)$. All the results in the proposition and the explanation that follows apply irrespective of the value of τ_l relative to τ_0^s and that of τ_{eo} relative to τ_0^b.

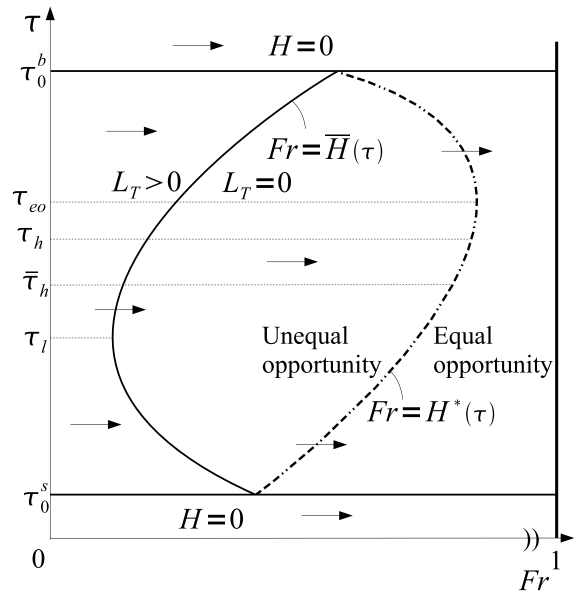

Figure 2: The dynamics when $\frac{A_T}{e}$ is high and τ is fixed

by superimposing the dynamics of Fr_t and positions of τ_h and $\overline{\tau}_h$ on Figure 1. In the figure, $H = 0$ for $\tau < \tau_0^s$ and $\tau > \tau_0^b$; $L_T > 0$ for $Fr < \overline{H}(\tau)$ and $L_T = 0$ for $Fr \geq \overline{H}(\tau)$; and the economy belongs to the unequal opportunity case and $H = Fr$ for $Fr < H^*(\tau)$, whereas it satisfies equal opportunity and $H = H^*(\tau)$ for $Fr \geq H^*(\tau)$. Directions of motion of Fr are represented with horizontal arrows.

The figure shows that, irrespective of the initial distribution of wealth (thus Fr_0) and τ, Fr_t increases over time and $Fr = 1$ holds in the long run. However, the tax rate makes significant differences in the dynamics and the long-run state of other aspects of the economy. When the tax rate is very low or very high, i.e. $\tau < \tau_0^s$ or $\tau > \tau_0^b$, the educational investment is not rewarding for any Fr, *everyone* becomes an unskilled worker, and thus only sector T is in operation, as shown in Lemma 2.[22] As explained in detail after the lemma, when the tax rate is too low, the productivity of sector M is low because task specialization among skilled workers is limited due to a lack of the governmental service that helps reducing the coordination cost. Further, as a result of the low productivity, many unskilled workers, who are complementary to skilled workers in sector M, choose sector T rather than sector M. Hence, the skilled wage is low and the educational investment is unprofitable. By contrast, when the tax rate is too high, what depresses the skilled wage and makes education unrewarding is the tax burden that more than offsets the resultant high degree of specialization and induces many unskilled workers to choose sector T for tax avoidance.

If the tax rate is not extreme, the economy goes through the following path of the structural change when it starts with $Fr_0 < \overline{H}(\tau)$. While $Fr_t < \overline{H}(\tau)$ is satisfied, unskilled workers are abundant and some of them have to take jobs in sector T. As more individuals gain access to education over time, the proportion of skilled workers increases and production and employment shift to sector M. Tax revenue increases with the growth of sector M and the government too expands. Note that unskilled

[22] From the equation of Assumption 1, an increase in A_T lowers τ_0^b and raises τ_0^s. That is, $H = 0$ is more likely to occur with 'bad' tax policies when the productivity of the traditional sector is higher.

workers too move to sector M (with the skilled-unskilled labor ratio of the sector unchanged), because higher numbers of skilled workers in the sector and in the government have, given other things equal, positive effects on the supply of the intermediate product and thus the sector's productivity. As long as $L_{T,t} > 0$, both wages remain unchanged because the unskilled wage is fixed by the productivity of sector T and the (total factor) productivity of sector M is unchanged.[23]

Once sector M becomes large enough to absorb all unskilled workers, the dynamics change. Because the wages are now independent of A_T (and $\frac{H_{G,t}}{H_t}$ is constant), the wages are determined by the ratio of total unskilled to skilled workers (see eqs. 25 and 26). Hence, as more individuals become skilled workers, the unskilled wage rises, while the skilled wage and wage inequality fall. The rising unskilled wage stimulates wealth accumulation of unskilled workers and the growth of Fr_t and H_t. The sectoral shift, the skill upgrading, and the associated changes in the wages end when the economy reaches the state of equal opportunity where both wages (net of the education cost) are equal.

When $\tau \in (\tau_0^s, \tau_0^b)$, wages change with the growth of H for $Fr \in (\overline{H}(\tau), H^*(\tau))$. How does net aggregate labor income $[w_h - (1+r)e]H + w_l(1-H)$, a measure of aggregate (private) consumption for given aggregate assets (see eq. 30), change with H?[24] The next proposition shows that it increases and then decreases with H. That is, *overeducation is inevitable in the long run.* Further, the proposition shows that, when attention is limited to steady states, there is a range of tax rates where *overeducation* occurs in the sense that H is higher but net aggregate labor income is lower than at the tax rate maximizing it. In the proposition, net aggregate labor income is denoted as $\widetilde{Y}_L$.

Proposition 2 *Suppose* $\tau \in (\tau_0^s, \tau_0^b)$.

(*i*) *For given* τ, *there exists unique* $H \in [\overline{H}(\tau), H^*(\tau))$ *maximizing net aggregate labor income, denoted* $H^\diamond(\tau)$, *and* $\frac{\partial \widetilde{Y}_L}{\partial H} < (>) 0$ *for* $H > (<) H^\diamond(\tau)$. *That is, overeducation is inevitable in the long run when* $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T > e$.

(*ii*) *In the equal opportunity case, net labor income (thus private consumption) is maximized at* $\tau = \overline{\tau}_h < min\{\tau_{eo}, \tau_0^b\}$ *and* $\frac{\partial \widetilde{Y}_L}{\partial \tau} > (<) 0$ *for* $\tau < (>) \overline{\tau}_h$, *while* $H = H^*(\tau)$ *is maximized at* $\tau = min\{\tau_{eo}, \tau_0^b\}$ *and* $\frac{\partial H^*(\tau)}{\partial \tau} > (<) 0$ *for* $\tau < (>) min\{\tau_{eo}, \tau_0^b\}$. *That is, when* $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T > e$, *among steady states with different* τ, *overeducation occurs in a range of* τ.

The first type of overeducation arises because only educated individuals benefit directly from employment opportunities at the government and thus the private return to education is higher than the social return.[25] The overeducation occurs only at a late stage of development since the social return is positive while not many can access education. The second type of overeducation is represented in

[23]Given the tax rate, the amount of the governmental service per production team does not change with H_t, hence the degree of task specialization and the total factor productivity remain constant (see equations 9, 14, 15, 18, and 23).

[24]Note that net aggregate labor income does *not* equal GDP net of the education cost: it equals $Y_T + Y_M - (1+r)eH$ from $w_h H_G = \tau Y_M$, while GDP equals net aggregate labor income plus the indirect tax revenue τY_M. However, since utility does not depend on the government consumption directly, it is a better measure of aggregate welfare.

[25]Since total labor income when $L_T = 0$ equals Y_M, from (18), the social return is $\frac{\partial Y_M}{\partial H} - (1+r)e - \frac{\partial Y_M}{\partial L_M}$, while $w_l = (1-\tau)\frac{\partial Y_M}{\partial L_M}$ from (19) and $w_h = \frac{1-\alpha(1-\tau)}{1-\alpha}\frac{\partial Y_M}{\partial H}$ from (20) and (23).

Figure 2: the tax rate maximizing $H^*(\tau)$, τ_{eo}, is higher than the net wage-maximizing rate, $\overline{\tau}_h$, thus there is a range of τ at which $H^*(\tau)$ is higher than at $\overline{\tau}_h$. This overeducation arises because *the return to taxation* is higher for skilled workers.[26]

The fact that only the educated benefit directly from governmental positions, together with the absence of taxation in sector T, also leads to *the oversized traditional sector* at an early stage of development: the unskilled wage of sector M is lower than the marginal productivity of unskilled workers, i.e. $w_l = (1-\tau)\frac{\partial Y_M}{\partial L_M}$ (footnote 25), thus too many unskilled workers choose sector T when $L_T > 0$, and $\overline{H}(\tau)$ is too high.

4.1.2 Endogenous tax rate

Now the determination of the tax rate is modeled. Suppose that the tax rate is chosen by a politically influential group so as to maximize their incomes in each period. Since the tax rate affects skilled and unskilled wages differently and only educated individuals can access skilled jobs, educated and uneducated individuals have different stakes in the tax policy.[27] Thus, situations in which either of the two groups have political power to determine the policy are examined.

As for the timing of the policy decision, two cases are considered. In the *commitment case*, the tax rate that is optimal to them before education is completed is implemented in the next period. By contrast, in the *non-commitment case*, the tax rate can be chosen after education is completed and thus an ex-ante optimal rate may not be implemented. This paper mainly focuses on the commitment case because one of the results of the non-commitment case (Proposition 5 (ii) in Section 4.1.3) is not robust under a more realistic assumption that parents care about effects of their policies on their descendants. The next lemma presents the equilibrium tax rate.

Lemma 5 (*i*) *When educated individuals determine the tax rate, it equals $\overline{\tau}_h$ for $Fr \leq \overline{H}(\overline{\tau}_h)$, τ satisfying $Fr = \overline{H}(\tau)$ for $Fr \in \left[\overline{H}(\overline{\tau}_h), \overline{H}(min\{\tau_h, \tau_0^b\})\right]$, and $min\{\tau_h, \tau_0^b\}$ for $Fr \geq \overline{H}(min\{\tau_h, \tau_0^b\})$. In the commitment case, there exists $Fr \in (H^*(\overline{\tau}_h), H^*(min\{\tau_h, \tau_0^b\}))$ above which it equals $\overline{\tau}_h$.*

(*ii*) *When the uneducated determine τ, any value can be an equilibrium rate for $Fr \leq \overline{H}(max\{\tau_l, \tau_0^s\})$. For $Fr > \overline{H}(max\{\tau_l, \tau_0^s\})$, the rate equals $max\{\tau_l, \tau_0^s\}$. In the commitment case, it equals τ satisfying $Fr = H^*(\tau)$ for $Fr \in [H^*(max\{\tau_l, \tau_0^s\}), H^*(\overline{\tau}_h)]$ and is $\overline{\tau}_h$ for higher Fr.*

Based on the lemma, Figure 3 shows the dynamics of Fr_t and τ_t when $\tau_l > \tau_0^s$ and $\tau_{eo} < \tau_0^b$, the same case as Figure 2.[28] The heavy solid (dashed) line represents the dynamics when educated (uneducated) individuals determine the tax policy, and the stars represent the long-run level of H for each case.

[26] w_h is maximized at $\tau = \tau_h$, while w_l is maximized at $\tau = \tau_l < \tau_h$, and $w_i (i = h, l)$ decreases (increases) with τ for $\tau > (<) \tau_i$ [Lemma 1]. Since $w_h - w_l$ decreases with H, τ_{eo} is the rate maximizing $w_h - w_l$ at $H = H^*(\tau_{eo})$ and thus $\tau_{eo} > \tau_h$ must hold. By contrast, $\overline{\tau}_h$ is the rate maximizing w_h (or w_l) under the constraint $w_h - (1+r)e = w_l$ and thus $\overline{\tau}_h \in (\tau_l, \tau_h)$ must be true.

[27] Although all educated individuals take skilled jobs in an equilibrium, distinguishing them from skilled workers is important. See the proof of Lemma 5 in Appendix for details.

[28] All the results that follow hold qualitatively regardless of the value of τ_l relative to τ_0^s and that of τ_{eo} relative to τ_0^b.

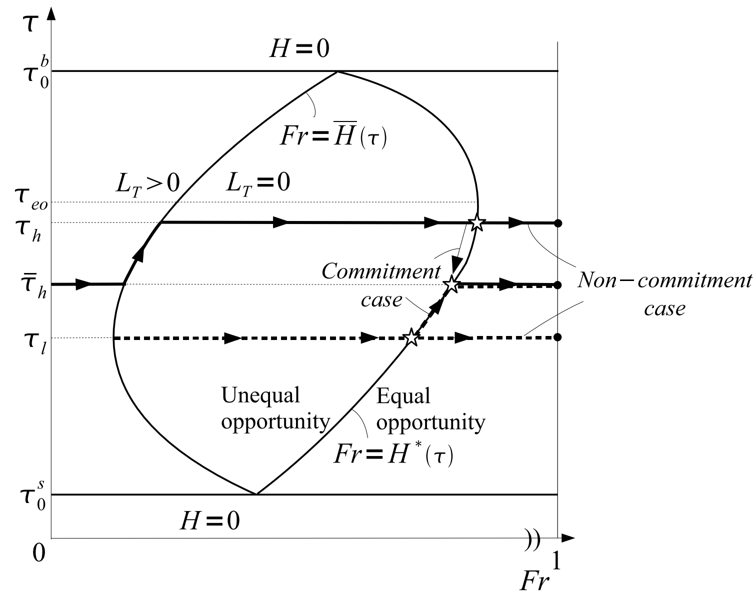

Figure 3: The dynamics when $\frac{A_T}{e}$ is high and τ is endogenous

The dashed line is not presented for $Fr \leq \overline{H}(\tau_l)$, since $w_l = A_T$ for any τ and thus the uneducated is indifferent on the tax rate. Whichever group decide on the tax rate, Fr_t increases and τ_t non-decreases over time (except when the educated choose the policy and commit to the ex-ante optimal rate, in which case τ_t declines from τ_h to $\overline{\tau}_h$ at some point), thus the qualitative dynamics of the structure of the economy are mostly as presented in Proposition 1.

The dynamics of the commitment case and of the non-commitment case are different when Fr is very high. In the commitment case, the effect of τ on H is taken into account in choosing τ, thus the chosen rate maximizes the wage of the group with decisive power. Hence, in the long run, both groups choose $\overline{\tau}_h$, the rate maximizing net wage (thus private consumption) in the equal opportunity case. (Remember, however, that $H^*(\overline{\tau}_h)$ is greater than H maximizing net aggregate labor income at $\tau = \overline{\tau}_h$.) By contrast, in the non-commitment case, the tax rate can be changed after education is completed and H is determined taking into account this, thus the wage-maximizing rate may not be realized, which is the case when Fr is high. In particular, when the educated determine the tax rate, those who can afford education cannot coordinate to select the optimal rate by restricting H to $H^*(\overline{\tau}_h) < Fr$, because, after education is completed, the educated have incentives to deviate from $\overline{\tau}_h$ and choose τ_h for higher wage (and anticipating this, too many individuals take education). In contrast, when the uneducated have power, the optimal rate cannot be an equilibrium, because after $H^*(\overline{\tau}_h)$ individuals take education, the uneducated have incentives to lower the tax rate for higher wage, which results in a negative return to education.

The tax rate is higher when it is chosen by the educated, except at the last stage of the commitment case, because the return to taxation is higher for them from employment opportunities at the government. The question is how far the chosen rates are from the best rate in the competitive economy and the socially optimal rate (in terms of efficiency).

The next proposition describes tax rates maximizing net aggregate labor income (thus aggregate private consumption) for given Fr in the competitive and the command economies, and shows that the rate chosen by the educated (the uneducated) is higher (lower) than the best-optimal rate when $L_T = 0$ and $w_l > A_T$ hold at the chosen rate (except at the last stage of the commitment case, which is best in the competitive economy), while when $L_T > 0$, the rate selected by *educated individuals* is best in the competitive economy but is *lower* than the optimal rate (see Figure 3).

Proposition 3 *(i) For given Fr, the tax rate maximizing net aggregate labor income* $\widetilde{Y}_L$ *in the competitive economy equals* $\overline{\tau}_h$ *for* $Fr \leq \overline{H}(\overline{\tau}_h)$ *and* τ *satisfying* $Fr = \overline{H}(\tau)$ *for* $Fr \in [\overline{H}(\overline{\tau}_h), \overline{H}(\tau_y)]$, *where* $\tau_y \equiv \frac{(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta}}{1-\alpha\frac{\rho\theta}{\delta-\rho-\theta}} \in (\overline{\tau}_h, min\{\tau_h, \tau_0^b\})$. *For* $Fr \geq \overline{H}(\tau_y)$, *the rate is* τ_y *until some* $Fr \in (H^*(\overline{\tau}_h), H^*(\tau_y))$, *above which it equals* $\overline{\tau}_h$. *The net income increases (decreases) with* τ *when* τ *is lower (higher) than the rate maximizing it.*

(ii) For given Fr, the tax rate maximizing $\widetilde{Y}_L$ *in the command economy equals* τ_y.

(iii) When $Fr \leq \overline{H}(\tau_y)$, τ *chosen by the educated is best in the competitive economy but is lower than the optimal rate. When* $Fr > \overline{H}(\tau_y)$, *the rate chosen by the educated (the uneducated) is higher (lower) than the best and optimal rate, except in the commitment case, in which it is best in the competitive economy (but is lower than the optimal rate) if Fr is greater than some* $Fr \in (H^*(\overline{\tau}_h), H^*(\tau_h))$ *(if* $Fr \geq H^*(\overline{\tau}_h)$*).*

When $L_T = 0$ and $w_l > A_T$ at the equilibrium tax rate, the educated overevaluate the contribution of the governmental activity (thus taxation) on output and the uneducated underevaluate it. When $L_T > 0$, by contrast, the rate chosen by the educated is best in the competitive economy, because the unskilled wage is fixed by the productivity of sector T, i.e. $w_l = A_T$, and thus maximizing the skilled wage amounts to maximizing net aggregte labor income. However, the best rate in such case is lower than the optimal rate τ_y, at which the 'efficiency' of sector M, $\Omega(\frac{H_G}{H})$, is maximized (see eq. 18). The reason is that, in the competitive economy, unskilled workers in sector M are overtaxed and thus, to keep them from escaping to sector T, the tax rate must be lower.

Results suggest that the socially desirable distribution of political power changes with development (see the figure). While the unskilled wage is not affected by the tax policy, i.e. $Fr \leq \overline{H}(\tau_l)$, educated individuals should have decisive power, unless a very high priority is given to equity: when they choose τ, net aggregate labor income is maximized in the competitive economy (although the wage inequality too is maximized), whereas when the uneducated determine the policy and, plausibly, they are concerned about the inequality too in choosing τ, there exists the non-negligible risk that $\tau \geq \tau_0^b$ or $\tau \leq \tau_0^s$ is chosen and $\widetilde{Y}_L$ (and the inequality) is minimized. Otherwise, i.e. $Fr > \overline{H}(\tau_l)$, it is difficult to state the desirable distribution precisely, but what is clear is that the uneducated should have some influence on the policy. The reason is that, when $L_T = 0$, as their influence becomes stronger, the unskilled wage increases, which leads to greater wealth accumulation of unskilled workers and thus faster structural change. (Further, the tax rate maximizing $\widetilde{Y}_L$ is between the rates preferred by the two groups when $Fr > \overline{H}(\tau_y)$.)

4.1.3 When the redistributive policy is available

The preceding analysis shows that both groups cannot choose the tax rate and thus the size of public service optimally, nor can they avoid overeducation at a late stage and oversized sector T at an early stage of development. If redistribution is feasible, by contrast, these inefficiencies can be resolved.

Assume that the government can impose a proportional tax on wages generated in the modern and the government sectors and use the revenue to provide a lump-sum transfer to workers with political power in the sectors. The transfer to the educated may be interpreted as benefits and privileges associated with pollical power. (When only non-targeted transfer is available, results when the educated [the uneducated] have power are same as the original economy without [with] redistribution.) Denote the skilled wage (20) and the unskilled wage (19) (with $\frac{H_G}{H}$ replaced by equation 23) by $w_h(\tau,H,L_M)$ and $w_l(\tau,H,L_M)$ respectively.

When educated individuals have decisive political power and can commit to policies that are optimal for them before H is fixed, the chosen policies are solutions to the following problem.

$$max_{\{x,T,\tau\}}\ (1-x)w_h(\tau,H,L_M)+T \tag{38}$$

$$s.t.\quad (1-x)w_h(\tau,H,L_M)+T-(1+r)e \geq (1-x)w_l(\tau,H,L_M), \tag{39}$$

$$(1-x)w_l(\tau,H,L_M) \geq A_T, \tag{40}$$

$$x[w_h(\tau,H,L_M)H+w_l(\tau,H,L_M)L_M]=TH, \tag{41}$$

$$H=Fr \text{ when (39) holds with such } H\text{; otherwise, (39) holds with } =, \tag{42}$$

$$L_M=1-H \text{ when (40) holds with such } L_M\text{; otherwise, (40) holds with } =, \tag{43}$$

where $x\in[0,1]$ is the wage tax rate and T is the lump-sum transfer. (39) is the incentive compatibility constraint for the educated (the net return to education must be non-negative), (40) is the individual rationality constraint for the unskilled, (41) is the governmental budget constraint for the redistributive policy, and (42) and (43) determine H and L_M, respectively. Note that feasibility conditions $H \leq Fr$ and $L_M \leq 1-H$ are reflected in (42) and (43). When they cannot commit to ex-ante optimal policies, x, T, and τ are determined given H.

The next proposition summarizes the chosen policies, H, and L_M. In the proposition, $\underline{H}^b(\tau)$ ($\tau\in(\tau_0^s,\tau_0^b)$) denotes the bigger H of two $H\in(0,1)$ at which net aggregate labor income $\widetilde{Y}_L$ when $L_T=0$ equals A_T, where $\underline{H}^b(\tau)>H^*(\tau)$.

Proposition 4 *Suppose that educated individuals determine τ and the redistributive policy.*

(*i*) *In the commitment case, τ coincides with the rate described in Proposition 3 (i).*

(*a*) *If $Fr \leq \overline{H}(\tau_y)$, $x=T=0$, $H=Fr$, and L_M is determined so that $w_l(\tau,Fr,L_M)=A_T$.*

(*b*) *If $Fr>\overline{H}(\tau_y)$, $L_M=1-H(L_T=0)$. There exists $Fr\in(H^*(\overline{\tau}_h),H^*(\tau_y))$ such that $x,T=(>)0$ for higher (lower) Fr, and when $x,T>0$, unskilled workers receive A_T, $\tau=\tau_y$, and $H=Fr$, while when $x=T=0$, $\tau=\overline{\tau}_h$, and $H=H^*(\overline{\tau}_h)$.*

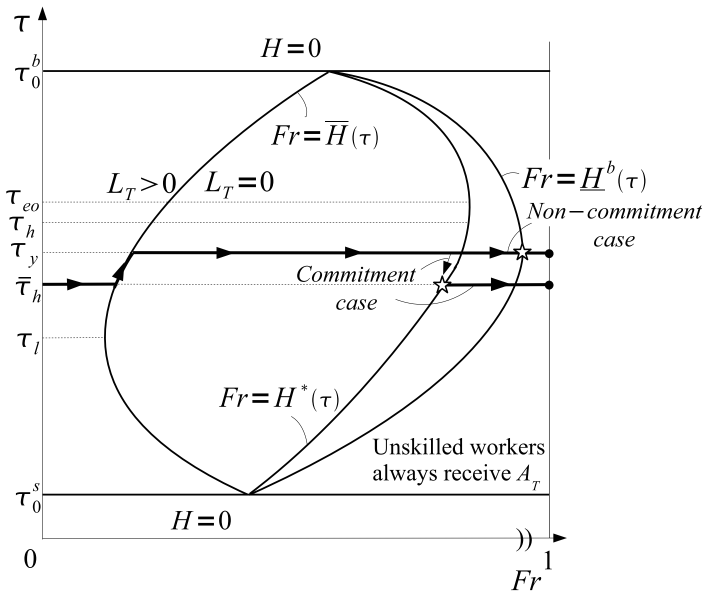

Figure 4: The dynamics when $\frac{A_T}{e}$ is high and the educated determine τ and the redistributive policy

(*ii*) *In the non-commitment case, the result is same as* (*i*) *if* $Fr \leq \overline{H}(\tau_y)$*. If* $Fr > \overline{H}(\tau_y)$*,* $x, T > 0$*, unskilled workers receive* A_T*,* $\tau = \tau_y$*,* $L_M = 1 - H$*, and* $H = Fr$ *for* $Fr \leq \underline{H}^b(\tau_y)(> H^*(\tau_y))$ *and* $H = \underline{H}^b(\tau_y)$ *otherwise.*

Figure 4 illustrates the dynamics of Fr_t and τ_t when $\tau_l > \tau_0^s$ and $\tau_{eo} < \tau_0^b$ (the stars represent the long-run level of H for each case). As before, τ and H determine $\widetilde{Y}_L$ completely.

In the commitment case, τ coincides with the rate of Proposition 3 (i) and thus is *best in the competitive economy without redistribution for any Fr*. When $w_l = A_T$ holds without redistribution ($Fr \leq \overline{H}(\tau)$), the redistributive policy is not implemented (thus the outcome is same as the economy without redistribution), since it induces unskilled workers to escape for sector T and lowers the skilled wage. When $w_l > A_T$ without redistribution, the policy is implemented unless Fr is very high, and the wage-tax rate is chosen so that the unskilled become indifferent between the sectors (their after-tax wage equals A_T). Then, the amount each skilled worker receives is $(\widetilde{Y}_L - A_T L)/H + (1+r)e$, thus maximizing their disposable income amounts to maximizing $\widetilde{Y}_L$ and the *socially optimal* rate τ_y is selected. When Fr is very large, however, the policy lowers the disposable income and thus is *not* implemented. Since $(\widetilde{Y}_L - A_T L)/H$ decreases with H and the net skilled wage when $H = H^*(\tau)$ is highest at $\tau = \overline{\tau}_h$ (Proposition 2), $x = T = 0$ and $\tau = \overline{\tau}_h$ are chosen and equal opportunity is realized from some $Fr \in (H^*(\overline{\tau}_h), H^*(\tau_y))$.[29]

By contrast, because the switch to $x = T = 0$ is not possible in the non-commitment case, the redistributive policy and $\tau = \tau_y$ continue to be enforced and H_t increases over time until equal opportunity is realized at $H = \underline{H}^b(\tau_y)(> H^*(\tau_y))$, where both groups end up receiving (net of the education cost) A_T. Thus, overeducation is exacerbated when redistribution is possible.

[29]Thus, inequality in non-capital income between skilled and unskilled workers *decreases* over time as before, although the total pie for skilled workers, $\widetilde{Y}_L - A_T L + (1+r)eH$, increases with H (see the proof of Proposition 4 (i)(b)). By contrast, the after-tax unskilled wage remains A_T until the last stage, when it jumps to $\widetilde{Y}$.

The allocation realized in the commitment case is efficient while sector T does not exist and overeducation is not the issue (not many can access education), but is *not* efficient when Fr is low, where τ is too low (the public service is undersupplied for given H) and sector T is oversized, and when Fr is very high, where τ is too low and H is too high (Propositions 2 (i) and 3 (i)(ii)). By contrast, as shown below, *efficient allocation is attained unless Fr is fairly low* when the *uneducated* have decisive power and commit to ex-ante optimal policies.

When they determine policies, the chosen policies in the commitment case are solutions to:

$$max_{\{x,T,\tau\}} \{(1-x)w_l(\tau,H,L_M)+T\} \tag{44}$$

$$s.t. \quad (1-x)w_h(\tau,H,L_M)-(1+r)e \geq (1-x)w_l(\tau,H,L_M)+T, \tag{45}$$

$$(1-x)w_l(\tau,H,L_M)+T \geq A_T, \tag{46}$$

$$x[w_h(\tau,H,L_M)H+w_l(\tau,H,L_M)L_M] = TL_M, \tag{47}$$

$$H = Fr \text{ when (45) holds with such } H\text{; otherwise, (45) holds with } =, \tag{48}$$

$$L_M = 1-H \text{ when (46) holds with such } L_M\text{; otherwise, (46) holds with } =, \tag{49}$$

where (45) is the incentive compatibility constraint for the educated and (46) is the individual rationality constraint for the unskilled.

The following assumption is imposed to ensure $\overline{H}(\overline{\tau}_h) < H^{\diamond}(\overline{\tau}_h)$, that is, highest $\widetilde{Y}_L$ when $w_l = A_T$, which is at $\tau = \overline{\tau}_h$, is lower than highest $\widetilde{Y}_L$ when $w_l > A_T$ (thus $L_T = 0$) and $\tau = \overline{\tau}_h$.

Assumption 2 $(1-\alpha)^{1-\frac{\rho\theta}{\delta-\rho-\theta}} A_M\left(\frac{\alpha A_M}{A_T}\right)^{\frac{\alpha}{1-\alpha}} \Omega_0(1-\overline{\tau}_h)^{\frac{1}{1-\alpha}}\left[\frac{\overline{\tau}_h}{1-\overline{\tau}_h}\right]^{\frac{\rho\theta}{\delta-\rho-\theta}} > \frac{1-\alpha(1-\overline{\tau}_h)}{1-\alpha}\left[\frac{A_T}{1-\overline{\tau}_h}+(1+r)e\right].$

Note that this assumption implies Assumption 1 and, together with Proposition 3 (i), implies $\overline{H}(\tau_y) < H^{\diamond}(\tau_y)$. It is equivalent to the net *social* return to education when $w_l = A_T$ and $\tau = \overline{\tau}_h$ being positive in the competitive economy.[30] The assumption is very weak since, it is very likely in real economy that, as long as the best tax rate is selected, education is socially productive when sector T exists.

The next proposition summarizes the chosen policies, H, and L_M. $\underline{H}^s(\tau)$ ($\tau \in (\tau_0^s, \tau_0^b)$) denotes the smaller H of two $H \in (0,1)$ at which $\widetilde{Y}_L$ when $L_T = 0$ equals A_T, where $\underline{H}^s(\tau) < \overline{H}(\tau)$.

Proposition 5 *Suppose that uneducated individuals choose τ and the redistributive policy.*

- *(i) In the commitment case,*
  - *(a) If $Fr \leq \underline{H}^s(\tau_y)(< \overline{H}(\tau_y))$, even with the redistributive policy, they cannot receive more than A_T, thus they are indifferent among any policies that assure A_T, including those with $x = T = 0$ and $\tau < \tau_0^s$ or $\tau > \tau_0^b$.*
  - *(b) If $Fr > \underline{H}^s(\tau_y)$, the redistributive policy is implemented and the net return to education becomes zero, $\tau = \tau_y$, and $L_M = 1-H$. When $Fr \in (\underline{H}^s(\tau_y), H^{\diamond}(\tau_y)]$, $H = Fr$, and when $Fr > H^{\diamond}(\tau_y)(\in (\overline{H}(\tau_y), H^*(\tau_y)))$, $H = H^{\diamond}(\tau_y)$.*
- *(ii) In the non-commitment case, $H = L_M = 0$ and $w_l = A_T$.*

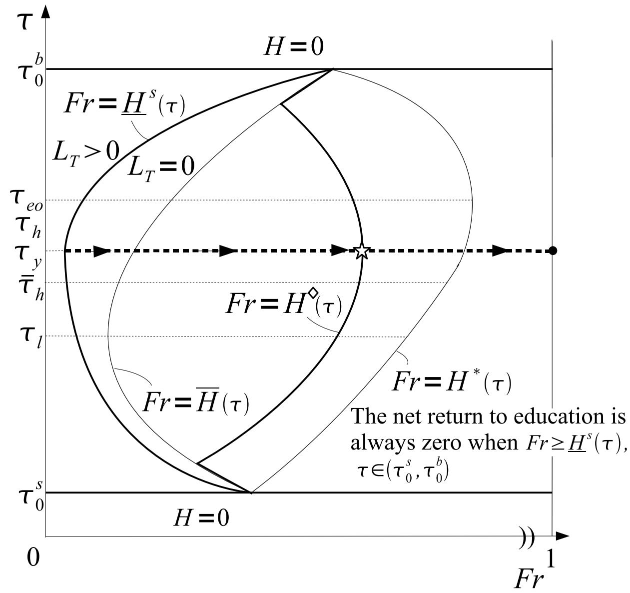

Figure 5: The dynamics when $\frac{A_T}{e}$ is high and the uneducated determine τ and the redistributive policy (the commitment case)

Figure 5 illustrates the dynamics of Fr_t and τ_t in the commitment case when $\tau_l > \tau_0^s$ and $\tau_{eo} < \tau_0^b$.[31] If the tax base of the redistributive policy, total labor income in the non-traditional sectors, does not exceed $A_T + (1+r)eH$ even when all unskilled workers are allocated to sector M, i.e. $Fr \leq \underline{H}^s(\tau)$, unskilled workers cannot obtain more than A_T with redistribution (since the net return to education must be non-negative), thus they are indifferent among any policies that assure A_T (thus, any dashed line is not presented in the figure), including ones with $x = T = 0$ and $\tau < \tau_0^s$ or $\tau > \tau_0^b$. By contrast, if the tax base exceeds $A_T + (1+r)eH$ with $L_M = 1 - H$, i.e. $Fr > \underline{H}^s(\tau)$, they implement the redistributive policy and extract from skilled workers until the net return to education becomes zero. Then, all workers receive the same level of disposable non-capital income (net of the education cost), thus maximizing unskilled workers' disposable income is equivalent to maximizing $\widetilde{Y}_L$ and the optimal rate τ_y is selected. Further, *unless Fr is fairly low, the allocation is efficient. Overeducation is avoided*, i.e. $H \leq H^\diamond(\tau_y)$, when *Fr* is very high, because unskilled workers choose x and T so that the net return to education becomes zero at optimal H. Intuitively, redistribution from the educated to the uneducated corrects the excessive private return. When *Fr* is low, *the overexpansion of sector T is averted*, because redistribution corrects the insufficient return to choosing sector M. However, when *Fr* is lower than $\overline{H}^{opt}(\tau_y) \in (\underline{H}^s(\tau_y), \overline{H}(\tau_y))$, where $\overline{H}^{opt}(\tau)$ is *Fr* satisfying $\frac{\partial Y_M}{\partial L_M} = A_T$ when $H = Fr$ and $L_T = 0$, *the redistributive policy overexpands sector M.*

In the non-commitment case, the effect on H is not taken into account when the policy is determined, thus the uneducated extract from the educated until the educated are indifferent between skilled and unskilled jobs *ex post*, i.e. $(1-x)w_h = (1-x)w_l + T$. Then, the net return to education

[30]Since the LHS of the equation is w_h from (24), the assumption can be expressed as $\frac{1-\alpha}{1-\alpha(1-\overline{\tau}_h)}w_h - (1+r)e - \frac{A_T}{1-\overline{\tau}_h} > 0$, where $w_h = \frac{1-\alpha(1-\overline{\tau}_h)}{1-\alpha}\frac{\partial Y_M}{\partial H}$ from (20) and (23) and $w_l = (1-\overline{\tau}_h)\frac{\partial Y_M}{\partial L_M} = A_T$ from (19) and $L_T > 0$.

[31]On the (Fr, τ) plane, $Fr = \underline{H}^s(\tau)$ is negatively (positively) sloped for $\tau < (>)\tau_y$, and $Fr = H^\diamond(\tau)$ when $H^\diamond(\tau) > \overline{H}(\tau)$ is positively (negatively) sloped for $\tau < (>)\tau_y$.

becomes negative, thus nobody takes education and only sector T is active.

As for the desirable distribution of political power, the uneducated should have power from an earlier stage of development than the economy without redistribution, where the educated should determine τ when the unskilled wage is not affected by τ ($Fr \leq \overline{H}(\tau_l)$) and the uneducated should have some influence on the policy for greater Fr. Now the uneducated should have *decisive* power for Fr higher than some $Fr \in (\underline{H}^s(\tau_y), \overline{H}^{opt}(\tau_y))$, that is, even when sector T is active without redistribution.[32] In particular, they *must* determine policies when $Fr > H^{\diamond}(\tau_y)$ (only then can overeducation be prevented) and when $Fr < \overline{H}(\tau_y)$ (only then is the overexpansion of sector T averted). When $Fr \in [\overline{H}(\tau_y), H^{\diamond}(\tau_y)]$, both groups choose optimal τ, but equality is realized and the speed of structural change is faster with the dominance of the uneducated. By contrast, the educated should control policies for $Fr \leq \underline{H}^s(\tau_y)(< \overline{H}^{opt}(\tau_y))$ due to the risk of destructive τ mentioned earlier.

4.2 When $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T < e$

So far, the initial distribution of wealth does not affect the long-run outcome, since Fr_t always increases over time when $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T > e$. Now effects of the initial distribution are examined by considering the case $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T < e$. Remember that unskilled (skilled) workers gain (lose) access to education over time if $b_l^*(w_{l,t}) = \frac{\gamma_b}{1-\gamma_b(1+r)} w_{l,t} > e$ (if $b_h^*(w_{h,t}) = \frac{\gamma_b}{1-\gamma_b(1+r)}[w_{h,t} - (1+r)e] < e$), as detailed in Section 3.3.2. Hence, when $\tau < \tau_0^s$ or $\tau > \tau_0^b$ and thus $H_t = 0$, $b_l^*(w_{l,t}) = \frac{\gamma_b}{1-\gamma_b(1+r)} A_T < e$ and Fr_t decreases. For $\tau \in [\tau_0^s, \tau_0^b]$, the direction of motion of Fr_t depends on Fr_t and the tax rate.

First, consider the region where $L_T > 0$ is satisfied, i.e. $Fr < \overline{H}(\tau)$ on the (Fr, τ) plane. By substituting (24) into the LHS of (37) and rearranging, $b_h^*(w_h) = e$ can be expressed as

$$(1-\alpha) A_M \left(\frac{\alpha A_M}{A_T}\right)^{\frac{\alpha}{1-\alpha}} \Omega_0 (1-\tau)^{\frac{1}{1-\alpha}} \left[\frac{\tau}{(1-\alpha)(1-\tau)}\right]^{\frac{\rho\theta}{\delta-\rho-\theta}} = \frac{e}{\gamma_b}. \tag{50}$$

The next lemma shows that, if the tax rate is very high or very small, $b_h^*(w_{h,t}) < e$ and thus Fr_t decreases over time, and it remains unchanged otherwise (see Figure 6 below).

Lemma 6 *Assume* $(1-\alpha)^{1-\frac{\rho\theta}{\delta-\rho-\theta}} A_M \left(\frac{\alpha A_M}{A_T}\right)^{\frac{\alpha}{1-\alpha}} \Omega_0 (1-\overline{\tau}_h)^{\frac{1}{1-\alpha}} \left[\frac{\overline{\tau}_h}{1-\overline{\tau}_h}\right]^{\frac{\rho\theta}{\delta-\rho-\theta}} > \frac{e}{\gamma_b}$. *Then, there exist* $\overline{\tau}_{h\downarrow}^s \in (\tau_0^s, \overline{\tau}_h)$ *and* $\overline{\tau}_{h\downarrow}^b \in (\overline{\tau}_h, \tau_0^b)$ *satisfying (50) such that, when* $L_{T,t} > 0$, Fr_t *decreases over time for* $\tau < \overline{\tau}_{h\downarrow}^s$ *and* $\tau > \overline{\tau}_{h\downarrow}^b$ *and is constant for* $\tau \in [\overline{\tau}_{h\downarrow}^s, \overline{\tau}_{h\downarrow}^b]$.

The result can be explained intuitively in a similar way to Lemma 2. Note that the assumption in the lemma implies Assumption 1 (since $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T < e$).

When $L_T = 0$, $b_h^*(w_h)$ and $b_l^*(w_l)$ depend on H ($= Fr$ in the unequal opportunity case) as well as τ. By plugging (25) with $H = Fr$ into the LHS of (37) and rearranging, $b_h^*(w_h) = e$ is expressed as

$$A_M {\Omega_0}^{1-\alpha} [1-\alpha(1-\tau)] \left(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}} [(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right)^{1-\alpha} \left(\frac{1-Fr}{Fr}\right)^{\alpha} = \frac{e}{\gamma_b}. \tag{51}$$

[32] The critical Fr is lower than $\overline{H}^{opt}(\tau_y)$ because while sector M is oversized for $Fr < \overline{H}^{opt}(\tau_y)$ when the uneducated determine policies, sector T is oversized for $Fr < \overline{H}(\tau_y)$ when the educated determine policies.

Given τ, when Fr is smaller (greater) than the value satisfying the equation, $b_h^*(w_h) > (<) e$ holds. Since the LHS of the equation equals w_h when $L_T = 0$, from Lemma 1, its locus is positively (negatively) sloped for $\tau < (>) \tau_h$ on the (Fr, τ) plane.

By plugging (26) with $H = Fr$ into the LHS of (36) and rearranging, $b_l^*(w_l) = e$ is expressed as

$$A_M \Omega_0^{1-\alpha} \alpha (1-\tau) \left(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}} [(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)} \right)^{1-\alpha} \left(\frac{Fr}{1-Fr} \right)^{1-\alpha} = \frac{[1-\gamma_b(1+r)]e}{\gamma_b}. \tag{52}$$

When Fr is greater (smaller) than the value satisfying the equation, $b_l^*(w_l) > (<)\, e$ holds. Since the LHS is w_l when $L_T = 0$, its locus is negatively (positively) sloped for $\tau < (>) \tau_l$ on the (Fr, τ) plane.

The critical loci $b_h^*(w_h) = e$ and $b_l^*(w_l) = e$ when $L_T = 0$ are *effective* at given Fr and τ only if $Fr \geq \overline{H}(\tau)$ and $Fr \leq H^*(\tau)$, i.e. $L_T = 0$ and $H = Fr$. In order for $b_l^*(w_l) = e$ to be effective for some Fr and τ, Assumption 1 needs to be replaced by the following assumption, which states that $b_l^*(w_l) = b_h^*(w_h) > e$ holds if the unskilled wage is highest (when $Fr \geq H^*(\tau)$ and $\tau = \overline{\tau}_h$).

Assumption 3 $(1-\alpha)^{(1-\alpha)(1-\frac{\rho\theta}{\delta-\rho-\theta})} A_M \Omega_0^{1-\alpha} \left(\frac{\alpha}{1-\gamma_b(1+r)} \right)^{\alpha} (1-\overline{\tau}_h) \left[\frac{\overline{\tau}_h}{1-\overline{\tau}_h} \right]^{\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}} > \frac{e}{\gamma_b}$

The assumption implies the one in Lemma 6, which states that $b_h^*(w_h) > e$ when $L_T > 0$ and $\tau = \overline{\tau}_h$, and thus Assumption 1.

The next lemma shows that the two loci intersect at two tax rates that are in between the two critical rates of Lemma 6, $\overline{\tau}_{h\downarrow}^s$ and $\overline{\tau}_{h\downarrow}^b$.

Lemma 7 *$b_h^*(w_h) = e$ and $b_l^*(w_l) = e$ intersect at $\tau = \tau^{*s} \in (\overline{\tau}_{h\downarrow}^s, \overline{\tau}_h)$ and $\tau = \tau^{*b} \in (\overline{\tau}_h, \overline{\tau}_{h\downarrow}^b)$.*

Clearly, the two loci intersect on $Fr = H^*(\tau)$. Based on Lemmas 6 and 7, the next lemma describes the shape of the *effective portion* of $b_h^*(w_h) = e$ when $L_T = 0$ and its intersection with $b_h^*(w_h) = e$ when $L_T > 0$ on the (Fr, τ) plane.

Lemma 8 *When $L_T = 0$, $b_h^*(w_h) = e$ is effective for $\tau \in [\overline{\tau}_{h\downarrow}^s, \tau^{*s}]$ and $\tau \in [\tau^{*b}, \overline{\tau}_{h\downarrow}^b]$ on the (Fr, τ) plane. For $\tau \in [\overline{\tau}_{h\downarrow}^s, \tau^{*s}]$, it is positively sloped. For $\tau \in [\tau^{*b}, \overline{\tau}_{h\downarrow}^b]$, its shape depends on the value of τ_h relative to τ^{*b} and $\overline{\tau}_{h\downarrow}^b$: (i) if $\tau_h \leq \tau^{*b}$, it is negatively sloped, (ii) otherwise, it is positively sloped for $\tau < min\{\tau_h, \overline{\tau}_{h\downarrow}^b\}$, and, when $\tau_h < \overline{\tau}_{h\downarrow}^b$, negatively sloped for $\tau > \tau_h$. The curve intersects with $b_h^*(w_h) = e$ when $L_T > 0$ on $Fr = \overline{H}(\tau)$.*

The curve is not effective for $\tau \in (\tau^{*s}, \tau^{*b})$ because $b_h^*(w_h) > e$ is satisfied for any Fr. Similarly, the next lemma describes the shape of the effective portion of $b_l^*(w_l) = e$.

Lemma 9 *When $L_T = 0$, $b_l^*(w_l) = e$ is effective for $\tau \in [\tau^{*s}, \tau^{*b}]$ on the (Fr, τ) plane. It is positively sloped for $\tau > max\{\tau_l, \tau^{*s}\}$, and, when $\tau_l > \tau^{*s}$, negatively sloped for $\tau < \tau_l$.*

Proof. Can be proved similarly to Lemma 8. □

The curve is not effective for $\tau < \tau^{*s}$ and $\tau > \tau^{*b}$ because $b_l^*(w_l) < e$ always holds.

As shown in the two lemmas, shapes of effective portions of $b_h^*(w_h) = e$ and $b_l^*(w_l) = e$ and relative positions of the two loci differ depending on the relative value of τ_h to τ^{*b} and $\overline{\tau}_{h\downarrow}^b$ and that of τ_l to

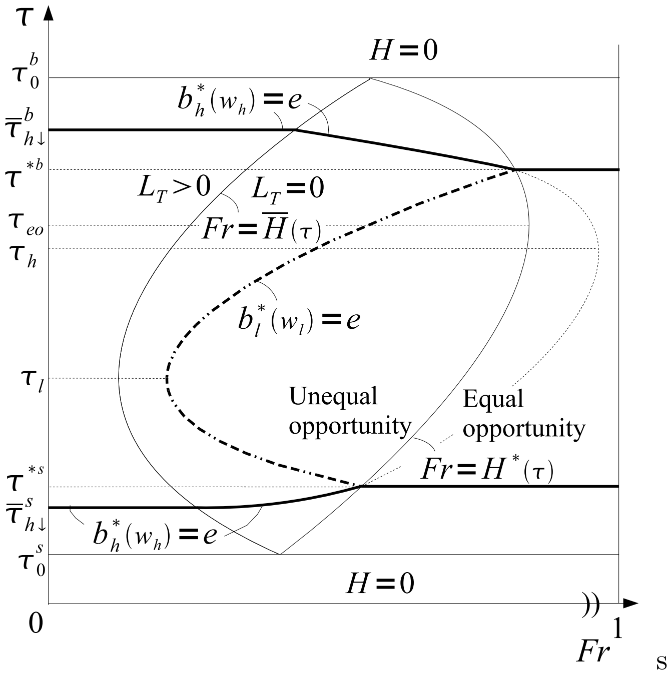

Figure 6: Locations of $b_h^*(w_h) = e$ and $b_l^*(w_l) = e$ when $\frac{A_T}{e}$ is low

τ^{*s}. Figure 6 superimposes positions of the two loci on Figure 1 when $\tau_h < \tau^{*b}(< \overline{\tau}_{h\downarrow}^b)$ and $\tau_l > \tau^{*s}$. A portion of $b_h^*(w_h) = e$ that is not effective (for $\tau \in (\tau^{*s},\tau^{*b})$) is represented with a dotted line.

4.2.1 Exogenous tax rate

Based on the above lemmas and discussions, the dynamics of Fr_t when the tax rate is fixed is summarized in the following proposition.

Proposition 6 *When $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T < e$, the dynamics of Fr_t for given τ are as follows.*

(i) For $\tau < \overline{\tau}_{h\downarrow}^s$ and $\tau > \overline{\tau}_{h\downarrow}^b$, $Fr_t > 0$ decreases over time.

(ii) For $\tau \in [\overline{\tau}_{h\downarrow}^s, \overline{\tau}_{h\downarrow}^b]$ and $Fr_t < \overline{H}(\tau)$, i.e. $L_{T,t} > 0$, Fr_t is constant.

(iii) For $\tau \in [\overline{\tau}_{h\downarrow}^s, \overline{\tau}_{h\downarrow}^b]$ and $Fr_t \geq \overline{H}(\tau)$, i.e. $L_{T,t} = 0$,

(a) For $\tau \in [\overline{\tau}_{h\downarrow}^s, \tau^{*s})$ and $\tau \in (\tau^{*b}, \overline{\tau}_{h\downarrow}^b]$, Fr_t decreases over time when $b_h^*(w_{h,t}) < e$, i.e. when the economy is located at the right side of $b_h^*(w_h) = e$ on the (Fr, τ) plane.

(b) Fr_t is time-invariant when $b_h^*(w_{h,t}) \geq e$ and $b_l^*(w_{l,t}) \leq e$, that is, when the economy is located at the left side (including the boundary) of $b_h^*(w_h) = e$ or $b_l^*(w_l) = e$.

(c) For $\tau \in (\tau^{*s}, \tau^{*b})$, Fr_t increases over time when $b_l^*(w_{l,t}) > e$, that is, when the economy is located at the right side of $b_l^*(w_l) = e$.

Figure 7 illustrates the dynamics of Fr_t and other aspects of the economic structure when $\tau_h < \tau^{*b}$ and $\tau_l > \tau^{*s}$ (the same case as Figure 6). The dynamics of sectoral and skill distributions of workers and the wages are as presented in Proposition 1 if Fr_t increases (and opposite if Fr_t decreases).

The dynamics differ greatly depending on the initial distribution of wealth (transfers) and the tax rate. When the tax rate is very low or very high, i.e. $\tau < \overline{\tau}_{h\downarrow}^s$ or $\tau > \overline{\tau}_{h\downarrow}^b$, from *any* initial distribution

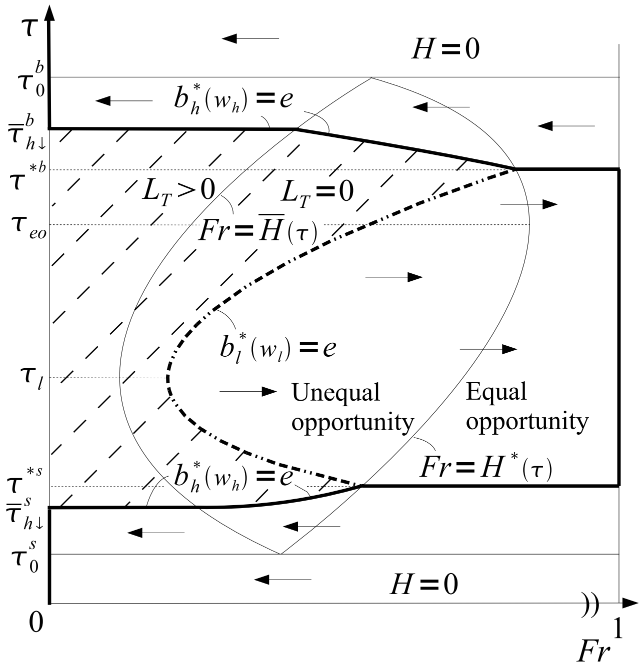

Figure 7: The dynamics when $\frac{A_T}{e}$ is low and τ is fixed

of wealth, Fr_t decreases over time and $Fr = 0$ holds in the long run.[33] In particular, the case in which $\tau \in [\tau_0^s, \overline{\tau}_{h\downarrow}^s)$ or $\tau \in (\overline{\tau}_{h\downarrow}^b, \tau_0^b]$ and $Fr_0 \geq H^*(\tau)$ are satisfied is worth mentioning. Initially all workers are in sector M and equal opportunity is attained, but the wage is low (due to the sector's low productivity or the heavy tax burden) and thus such desirable state is not sustained. Less wealthy individuals lose access to education gradually. After equal opportunity is lost and the proportion of unskilled workers starts to increase, the unskilled wage falls and the skilled wage and wage inequality rise over time. Eventually, unskilled workers become so abundant that some of them must seek jobs in sector T. After that, the wages stabilize, but Fr_t continues to decrease due to low $w_{h,t}$. In the long run, nobody can afford education and only sector T is in operation.

By contrast, when the tax rate is in the intermediate range, i.e. $\tau \in [\overline{\tau}_{h\downarrow}^s, \overline{\tau}_{h\downarrow}^b]$, the dynamics of Fr_t are affected by the initial distribution of wealth as well. If Fr_0 is small either because the economy's wealth is concentrated in the few rich or because the wealth is dispersed among the many poor (the region with hatched lines in the figure), the proportion of unskilled workers is high and thus the unskilled wage is too low for their descendants to gain access to education. As a result, Fr_t remains unchanged, so as other aspects of the economy.

Alternatively, if Fr_0 is relatively high, i.e. Fr_0 is located at the right side of $b_h^*(w_h) = e$ or $b_l^*(w_l) = e$ in the figure, the long-run outcome of the economy critically depends on the tax rate. When $\tau \in [\overline{\tau}_{h\downarrow}^s, \tau^{*s})$ or $\tau \in (\tau^{*b}, \overline{\tau}_{h\downarrow}^b]$, with high H_t, the skilled wage is not high enough for all progenies of current skilled workers to access education, i.e. $b_h^*(w_{h,t}) < e$. Hence, Fr_t decreases over time, and

[33] Because an increase in A_T lowers w_h when $L_T > 0$ (equation 24), higher A_T raises $\overline{\tau}_{h\downarrow}^s$ and lowers $\overline{\tau}_{h\downarrow}^b$. That is, in the long run, $Fr = 0$ is more likely to happen with a 'bad' tax policy, when the productivity of sector T is *higher*. This result and the one in footnote 22 suggest that an appropriate choice of tax rate is particularly important in an economy where the productivity of the traditional sector is *not* very low.

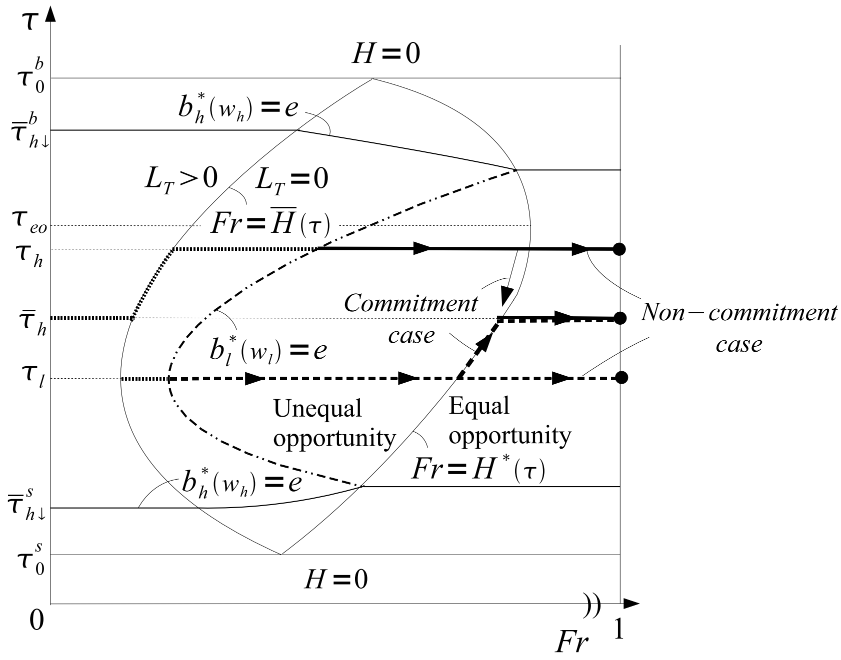

Figure 8: The dynamics when $\frac{A_T}{e}$ is low and τ is endogenous

in the unequal opportunity case, H_t and $w_{l,t}$ fall, while $w_{h,t}$ and the wage inequality go up. Decreases in Fr_t and H_t stop once $w_{h,t}$ becomes high enough that $b_h^*(w_{h,t}) \geq e$ is satisfied. By contrast, when $\tau \in (\tau^{*s}, \tau^{*b})$, because of high productivity but moderate tax burden in sector M, $b_l^*(w_{l,0}) > e$ is satisfied initially. Thus, Fr_t increases over time, the economy goes through the structural change as described in Proposition 1, and $Fr = 1$, $H = H^*(\tau)$, and $w_h - (1+r)e = w_l$ hold in the long run.

Observe that the long run outcome can be very different depending on the tax rate even when the economy starts with an *identical* distribution of wealth.[34,35] This suggests the critical importance of the distribution of political power, which is analyzed next.

4.2.2 Endogenous tax rate

Based on Figure 7 and Lemma 5 of Section 4.1.2, Figure 8 presents the dynamics of Fr_t and τ_t when $\tau_h < \tau^{*b}$ and $\tau_l > \tau^{*s}$. The heavy solid (heavy dashed) line represents the dynamics when educated (uneducated) individuals determine the tax rate. Short dashed lines represent equilibrium combinations of Fr and τ when Fr_t is time-invariant.

When educated (uneducated) individuals determine the policy, Fr_t increases over time and equal

[34] The efficiency of the governmental service also affects the outcome. If the efficiency declines, i.e. Ω_0 decreases, from (52), $b_l^*(w_l) = e$ shifts to the right (and $Fr = \overline{H}(\tau)$ shifts to the right, $Fr = H^*(\tau)$ and $b_h^*(w_h) = e$ shift to the left, τ_0^s and $\overline{\tau}_{h\downarrow}^s$ increase, τ_0^b and $\overline{\tau}_{h\downarrow}^b$ decrease) in the figure, and thus the economy is more likely to stagnate. Further, when unproductive public services are included into the model (see footnote 14), an increase in the proportion of such services has the same qualitative effects. Consistent with this result, Keefer and Knack (1997) find that the growth of poor countries is negatively affected by institutional quality, such as bureaucratic quality, the pervasiveness of corruption, the risk of expropriation and contract repudiation by the government, and the rule of law.

[35] If sector T is not incorporated in the model, unskilled workers cannot escape from sector M when the tax rate is very high, and thus the difference in the outcome becomes much smaller. In particular, from (51), if $\alpha > (1-\alpha)\max\left\{\frac{\rho\theta}{\sigma-\rho\theta}, 1-\frac{\rho\theta}{\sigma-\rho\theta}\right\}$, $H > 0$ for *any* τ. Hence, ignoring the presence of the informal/traditional sector would underestimate negative effects of a bad tax policy.

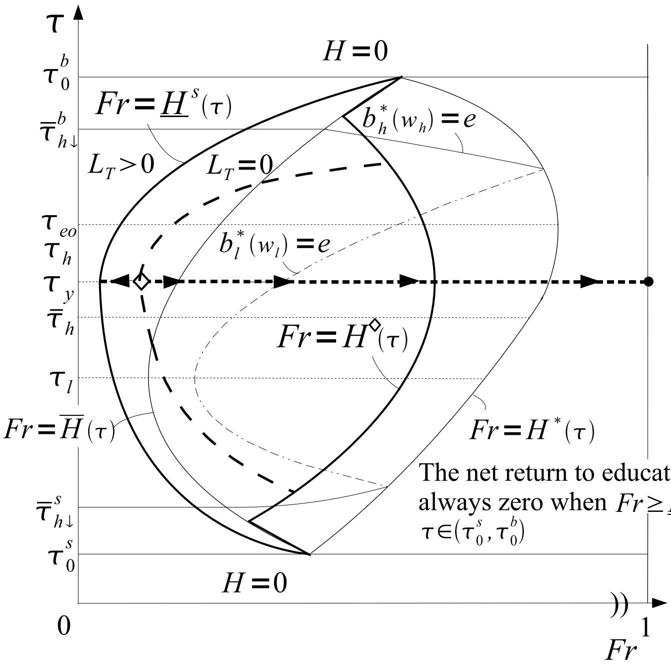

Figure 9: The dynamics when $\frac{A_T}{e}$ is low and the uneducated determine τ and the redistributive policy (the commitment case)

opportunity is attained in the long run, if Fr_0 is located at the right side of $b_l^*(w_l) = e$ at $\tau = \tau_h$ ($\tau = \tau_l$). The minimum level of Fr_0 for reaching equal opportunity is lower, when the uneducated choose the tax rate, because of higher w_l for given Fr. In particular, if Fr_0 is at the left side of $b_l^*(w_l) = e$ at $\tau = \tau_h$ but at the right side of the locus at $\tau = \tau_l$, Fr_t increases (is constant) over time when the uneducated (the educated) determine the policy. Thus, unlike the case $\frac{\gamma_b}{1-\gamma_b(1+r)}A_T > e$, the uneducated should control the policy while Fr_t is in this range. By contrast, when $Fr_t \leq \overline{H}(\tau_l)$, allocating decisive power to the educated is even more important compared to the high productivity case: if the uneducated have decisive power and are concerned about the wage inequality, there exists non-negligible risk that $\tau > \overline{\tau}_{h\downarrow}^b$ or $\tau < \overline{\tau}_{h\downarrow}^s$ is selected and the economy ends up with $Fr = H = 0$.

4.2.3 When the redistributive policy is available

When the redistributive policy is available, the distribution of political power is even more critical for the long-run outcome. When the educated determine policies, unskilled workers always receive A_T from Proposition 4 of Section 4.1.3, thus Fr_t and other aspects of the economy *remain unchanged*.[36] By contrast, when the uneducated choose the policy, all workers receive the same amount (net of the education cost) from Proposition 5, thus the dynamics of Fr_t differ greatly depending on Fr_0 in the commitment case. Figure 9 illustrate the dynamics (the dashed line) when $\tau_h < \tau^{*b}$ and $\tau_l > \tau^{*s}$. For given Fr, τ and the redistributive policy are determined as in Figure 5 of Section 4.1.3. Remember that, when $Fr \leq \underline{H}^s(\tau)$, the uneducated are indifferent among any policies that assure A_T (thus any dashed line is not presented in the figure). By contrast, when $Fr > \underline{H}^s(\tau)$, Fr_t increases (*decreases*)

[36] To be more precise, in the non-commitment case, Fr_t decreases when Fr_t is very high, because all workers receive A_T at $H = \underline{H}^b(\tau_y)$ (see Figure 4 of Section 4.1.3). In the commitment case, Fr_t never decreases from Assumption 3.

over time at the right (left) side of the broken line, which is a combination of $H = Fr$ and τ at which workers receive $\frac{1-\gamma_b(1+r)}{\gamma_b}e$, net of the education cost.[37] Hence, if Fr_0 is at the right side of the broken line at $\tau = \tau_y$, Fr_t increases over time and $H = H^{\diamond}(\tau_y)$ and $Fr = 1$ in the long run, while if Fr_0 is at the left side of the line, Fr_t *decreases* until $Fr_t = \underline{H}^s(\tau_y)$, after which whether Fr_t decreases further or not (constant) depends on chosen policies. The economy can attain successful development from lower Fr_0 than the economy without redistribution (see Figure 8).

The desirable distribution of political power is more evident, compared to the corresponding economy when $\frac{\gamma_b}{1-\gamma_b(1+r)}A_T > e$ (Figures 4 and 5) and the economy without redistribution: the uneducated (the educated) should determine policies for *any* Fr weakly higher (lower) than the level on the broken line at $\tau = \tau_y$.[38] In particular, since the economy *inevitably* stagnates when the educated control policies, the political dominance of the uneducated is a *must* for successful development.

5 Conclusion

This paper has analyzed interactions among taxation, the provision of a productive public service, human capital accumulation, and modernization based on a dynamic dual economy model, which draws on the Becker and Murphy (1992) model of skill and task specialization, and examined conditions for successful development. Distributions of political power and wealth as well as sectoral productivities and the education cost have been found to affect the outcome qualitatively. In particular, the socially desirable distribution of political power is such that educated (uneducated) individuals should have dominant power at an early (late) stage of development. Further, it has been shown that several novel or overlooked inefficiencies arise naturally from realistic features of the model and appropriate redistribution can correct them except at a fairly early stage of development.

Several limitations remain in the analysis. First, the paper has focused mostly on the productive governmental service, although qualities of public services such as the degree of corruption and quality of bureaucracy are found to be important in development empirically (see Mo, 2001, and Easterly, 2007, for example). While footnote 14 has shown that qualitative results are unchanged even when a portion of officers are engaged in unproductive services and footnotes 19 and 34 have shown that the economy is more likely to be underdeveloped and stagnate with an exogenous increase in such services, analyzing endogenous determination of qualities of public services would be valuable. Second, the paper has not endogenized transitions of political power between educated and uneducated individuals, although whether or not and when transitions occur are crucial for the long-run outcome. While modeling transitions within the present framework is straightforward,[39] constructing a plausible model may not

[37] From the definition of the dashed line, it must not intersect with $b_h^*(w_h) = e$, and it must be located at the right side of $Fr = \underline{H}^s(\tau)$ and at the left side of $b_l^*(w_l) = e$, but, unlike the figure, it may be located at the right side of $Fr = \overline{H}(\tau)$.

[38] To be more exact, if the critical $Fr \in (\underline{H}^s(\tau_y), \overline{H}^{opt}(\tau_y))$ of Section 4.1.3 is at the right side of the broken line and a high priority is placed on present efficiency, both groups should have some power when Fr is lower than the critical level and is weakly higher than the level on the broken line at $\tau = \tau_y$.

[39] Suppose that the probability of the power transition from the educated minority (they are the minority when education is appropriately defined) to the uneducated majority increases with the relative per capita wealth and the

be, since Acemoglu et al. (2005, 2008) find no causal effects of income per capita and average years of education on democracy empirically. Improvements in these respects are left for future work.

Appendix: Proofs of lemmas and propositions

Proof of Lemma 1: (i) As for w_l when $L_T=0$,

$$\frac{\partial w_l}{\partial \tau} \gtreqless 0 \Leftrightarrow \frac{d\left[(1-\tau)\left(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}(1-\tau)^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right)^{1-\alpha}\right]}{d\tau} \gtreqless 0, \tag{53}$$

$$\Leftrightarrow \alpha\tau^2+(1-\alpha)\left[2-\alpha\frac{\rho\theta}{\delta-\rho-\theta}\right]\tau-(1-\alpha)^2\frac{\rho\theta}{\delta-\rho-\theta} \lesseqgtr 0 \quad (\because \tau\in[0,1]). \tag{54}$$

At $\tau=0$, the LHS of the last equation is negative, and at $\tau=1$, the LHS$=1+(1-\alpha)\left(1-\frac{\rho\theta}{\delta-\rho-\theta}\right)>0$, hence there exists unique $\tau\in(0,1)$ satisfying $\frac{\partial w_l}{\partial\tau}=0$ and the statement holds. The results for w_h when $L_T>0$ and w_h when $L_T=0$ can be proved in a similar way.

(ii) Regarding w_h when $L_T>0$,

$$\frac{\partial w_h}{\partial\tau}=0 \Leftrightarrow \frac{d\left[(1-\tau)^{\frac{1}{1-\alpha}}\left(\frac{\tau}{1-\tau}\right)^{\frac{\rho\theta}{\delta-\rho-\theta}}\right]}{d\tau}=0 \Leftrightarrow \tau=\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}. \tag{55}$$

It can be proved that, at $\tau=\overline{\tau}_h\equiv\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}$, $\frac{\partial w_l}{\partial\tau}<0$ and $\frac{\partial w_h}{\partial\tau}>0$ when $L_T=0$. □

Proof of Lemma 2: When $L_T>0$, from (24) and $w_l=A_T$, the wages are independent of H and $w_h=0$ at $\tau=0,1$. Thus, when τ is very large or very small, $w_h-(1+r)e<w_l$ holds and $H=0$. $w_h-(1+r)e>w_l$ and thus $H>0$ for an intermediate range of τ iff $w_h-(1+r)e>w_l$ holds at τ maximizing w_h, i.e. $\tau=\overline{\tau}_h$, which is Assumption 1. From Lemma 1, $\frac{\partial w_h}{\partial\tau}>(<)\ 0$ for $\tau<(>)\ \overline{\tau}_h$, hence there exist $\tau_0^s<\overline{\tau}_h$ and $\tau_0^b>\overline{\tau}_h$ satisfying $w_h-(1+r)e=w_l$ and the statement holds. Further, $L_T=0$ and thus $H>0$ cannot hold for $\tau<\tau_0^s$ and $\tau>\tau_0^b$, because, when $L_T=0$, from (19) and $w_l\geq A_T$, L_M is smaller and thus w_h is lower (from equation 20) and w_l is higher compared to the case $L_T>0$. □

Proof of Lemma 3: For any $\tau\in[0,1]$, there exists a single $Fr\in(0,1]$ satisfying (32), and when $\tau=0,1$, $Fr=1$. Since $\left[\frac{(1-\tau)\alpha A_M}{A_T}\right]^{\frac{1}{1-\alpha}}\Omega\left(\frac{H_G}{H}\right)$ in the denominator of the RHS of (32) is proportional to $(w_l)^{\frac{1}{1-\alpha}}$ when $L_T=0$ (from eq. 26), from Lemma 1, $Fr=\overline{H}(\tau)$ decreases (increases) with τ for $\tau<(>)\tau_l$. However, from Lemma 2, $H=0$ when $\tau<\tau_0^s$ or $\tau>\tau_0^b$, hence $Fr=\overline{H}(\tau)$ is defined only for $\tau\in[\tau_0^s,\tau_0^b]$ and its shape is as stated in the lemma. □

Proof of Lemma 4: [Existence of $Fr=H^*(\tau)$] Given $\tau\in(0,1)$, the LHS of (33) is decreasing in Fr, equals $+\infty$ at $Fr=0$, and equals $-\infty$ at $Fr=1$. Thus, for any $\tau\in(0,1)$, there exists a single $Fr\in(0,1)$ satisfying (33). From Lemma 2, $H=0$ for $\tau<\tau_0^s$ and $\tau>\tau_0^b$, so the dividing line is defined only for $\tau\in[\tau_0^s,\tau_0^b]$.

[Existence of $\tau_{eo}\in(0,1)$ and the shape of $Fr=H^*(\tau)$]

relative size of the latter group and the effect of the per capital wealth is stronger. Then, the probability is low and *decreases* with Fr when $L_T>0$, whereas it increases with Fr when $L_T=0$. Given Fr, the probability is lower when the educated implement redistribution.

$$\frac{\partial(\text{LHS of (33)})}{\partial\tau} \gtreqless 0 \;\Leftrightarrow\; \frac{\partial\left\{\left[\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}(1-\tau)^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right]^{1-\alpha}[1-Fr-\alpha(1-\tau)]\right\}}{\partial\tau} \gtreqless 0$$

$$\Leftrightarrow (1-\alpha)\left[\frac{\rho\theta}{\delta-\rho-\theta}\frac{1}{\tau}-\left(1-\frac{\rho\theta}{\delta-\rho-\theta}\right)\frac{1}{1-\tau}-\frac{\alpha}{1-\alpha(1-\tau)}\right]+\frac{\alpha}{1-Fr-\alpha(1-\tau)} \gtreqless 0 \tag{56}$$

$$\Leftrightarrow -(1-\alpha)[1-Fr-\alpha(1-\tau)]\left\{\tau-\frac{\rho\theta[1-\alpha(1-\tau)]}{\delta-\rho-\theta}\right\}+\alpha\tau(1-\tau)[1-\alpha(1-\tau)] \gtreqless 0. \tag{57}$$

When $\tau-\frac{\rho\theta}{\delta-\rho-\theta}[1-\alpha(1-\tau)] \le 0 \Leftrightarrow \tau \le \frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}/[1-\frac{\alpha\rho\theta}{\delta-\rho-\theta}] \in (\overline{\tau}_h,1)$, $\frac{\partial(\text{LHS of (33)})}{\partial\tau} > 0$ for any Fr on the dividing line (where $1-Fr-\alpha(1-\tau)>0$ is satisfied), and thus the dividing line is positively sloped in this interval. When $\tau-\frac{\rho\theta}{\delta-\rho-\theta}[1-\alpha(1-\tau)] > 0$, from (57),

$$\frac{\partial(\text{LHS of (33)})}{\partial\tau} \gtreqless 0 \;\Leftrightarrow Fr \gtreqless [1-\alpha(1-\tau)]\left\{1-\frac{\alpha(1-\tau)\tau}{(1-\alpha)\left\{\tau-\frac{\rho\theta}{\delta-\rho-\theta}[1-\alpha(1-\tau)]\right\}}\right\}$$

$$=[1-\alpha(1-\tau)]\frac{Q(\tau)}{(1-\alpha)\left\{\tau-\frac{\rho\theta}{\delta-\rho-\theta}[1-\alpha(1-\tau)]\right\}}, \tag{58}$$

$$\text{where}\quad Q(\tau)\equiv(1-\alpha)\left\{\tau-\frac{\rho\theta}{\delta-\rho-\theta}[1-\alpha(1-\tau)]\right\}-\alpha(1-\tau)\tau. \tag{59}$$

Since $Q(0)<0$, $Q(1)>0$, and $Q\left(\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}/[1-\frac{\alpha\rho\theta}{\delta-\rho-\theta}]\right)<0$, there exists a single $\tau\equiv\tau_Q\in\left(\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}/[1-\frac{\alpha\rho\theta}{\delta-\rho-\theta}],1\right)$ satisfying $Q(\tau)=0$. For $\tau\le\tau_Q$, the RHS of (58) is non-positive, hence $\frac{\partial(\text{LHS of (33)})}{\partial\tau}>0$ for any $Fr\in(0,1)$ and $Fr=H^*(\tau)$ is positively sloped. For $\tau>\tau_Q$, on the other hand, the RHS of (58) is positive, hence the sign of $\frac{\partial(\text{LHS of (33)})}{\partial\tau}$ must be checked for each combination of (Fr,τ) on the dividing line. By substituting (58) into (33),

$$\frac{\partial(\text{LHS of (33)})}{\partial\tau} \gtreqless 0 \quad \text{at } (Fr,\tau) \text{ on } Fr=H^*(\tau) \text{ (eq. 33)}$$

$$\Leftrightarrow A_M{\Omega_0}^{1-\alpha}\left(\frac{\alpha(1-\tau)}{1-\alpha(1-\tau)}\frac{Q(\tau)+\tau}{Q(\tau)}\right)^{\alpha}\left(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\right)^{1-\alpha}\frac{[1-\alpha(1-\tau)]\tau}{Q(\tau)+\tau}\gtreqless(1+r)e \tag{60}$$

$$\Leftrightarrow A_M{\Omega_0}^{1-\alpha}\alpha^{\alpha}(1-\alpha)^{(1-\alpha)(1-\frac{\rho\theta}{\delta-\rho-\theta})}\frac{\tau^{1+(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta}}(1-\tau)^{1-(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta}}}{\{Q(\tau)\}^{\alpha}\{Q(\tau)+\tau\}^{1-\alpha}} \gtreqless (1+r)e. \tag{61}$$

Now it is proved that there exists a unique $\tau_{eo}\in(\tau_Q,1)$ satisfying the above equation with equality. The shape of the LHS of the equation must be examined.

$$\frac{d\left[\frac{\tau^{1+(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta}}(1-\tau)^{1-(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta}}}{\{Q(\tau)\}^{\alpha}\{Q(\tau)+\tau\}^{1-\alpha}}\right]}{d\tau} \gtreqless 0$$

$$\Leftrightarrow \left[1+\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\right]\frac{1}{\tau}-\left[1-\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\right]\frac{1}{1-\tau}-\alpha\frac{Q'(\tau)}{Q(\tau)}-(1-\alpha)\frac{Q'(\tau)+1}{Q(\tau)+\tau}\gtreqless 0, \tag{62}$$

$$\Leftrightarrow Q(\tau)[Q(\tau)+\tau]\left[1+\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}-2\tau\right]-\tau(1-\tau)\{\alpha[Q(\tau)+\tau]Q'(\tau)+(1-\alpha)Q(\tau)[Q'(\tau)+1]\}\gtreqless 0. \tag{63}$$

Since $1+(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta}-2\tau=\frac{1-\alpha-Q'(\tau)}{\alpha}$, (63) is equivalent to

$$Q(\tau)[Q(\tau)+\tau]\{1-\alpha-Q'(\tau)\}-\alpha\tau(1-\tau)\left\{Q(\tau)Q'(\tau)+\alpha\tau Q'(\tau)+(1-\alpha)Q(\tau)\right\}\gtreqless 0, \tag{64}$$

$$\Leftrightarrow [Q(\tau)+\alpha\tau]\{Q(\tau)[1-\alpha-Q'(\tau)]-\alpha\tau(1-\tau)Q'(\tau)\}+(1-\alpha)\tau Q(\tau)\{1-\alpha-Q'(\tau)-\alpha(1-\tau)\}\gtreqless 0. \tag{65}$$

Because $Q(\tau)+\alpha\tau=[1-\alpha(1-\tau)]\left\{\tau-\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\right\}$ and $1-\alpha-Q'(\tau)-\alpha(1-\tau)=-\alpha\{\tau-\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\}$, the above equation is equivalent to

$$\Big[\tau-\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\Big]\Big\{[1-\alpha(1-\tau)][Q(\tau)(1-\alpha-Q'(\tau))-\alpha\tau(1-\tau)Q'(\tau)]-\alpha(1-\alpha)\tau Q(\tau)\Big\}\gtreqless 0, \tag{66}$$

$$\Leftrightarrow\Big[\tau-\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\Big]\Big\{(1-\alpha)Q(\tau)(1-\alpha-Q'(\tau))-\alpha\tau Q'(\tau)\big\{Q(\tau)+(1-\tau)[1-\alpha(1-\tau)]\big\}\Big\}\gtreqless 0, \tag{67}$$

$$\Leftrightarrow(1-\alpha)\Big[\tau-\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\Big]\Big\{Q(\tau)(1-\alpha-Q'(\tau))-\alpha\tau Q'(\tau)\Big(1-\tfrac{[1-\alpha(1-\tau)]\rho\theta}{\delta-\rho-\theta}\Big)\Big\}\gtreqless 0, \tag{68}$$

$$\Leftrightarrow R(\tau)\equiv\alpha(1-\alpha)\Big[\tau-\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\Big]\Big([1-\alpha(1-\tau)]\Big[1+\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}-2\tau\Big]\Big(\tau-\tfrac{[1-\alpha(1-\tau)]\rho\theta}{\delta-\rho-\theta}\Big)-(1-\alpha)\tau\Big\{1-\tfrac{[1-\alpha(1-\tau)]\rho\theta}{\delta-\rho-\theta}\Big\}\Big)\gtreqless 0. \tag{69}$$

Since $Q(\tau)>0$ for $\tau>\tau_Q$, from (59),

$$(1-\alpha)\big\{1-\tfrac{[1-\alpha(1-\tau)]\rho\theta}{\delta-\rho-\theta}\big\}>(1-\tau)[1-\alpha(1-\tau)]. \tag{70}$$

By using the above inequality into the LHS of (69),

$$R(\tau)<\alpha(1-\alpha)\Big[\tau-\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\Big][1-\alpha(1-\tau)]\Big\{\Big[1+\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}-2\tau\Big]\Big(\tau-\tfrac{[1-\alpha(1-\tau)]\rho\theta}{\delta-\rho-\theta}\Big)-\tau(1-\tau)\Big\}\Big(\because\ \tau>\tau_Q>\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\Big) \tag{71}$$

$$<\alpha(1-\alpha)\Big[\tau-\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}\Big][1-\alpha(1-\tau)]\Big\{\Big[\tfrac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}-\tau\Big]\Big(\tau-\tfrac{[1-\alpha(1-\tau)]\rho\theta}{\delta-\rho-\theta}\Big)-(1-\tau)\tfrac{[1-\alpha(1-\tau)]\rho\theta}{\delta-\rho-\theta}\Big\}\Bigg(\because\ \tau>\tau_Q>\frac{\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}}{1-\frac{\alpha\rho\theta}{\delta-\rho-\theta}}\Bigg) \tag{72}$$

$$<0. \tag{73}$$

Hence, the LHS of (61) decreases with τ for $\tau>\tau_Q$. Further, LHS $\to+\infty$ as $\tau\to\tau_Q$ from above and equals 0 at $\tau=1$. Therefore, there exists a unique $\tau_{eo}\in(\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}/[1-\frac{\alpha\rho\theta}{\delta-\rho-\theta}],1)$ satisfying $\frac{\partial(\text{LHS of (33)})}{\partial\tau}=0$ on the dividing line. From Lemma 2, $H=0$ at $\tau<\tau_0^s$ and $\tau>\tau_0^b$, so the dividing line is defined for $\tau\in[\tau_0^s,\tau_0^b]$. Since $\frac{\partial(\text{LHS of (33)})}{\partial\tau}\gtreqless 0\Leftrightarrow\tau\lesseqgtr\tau_{eo}$ and the LHS of (33) is decreasing in Fr, the shape of the dividing line is as stated in the lemma.

[$\tau_{eo}>\tau_h$] Since $\frac{\partial(\text{LHS of (33)})}{\partial\tau}=\frac{\partial w_h}{\partial\tau}-\frac{\partial w_l}{\partial\tau}=0$ and thus $\frac{\partial w_h}{\partial\tau}$ and $\frac{\partial w_l}{\partial\tau}$ must have a same sign at $\tau=\tau_{eo}$, $\tau_{eo}>\tau_h$ or $\tau_{eo}<\tau_l$. If $\tau_{eo}<\tau_l$, $\frac{\partial w_h}{\partial\tau}>0$ and $\frac{\partial w_l}{\partial\tau}<0$ and thus the dividing line would be positively sloped for $\tau\in(\tau_l,\tau_h)$, which is inconsistent with the shape of the line proved just above.

[The intersection with $Fr=\overline{H}(\tau)$] Straightforward from the fact that $w_l=A_T$ and $w_h-(1+r)e=w_l$ are satisfied at the intersection. □

Proof of Proposition 1: Except (ii)(a), the results are straightforward. (ii)(a) Since $\frac{H_{G,t}}{H_t}=\frac{\tau}{1-\alpha(1-\tau)}$ (equation 23), when H_t increases, both $H_{M,t}$ and $H_{G,t}$ increase proportionally. Further, since the RHS of (19) equals A_T in this case, H_t and $L_{M,t}$ increase proportionally, while $L_{T,t}$ decreases. □

Proof of Proposition 2: (i) When $\tau\in(\tau_0^s,\tau_0^b)$, if $H\le\overline{H}(\tau)$ and thus $w_l=A_T$, $\widetilde{Y}_L=[w_h-(1+r)e]H+A_T(1-H)$, where w_h is given by (24) and thus independent of H. Hence, $\frac{\partial\widetilde{Y}_L}{\partial H}=w_h-(1+r)e-A_T>0$ for $H<\overline{H}(\tau)$ from Lemma 2 and (24), which implies $H^{\diamond}(\tau)\ge\overline{H}(\tau)$.

Let $\Psi(\tau)\equiv A_M{\Omega_0}^{1-\alpha}\Big(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\Big)^{1-\alpha}$. If $H>\overline{H}(\tau)$ and thus $L_T=0$, $\frac{\partial\widetilde{Y}_L}{\partial H}=[w_h-(1+r)e-w_l]+\frac{\partial w_h}{\partial H}H+\frac{\partial w_l}{\partial H}(1-H)$, where $\frac{\partial w_h}{\partial H}=-\alpha[1-\alpha(1-\tau)]\Psi(\tau)(H)^{-\alpha-1}(1-H)^{\alpha-1}$ from

(25), $\frac{\partial w_l}{\partial H} = (1-\alpha)\alpha(1-\tau)\Psi(\tau)(H)^{-\alpha}(1-H)^{\alpha-2}$ from (26), and $w_h - (1+r)e - w_l = [1-\alpha(1-\tau) - H]\Psi(\tau)(H)^{-\alpha}(1-H)^{\alpha-1} - (1+r)e$ from (25) and (26). Thus,

$$\frac{\partial \widetilde{Y}_L}{\partial H} = (1-\alpha-H)\Psi(\tau)(H)^{-\alpha}(1-H)^{\alpha-1} - (1+r)e, \tag{74}$$

$$\frac{\partial^2 \widetilde{Y}_L}{\partial H^2} = -\alpha(1-\alpha)\Psi(\tau)(H)^{-\alpha-1}(1-H)^{\alpha-2} < 0. \tag{75}$$

Besides (75), since $lim_{H\to 0}\frac{\partial \widetilde{Y}_L}{\partial H} = +\infty, lim_{H\to 1-\alpha}\frac{\partial \widetilde{Y}_L}{\partial H} = -(1+r)e < 0$, and $\frac{\partial \widetilde{Y}_L}{\partial H} < 0$ when $H > 1-\alpha$, for given τ, there exists unique $H \in (0, 1-\alpha)$ satisfying $\frac{\partial \widetilde{Y}_L}{\partial H} = 0$. If such H is greater than $\overline{H}(\tau)$, it is $H^{\diamond}(\tau)$, otherwise, $H^{\diamond}(\tau) = \overline{H}(\tau)$.

When $H^{\diamond}(\tau) > \overline{H}(\tau)$, since $\frac{\partial w_h}{\partial H}H + \frac{\partial w_l}{\partial H}(1-H) = -\alpha\tau\Psi(\tau)(H)^{-\alpha}(1-H)^{\alpha-1} < 0$, $w_h - (1+r)e - w_l > 0$ when $H = H^{\diamond}(\tau)$ and thus $H^{\diamond}(\tau) < H^*(\tau)$.

(ii) The result on H is from Lemma 4 and Proposition 1 (ii) and the rest of the proof is for the result on $\widetilde{Y}_L$. In the equal opportunity case, $\widetilde{Y}_L = w_h - (1+r)e$ and (33) with Fr replaced by H is satisfied. From this equation and (25),

$$\frac{w_h}{1-\alpha(1-\tau)} = (1+r)e\frac{1-H}{1-H-\alpha(1-\tau)} \Leftrightarrow H = \frac{[1-\alpha(1-\tau)][w_h-(1+r)e]}{w_h-[1-\alpha(1-\tau)](1+r)e}. \tag{76}$$

Substituting the above equation into (25),

$$w_h = A_M{\Omega_0}^{1-\alpha}\left(\frac{\alpha(1-\tau)w_h}{w_h-(1+r)e}\right)^{\alpha}\left(\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}\left[(1-\tau)(1-\alpha)\right]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}\right)^{1-\alpha} \tag{77}$$

$$\Leftrightarrow (w_h)^{1-\alpha}[w_h-(1+r)e]^{\alpha} = A_M{\Omega_0}^{1-\alpha}\alpha^{\alpha}(1-\alpha)^{(1-\alpha)(1-\frac{\rho\theta}{\delta-\rho-\theta})}\tau^{\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}}(1-\tau)^{1-\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}}. \tag{78}$$

The LHS of (78) increases with w_h and $\frac{\partial \text{RHS}}{\partial\tau} \gtreqless 0 \Leftrightarrow \tau \lesseqgtr \overline{\tau}_h = \frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}$. □

Proof of Lemma 5: [*Non-commitment case*] The (subgame perfect Nash) equilibrium tax rate is obtained by examining the determination of the rate given the number of educated workers first and then considering education decisions in childhood. Denote the number of educated workers by E.

(i) When $E \le H^*(min\{\tau_h, \tau_0^b\})$, the wage-maximizing tax rate equals the rate in the lemma at $Fr = E$ (and $H = E$), which is proved from the shape of $Fr = \overline{H}(\tau)$ (Lemma 3) and the association of w_h with τ and $\tau_l < \overline{\tau}_h < \tau_h$ (Lemma 1). When $E > H^*(min\{\tau_h, \tau_0^b\})$, $\tau = min\{\tau_h, \tau_0^b\}$ and $H = E$ if $w_h \ge w_l$ at such τ and E, otherwise, $\tau = \overline{\tau}_h$ and H ($\in (H^*(\overline{\tau}_h), E)$) is determined so that $w_h = w_l$, which is proved from Lemma 1, the shape of $Fr = H^*(\tau)$ (Lemma 4), $\frac{\partial w_h}{\partial H} < 0$ when $L_T = 0$, and H and τ satisfying $w_h = w_l$ is given by $H = 1-\alpha(1-\tau)$ and $w_h = w_l$ is maximized at $\tau = \overline{\tau}_h$. As for education decisions, if $Fr \le H^*(min\{\tau_h, \tau_0^b\})$, the net return to education at $E = Fr$ is non-negative, thus $E(= H) = Fr$ and the equilibrium tax rate equals the wage-maximizing rate at $E = Fr$. Otherwise, $E = H^*(min\{\tau_h, \tau_0^b\}) < Fr$ and $\tau = min\{\tau_h, \tau_0^b\}$ because the net return is negative if $E > H^*(min\{\tau_h, \tau_0^b\})$.

(ii) When $E \le H^*(max\{\tau_l, \tau_0^s\})$, the wage-maximizing tax rate equals the rate in the lemma at $Fr = E$, which can be shown based on shapes of the curves (Lemmas 3 and 4), $w_l = A_T$ when $L_T > 0$, and the association of w_l (when $L_T = 0$) with τ and $\tau_l < \overline{\tau}_h$ (Lemma 1). When $E > H^*(max\{\tau_l, \tau_0^s\})$, $\tau = max\{\tau_l, \tau_0^s\}$ and $H = E$ if $w_h \ge w_l$ at such τ and E, otherwise, τ equals the minimum of $\overline{\tau}_h$

and $\tau(> max\{\tau_l, \tau_0^s\})$ satisfying $w_h = w_l$ at E (and H is determined so that $w_h = w_l$), which can be proved based on Lemmas 1 and 4, $\frac{\partial w_l}{\partial H} > 0$ when $L_T = 0$, H and τ satisfying $w_h = w_l$ is given by $H = 1-\alpha(1-\tau)$, and $w_h = w_l$ increases (decreases) with τ for $\tau < (>)\overline{\tau}_h$. As for education decisions in childhood, if $Fr \leq H^*(max\{\tau_l, \tau_0^s\})$, the net return to education at $E = Fr$ is non-negative, thus $E = Fr$ and the equilibrium tax rate equals the wage-maximizing rate at $E = Fr$. Otherwise, $E = H^*(max\{\tau_l, \tau_0^s\})(< Fr)$ and $\tau = max\{\tau_l, \tau_0^s\}$ since, if $E > H^*(max\{\tau_l, \tau_0^s\})$, the net return is negative with the wage-maximizing rate.

[*Commitment case*] In this case, τ and E are determined simultaneously.

(i) If $Fr \leq H^*(min\{\tau_h, \tau_0^b\})$, w_h when $H = E = H^*(\tau) < Fr$, if such τ exists, is lower than w_h when $H = E = Fr$ and τ equals the rate in the commitment case, because w_h is maximized at such rate when $H = Fr$ and w_h when $H = H^*(\tau)$ increases (decreases) with τ for $\tau < (>)\overline{\tau}_h$ from Proposition 2 (ii); thus, the result is same as the non-commitment case. If $Fr > H^*(min\{\tau_h, \tau_0^b\})$, since w_h when $H = H^*(\tau)$ decreases with τ for $\tau > \overline{\tau}_h$, w_h when $H = H^*(\tau)$ and $\tau = min\{\tau_h, \tau_0^b\}$ is lower than the one when $H = H^*(\tau)$ and $\tau = \overline{\tau}_h$, thus the result is as stated in the lemma.

(ii) If $Fr \leq \overline{H}(max\{\tau_l, \tau_0^s\})$, $w_l = A_T$ for any τ and $H \leq Fr$, thus the result is same as before. If $Fr \in (\overline{H}(max\{\tau_l, \tau_0^s\}), H^*(max\{\tau_l, \tau_0^s\})]$, since w_l when $H = E = H^*(\tau) < Fr$ (such τ exists only when $\tau_l > \tau_0^s$ and thus $\tau < \tau_l$) is lower than w_l when $H = E = Fr$ and $\tau = \tau_l$ from the fact that w_l is maximized at $\tau = \tau_l$ when $H = Fr$ and w_l when $H = H^*(\tau)$ increases with τ for $\tau < \overline{\tau}_h$ (from Proposition 2 (ii)), the result is same as before. When $Fr > H^*(max\{\tau_l, \tau_0^s\})$, since w_l when $H = H^*(\tau)$ increases(decreases) with τ for $\tau < (>)\overline{\tau}_h$, the result is as stated in the lemma. $\square$

Proof of Proposition 3: (i) When $L_T = 0$, $\widetilde{Y}_L = A_M(\Omega_0\Psi_0(\tau))^{1-\alpha}(H)^{1-\alpha}(1-H)^{\alpha}-(1+r)eH$, where $\Psi_0(\tau) \equiv \frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\alpha)(1-\tau)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}$, from (25) and (26). Then, $\frac{\partial \widetilde{Y}_L}{\partial \tau} = \frac{\partial \Psi_0(\tau)}{\partial \tau}(1-\alpha)(\Psi_0(\tau))^{-\alpha}A_M{\Omega_0}^{1-\alpha}$ $(H)^{1-\alpha}(1-H)^{\alpha}$, where

$$\frac{\partial \Psi_0(\tau)}{\partial \tau} = \Psi_0(\tau)\Big[\frac{\rho\theta}{\delta-\rho-\theta}\frac{1}{\tau} - \Big(1 - \frac{\rho\theta}{\delta-\rho-\theta}\Big)\frac{1}{1-\tau} - \frac{\alpha}{1-\alpha(1-\tau)}\Big] = \frac{(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta} - \Big(1-\alpha\frac{\rho\theta}{\delta-\rho-\theta}\Big)\tau}{\tau(1-\tau)[1-\alpha(1-\tau)]}. \tag{79}$$

Thus, $\frac{\partial \Psi_0(\tau)}{\partial \tau} \gtreqless 0 \Leftrightarrow \tau \lesseqgtr \tau_y \equiv \frac{(1-\alpha)\frac{\rho\theta}{\delta-\rho-\theta}}{1-\alpha\frac{\rho\theta}{\delta-\rho-\theta}}$. From Lemma 1 (ii), $\tau_y > \overline{\tau}_h$, and from (25), $w_h \propto$ $[1-\alpha(1-\tau)](\Psi_0(\tau))^{1-\alpha}$ and thus $\tau_y < \tau_h$. Further, $\tau_y < \tau_0^b$ must be true because $\widetilde{Y}_L \geq A_T$ for any τ and $\widetilde{Y}_L = A_T$ at $\tau = \tau_0^b$. By contrast, when $L_T > 0$, $\widetilde{Y}_L = [w_h - (1+r)e]H + A_T(1-H)$, where w_h is given by (24), and $\frac{\partial \widetilde{Y}_L}{\partial \tau} > (<)0$ for $\tau < (>)\overline{\tau}_h$ from Lemma 1. Based on these results, the tax rate maximizing $\widetilde{Y}_L$ for given Fr can be obtained as in the proof of the commitment case of Lemma 5 (i). Finally, the association between $\widetilde{Y}_L$ and τ can be proved from these results and the shape of $Fr = \overline{H}(\tau)$ (Lemma 3, see Figure 3).

(ii) $\widetilde{Y}_L = A_M {L_M}^{\alpha}\big[\Omega\big(\frac{H_G}{H}\big)H\big]^{1-\alpha} + A_T(1-H-L_M) - (1+r)eH$ from (18), where $L_M \leq 1-H$ and $H \leq Fr$ are determined so as to maximize $\widetilde{Y}_L$. Thus, for given Fr, the optimal $\frac{H_G}{H}$ is the one maximizing $\Omega\big(\frac{H_G}{H}\big)$, which equals $\frac{\rho\theta}{\delta-\rho-\theta}$ from (15) and the corresponding τ in a decentralized economy is τ_y from (23). (iii) From (i), (ii), and Lemma 5. $\square$

Proof of Proposition 4: [Existence of $\underline{H}^b(\tau)$ and $\underline{H}^b(\tau) > H^*(\tau)$]When $L_M = 1-H$, $\widetilde{Y}_L$ equals $A_M(\Omega_0\Psi_0(\tau))^{1-\alpha}(H)^{1-\alpha}(1-H)^{\alpha} - (1+r)eH$ from the proof of Proposition 3 (i). From properties of $\widetilde{Y}_L$ in the proof of Proposition 2 (i), $lim_{H\to 0}\widetilde{Y}_L = 0$, and $lim_{H\to 1}\widetilde{Y}_L = -(1+r)e < 0$, for given $\tau \in (\tau_0^s, \tau_0^b)$, there exist two $H \in (0,1)$ satisfying $\widetilde{Y}_L = A_T$. The greater one, $\underline{H}^b(\tau)$, is higher than $H^*(\tau)$, since $w_l(\tau,H,1-H) > A_T$ and thus $w_h(\tau,H,1-H) - (1+r)e < w_l(\tau,H,1-H)$ at $H = \underline{H}^b(\tau)$, which can be shown from $\widetilde{Y}_L > A_T$ at H satisfying $\frac{\partial\widetilde{Y}_L}{\partial H} = 0$ and $\frac{\partial\widetilde{Y}_L}{\partial H} < (>)0$ for higher (lower) H (from the proof of Proposition 2 (i)) and $\widetilde{Y}_L > A_T$ at $H = \overline{H}(\tau)$.

(i) As long as the redistributive policy is feasible, i.e. $w_l(\tau,H,L_M) > A_T$, and is implemented, i.e. $x, T > 0$, (40) holds with equality. Then, $xw_l(\tau,H,L_M) = w_l(\tau,H,L_M) - A_T$ and the objective function equals $w_h(\tau,H,L_M) + \frac{L_M}{H}[w_l(\tau,H,L_M) - A_T]$. Hence, in this case, the problem is simplified as:

$$max_{\{L_M,\tau\}}\left\{w_h(\tau,H,L_M) + \tfrac{L_M}{H}[w_l(\tau,H,L_M) - A_T]\right\} \tag{80}$$

$$s.t. \quad w_h(\tau,H,L_M) + \tfrac{L_M}{H}[w_l(\tau,H,L_M) - A_T] - (1+r)e \geq A_T, \tag{39'}$$

$$(42) \text{ and } L_M \leq 1-H.$$

The derivative of the objective function with respective to L_M is, from (19) and (20),

$$\frac{\partial w_h(\tau,H,L_M)}{\partial L_M} + \frac{L_M}{H}\frac{\partial w_l(\tau,H,L_M)}{\partial L_M} + \frac{1}{H}[w_l(\tau,H,L_M) - A_T] = \frac{1}{H}\left[\frac{w_l(\tau,H,L_M)}{1-\tau} - A_T\right], \tag{81}$$

which is positive whenever $w_l(\tau,H,L_M) > A_T$. Thus, if $w_l(\tau,H,1-H) > A_T$, i.e. $H > \overline{H}(\tau)$, $L_M = 1-H$ and $x = \frac{w_l(\tau,H,1-H)-A_T}{w_l(\tau,H,1-H)}$; otherwise, L_M is determined so that $w_l(\tau,H,L_M) = A_T$ and $x = T = 0$. Hence, when $H > \overline{H}(\tau)$, the objective function is expressed as $\frac{1}{H}\{w_h(\tau,H,1-H)H + [w_l(\tau,H,1-H) - A_T](1-H)\} = \frac{1}{H}\{\widetilde{Y}_L - A_T(1-H)\} + (1+r)e$, which increases (decreases) with τ for $\tau < (>)\tau_y \in (\overline{\tau}_h, min\{\tau_h, \tau_0^b\})$ and is maximized at $\tau = \tau_y$ from the proof of Proposition 3 (i), and approaches $w_h(\tau,H,1-H)$ as $w_l(\tau,H,1-H) \to A_T$. When $H \leq \overline{H}(\tau)$ and thus $x = T = 0$, $\frac{\partial w_h}{\partial\tau} > (<)0$ for $\tau < (>)\overline{\tau}_h$ from Lemma 1.

(a) From the above result and the shape of $Fr = \overline{H}(\tau)$, when $w_l(\tau_y,Fr,1-Fr) \leq A_T$, i.e. $Fr \leq \overline{H}(\tau_y)$, $x = T = 0$, $H = Fr$, L_M is determined so that $w_l(\tau,Fr,L_M) = A_T$, and τ coincides with the rate described in Proposition 3 (i).

(b) By contrast, when $Fr > \overline{H}(\tau_y)$, if the redistributive policy is implemented, $\tau = \tau_y$, $L_M = 1-H$, and $H = Fr$ for $Fr \leq \underline{H}^b(\tau_y)$ and $H = \underline{H}^b(\tau_y)$ otherwise (note eq. 42). If the policy is not implemented, $\tau = \overline{\tau}_h$, $L_M = 1-H$, and $H = H^*(\overline{\tau}_h)$ for large $Fr(> H^*(\overline{\tau}_h))$ from Lemma 5 (i) and the skilled wage is highest among H and τ satisfying $H = H^*(\tau)$ from Proposition 2 (ii). The objective function when $x, T > 0$, $\frac{1}{H}\{\widetilde{Y}_L - A_T(1-H)\} + (1+r)e$, decreases with H since

$$\frac{\partial\left[\frac{\widetilde{Y}_L - A_T(1-H)}{H}\right]}{\partial H} = \frac{\frac{\partial\widetilde{Y}_L}{\partial H}H - \widetilde{Y}_L + A_T}{H^2} = \frac{A_T - w_l}{H^2} < 0 \tag{82}$$

from (74) in the proof of Proposition 2 (i). Thus, from $H^*(\overline{\tau}_h) < H^*(\tau_y)$ (from Lemma 4) and Proposition 2 (ii), there exists $Fr \in (H^*(\overline{\tau}_h), H^*(\tau_y))$ such that the educated strictly prefer the policy without (with) redistribution for higher (lower) Fr.

(ii) The maximization problem determining policies is different from the commitment case in that (39) is replaced by $(1-x)w_h(\tau,H,L_M)+T \geq (1-x)w_l(\tau,H,L_M)$ and (42) does not appear. As in the previous case, when $w_l(\tau,H,L_M) > A_T$ and $x, T > 0$, $(1-x)w_l(\tau,H,L_M) = A_T$ holds and the problem can be simplified. The difference is that $-(1+r)e$ does not appear in the LHS of the constraint for skilled workers (the constraint is looser) and (42) does not appear in the problem. When $H \leq \overline{H}(\tau_y)$ or when $H > \overline{H}(\tau_y)$ and the redistributive policy is implemented, the proof of the commitment case applies and chosen policies are same. When $H > \overline{H}(\tau_y)$ and $x = T = 0$, by contrast, $L_M = 1-H$ and $\tau = min\{\tau_h, \tau_0^b\}(> \tau_y > \overline{\tau}_h)$ from Lemma 5 (i). Since τ_y is the rate maximizing net aggregate wage and $w_l(min\{\tau_h, \tau_0^b\},H,1-H) \geq A_T(=$ the amount unskilled workers receive when $x, T > 0)$, for given H, the policy with redistribution is always preferred to the one with $x = T = 0$.

Finally, taking into account the policy choice, H is determined. When $Fr \leq \overline{H}(\tau_y)$, $H = Fr$ as before. When $Fr > \overline{H}(\tau_y)$, since policies with $x, T > 0$ are always selected, $H = Fr$ for $Fr \leq \underline{H}^b(\tau_y)$ and $H = \underline{H}^b(\tau_y)$ otherwise. □

Proof of Proposition 5: (ii) is straightforward and is as explained in the main text.

(i) As long as the redistributive policy is feasible, i.e. $w_h(\tau,H,L_M)-(1+r)e-w_l(\tau,H,L_M) > 0$, and is implemented, i.e. $x, T > 0$, (45) holds with equality. Then, from (45) and (47), $(1-x)w_h(\tau,H,L_M)-(1+r)e = w_l(\tau,H,L_M)+\frac{H}{L_M}xw_h(\tau,H,L_M) \Leftrightarrow xw_h(\tau,H,L_M) = \frac{L_M}{H+L_M}[w_h(\tau,H,L_M)-(1+r)e-w_l(\tau,H,L_M)]$ and thus the objective function becomes $\frac{H}{H+L_M}[w_h(\tau,H,L_M)-(1+r)e]+\frac{L_M}{H+L_M}w_l(\tau,H,L_M)$.

Hence, in this case, the maximization problem is simplified as:

$$max_{\{H,\tau\}}\left\{\frac{H}{H+L_M}[w_h(\tau,H,L_M)-(1+r)e]+\frac{L_M}{H+L_M}w_l(\tau,H,L_M)\right\} \tag{83}$$

$$s.t. \quad \frac{H}{H+L_M}[w_h(\tau,H,L_M)-(1+r)e]+\frac{L_M}{H+L_M}w_l(\tau,H,L_M) \geq A_T, \tag{46'}$$

(49) and $H \leq Fr$.

Suppose (46') does not bind. Then, $L_M = 1-H$ and the objective function equals $\widetilde{Y}_L = A_M(\Omega_0\Psi_0(\tau))^{1-\alpha}(H)^{1-\alpha}(1-H)^{\alpha}-(1+r)eH$, where $\Psi_0(\tau) \equiv \frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\alpha)(1-\tau)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}$, from (25) and (26). From the proof of Proposition 3 (i), $\frac{\partial\widetilde{Y}_L}{\partial\tau} \gtreqless 0$ for $\tau \lesseqgtr \tau_y \in (\overline{\tau}_h, min\{\tau_h, \tau_0^b\})$, thus, as long as $\widetilde{Y}_L > A_T$ at $\tau = \tau_y$, τ_y is the chosen rate. From the proof of Proposition 2 (i), for given $\tau \in (\tau_0^s, \tau_0^b)$, there exists unique $H \in (0, H^*(\tau))$ satisfying $\frac{\partial\widetilde{Y}_L}{\partial H} = 0$ and $\frac{\partial\widetilde{Y}_L}{\partial H} > (<)0$ for lower (higher) H. From Assumption 2 and Proposition 2 (i), at $\tau = \tau_y$, such H equals $H^{\diamond}(\tau_y)$ from $H^{\diamond}(\tau_y) > \overline{H}(\tau_y)$. Hence, as long as $\widetilde{Y}_L > A_T$ at $\tau = \tau_y$, $H = Fr$ when $Fr \leq H^{\diamond}(\tau_y)$ and $H = H^{\diamond}(\tau_y)$ otherwise. Such policy is feasible since $H^{\diamond}(\tau) < H^*(\tau)$. Further, the policy is preferred to $x = T = 0$, because, given Fr, it maximizes net aggregate labor income and $(1-x)w_l+T = \widetilde{Y}_L$.

Now binding (46') is taken into account. The existence of H satisfying (46') with equality when $L_M = 1-H$, that is, the existence of $\underline{H}^s(\tau)$, is from the proof of Proposition 4. Clearly, $\underline{H}^s(\tau) < H^{\diamond}(\tau)$. Further, $\underline{H}^s(\tau) < \overline{H}(\tau)$, that is, $w_l(\tau,\underline{H}^s(\tau),1-\underline{H}^s(\tau)) < A_T$, from $\widetilde{Y}_L = A_T$ and $w_h(\tau,H,1-H)-(1+r)e > w_l(\tau,H,1-H)$ at $H = \underline{H}^s(\tau)$ (note $\underline{H}^s(\tau) < H^{\diamond}(\tau) < H^*(\tau)$). Since $\frac{\partial\widetilde{Y}_L}{\partial H} > 0$ at $H = \underline{H}^s(\tau)$, $Fr =$

$\underline{H}^s(\tau)$ is negatively (positively) sloped for $\tau < (>)\tau_y$ on the (Fr, τ) plane and thus $Fr = \underline{H}^s(\tau)$ is lowest at $\tau = \tau_y$. Hence, when $Fr > \underline{H}^s(\tau_y)$, (46') does not bind, thus the redistributive policy is implemented, and τ, H, and L_M are determined as stated above. When $Fr \leq \underline{H}^s(\tau_y)$, (46') binds and thus the uneducated cannot obtain more than A_T through policies, hence they are indifferent among any policies with redistribution satisfying $\tau \in (\tau_0^s, \tau_0^b)$, $\frac{H}{H+L_M}[w_h(\tau, H, L_M) - (1+r)e] + \frac{L_M}{H+L_M} w_l(\tau, H, L_M) = A_T$, $H \leq Fr$, and $L_M \leq 1 - H$ and any policies with $x = T = 0$, including $\tau < \tau_0^s$ or $\tau > \tau_0^b$. □

Proof of Lemma 6: Mostly, the proof of Lemma 2 can be applied with minor modifications. $\overline{\tau}_{h\downarrow}^s > \tau_0^s$ and $\overline{\tau}_{h\downarrow}^b < \overline{\tau}_h$ are from $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T < e$. Since $b_l^*(w_{l,t}) < e$, $b_h^*(w_{h,t}) \geq e$ implies that Fr_t is constant. □

Proof of Lemma 7: [Existence of the intersection] By plugging (52) into (51) and deleting $A_M {\Omega_0}^{1-\alpha} \times \Big(\frac{\tau^{\frac{\rho\theta}{\delta-\rho-\theta}}[(1-\tau)(1-\alpha)]^{1-\frac{\rho\theta}{\delta-\rho-\theta}}}{1-\alpha(1-\tau)}\Big)^{1-\alpha} \big(\frac{1-Fr}{Fr}\big)^\alpha$, $\frac{1-Fr}{Fr} = \frac{\alpha(1-\tau)}{1-\alpha(1-\tau)} \frac{1}{1-\gamma_b(1+r)}$. By plugging this equation into (51),

$$(1-\alpha)^{(1-\alpha)(1-\frac{\rho\theta}{\delta-\rho-\theta})} A_M {\Omega_0}^{1-\alpha} \left[\frac{\alpha}{1-\gamma_b(1+r)}\right]^\alpha \tau^{\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}} (1-\tau)^{1-\frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}} = \frac{e}{\gamma_b}. \tag{84}$$

τ^{*s} and τ^{*b} are solutions to this equation. For the two loci to intersect at two distinct $\tau \in (0,1)$, the LHS of the equation at $\tau = \overline{\tau}_h = \frac{(1-\alpha)\rho\theta}{\delta-\rho-\theta}$ (the rate maximizing the LHS) must be greater than the RHS, which is Assumption 3.

[$\tau^{*s} \in (\overline{\tau}_{h\downarrow}^s, \overline{\tau}_h)$ and $\tau^{*b} \in (\overline{\tau}_h, \overline{\tau}_{h\downarrow}^b)$] $\tau^{*s} < \overline{\tau}_h < \tau^{*b}$ is obvious from above. As for $\tau^{*s} > \overline{\tau}_{h\downarrow}^s$ and $\tau^{*b} < \overline{\tau}_{h\downarrow}^b$, by comparing the LHS of the equation determining τ^{*s} and τ^{*b} (eq. 84) with the one determining $\overline{\tau}_{h\downarrow}^s$ and $\overline{\tau}_{h\downarrow}^b$ (the LHS of the condition in Lemma 6 with $\overline{\tau}_h$ replaced by τ),

$$\begin{aligned}
&\big(\text{LHS of the equation for } \tau^{*s} \text{ and } \tau^{*b}\big) \gtreqless \big(\text{LHS of the equation for } \overline{\tau}_{h\downarrow}^s \text{ and } \overline{\tau}_{h\downarrow}^b\big) \\
\Leftrightarrow\ & \left[\frac{\alpha}{1-\gamma_b(1+r)}\right]^\alpha \gtreqless (1-\alpha)^{\alpha(1-\frac{\rho\theta}{\delta-\rho-\theta})} {\Omega_0}^\alpha \left(\frac{\alpha A_M}{A_T}\right)^{\frac{\alpha}{1-\alpha}} \tau^{\frac{\alpha\rho\theta}{\delta-\rho-\theta}} (1-\tau)^{\alpha(\frac{1}{1-\alpha}-\frac{\rho\theta}{\delta-\rho-\theta})} &(85)\\
\Leftrightarrow\ & \frac{A_T}{1-\gamma_b(1+r)} \gtreqless (1-\alpha)^{1-\frac{\rho\theta}{\delta-\rho-\theta}} A_M \left(\frac{\alpha A_M}{A_T}\right)^{\frac{\alpha}{1-\alpha}} \Omega_0 (1-\tau)^{\frac{1}{1-\alpha}} \left[\frac{\tau}{1-\tau}\right]^{\frac{\rho\theta}{\delta-\rho-\theta}}, &(86)
\end{aligned}$$

where the RHS is same as the LHS of the equation in Assumption 1 (with $\overline{\tau}_h$ replaced by τ). Since $\frac{\gamma_b}{1-\gamma_b(1+r)} A_T < e$, Assumption 1 implies that the RHS is greater than the LHS. That is, the LHS of the equation for $\overline{\tau}_{h\downarrow}^s$ and $\overline{\tau}_{h\downarrow}^b$ is greater than the one for τ^{*s} and τ^{*b} and $\tau^{*s} > \overline{\tau}_{h\downarrow}^s$ and thus $\tau^{*b} < \overline{\tau}_{h\downarrow}^b$. □

Proof of Lemma 8: $b_h^*(w_h) = e$ when $L_T = 0$ is not defined for $\tau < \overline{\tau}_{h\downarrow}^s$ and $\tau > \overline{\tau}_{h\downarrow}^b$, because $b_h^*(w_h) < e$ when $L_T > 0$ from Lemma 6. From Lemma 7, the locus intersects with $b_l^*(w_l) = e$ at $\tau = \tau^{*s} \in (\overline{\tau}_{h\downarrow}^s, \overline{\tau}_h)$ and at $\tau = \tau^{*b} \in (\overline{\tau}_h, \overline{\tau}_{h\downarrow}^b)$, so $b_h^*(w_h) = e$ is located to the right side of $b_l^*(w_l) = e$ for $\tau \in (\tau^{*s}, \tau^{*b})$ on the (Fr, τ) plane. Then, if $b_h^*(w_h) = e$ were effective in this tax range, $b_l^*(w_l) > b_h^*(w_h) \Leftrightarrow w_l > w_h - (1+r)e$ on $b_h^*(w_h) = e$, which cannot happen. Hence, $Fr = H^*(\tau)$ must be located between the two loci and $b_h^*(w_h) > e$ for any H when $\tau \in (\tau^{*s}, \tau^{*b})$. The rest of the result is straightforward. □

References

[1] Acemoglu, D. (2005), "Politics and economics in weak and strong states," *Journal of Monetary Economics* 52 (8), 1199-1226.

[2] Acemoglu, D. (2008), "Oligarchic vs. democratic societies," *Journal of the European Economic Association* 6 (1): 1-44.

[3] Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2005), "From education to democracy?," *American Economic Association, Papers and Proceedings*, 95 (2), 44-49.

[4] Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2008), "Income and democracy," *American Economic Review* 98 (3), 808-42.

[5] Asplund, R. (2005), "The provision and effects of company training: a brief review of the literature", *Nordic Journal of Political Economy* 31 (1), 47-73.

[6] Baum, M. and D. Lake (2003), "The political economy of growth: democracy and human capital", *American Journal of Political Science* 47 (2), 333-347.

[7] Becker, G. (1981), *A Treatise on the Family*, Cambridge, Mass.: Harvard University Press.

[8] Becker, G. and K. Murphy (1992), "The division of labor, coordination costs, and knowledge", *Quarterly Journal of Economics* 107 (4), 1137-1160.

[9] Benabou, R. (1996), "Heterogeneity, stratification, and growth: macroeconomic implications of community structure and school finance", *American Economic Review* 86 (3), 584-609.

[10] Besley, T. and T. Persson (2009), "The origins of state capacity: property rights, taxation, and politics", *American Economic Review* 99 (4), 1218-1244.

[11] Bourguignon, F. and T. Verdier (2000), "Oligarchy, democracy, inequality and growth," *Journal of Development Economics* 62 (2), 285-313.

[12] Burgess, R. and N. H. Stern (1993), "Taxation and development", *Journal of Economic Literature* 31 (2), 762-830.

[13] Davis, L. (2003), "The division of labor and the growth of government", *Journal of Economic Dynamics and Control* 27 (7), 1217-35.

[14] Deininger, K. and P. Olinto (2000), "Asset distribution, inequality, and growth", World Bank, mimeo.

[15] Doucouliagos, H. and M. A. Ulubaşoğlu (2008), "Democracy and economic growth: a meta-analysis", *American Journal of Political Science* 52 (1), 61-83.

[16] Easterly, W. (2007), "Inequality does cause underdevelopment: insights from a new instrument," *Journal of Development Economics* 84 (2), 755-776.

[17] Friedman, E., S. Johnson, D. Kaufmann, and P. Zoido-Lobatón (2000), "Dodging the grabbing hand: the determinants of unofficial activity in 69 countries", *Journal of Public Economics* 76 (3), 459-493.

[18] Kaufmann, D., A. Kraay, and P. Zoido-Lobatón (2002), "Governance matters II? Updated indicators for 2000/01", World Bank Policy Research Dept. WP No. 2772.

[19] Keefer, P. and S. Knack (1997), "Why don't poor countries catch up? A cross-national test of institutional explanation", *Economic Inquiry* 35 (3), 590-602.

[20] Galor, O. and J. Zeira (1993), "Income distribution and macroeconomics", *Review of Economic Studies* 60 (1), 35-52.

[21] Ljungqvist, L. (1993), "Economic underdevelopment: the case of missing market for human capital", *Journal of Development Economics* 40 (2), 219-239.

[22] Mo, P. H. (2001), "Corruption and economic growth", *Journal of Comparative Economics* 29 (1), 66-79.

[23] Rodrik, D., A. Subramanian, and F. Trebbi (2004), "Institutions rule: the primacy of institutions over geography and integration in economic development", *Journal of Economic Growth* 9 (2), 131-165.

[24] Rosen, S. (1983), "Specialization and human capital", *Journal of Labor Economics* 1 (1), 43-49.

[25] Schneider, F. (2005), "Shadow economies around the world: what do we really know?", *European Journal of Political Economy* 21 (3), 598-642.

[26] Tan, H. W. and G. Batra (1996), "Enterprise training in developing countries: overview of incidence, determinants, and productivity outcomes", World Bank Private Sector Development Dept. Occasional Paper No. 9.

[27] Tamura, R. (1996), "Regional economies and market integration", *Journal of Economic Dynamics and Control* 20 (5), 825-45.

[28] Yuki, K. (2006), "Roles of the extent of the market and skill specialization in development", Kyoto University, mimeo.

[29] Yuki, K. (2007), "Urbanization, informal sector, and development", *Journal of Development Economics* 84 (1), 76-103.

[30] Yuki, K. (2008), "Sectoral shift, wealth distribution, and development", *Macroeconomic Dynamics* 12 (4), 527-559.