# HUMAN CAPITAL ALLOCATION AND POLICY INTERVENTION WHEN THERE IS EXTERNALITY IN CITIES

by

Neville Nien-Heui Jiang and Rui Zhao

**Working Paper No. 03-W25**

December 2003

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

# Human Capital Allocation and Policy Intervention when there is Externality in Cities

Neville Jiang*and Rui Zhao†

January 31, 2003

**Abstract**

This paper studies the allocation of skilled and unskilled workers with different human capital levels between two locations: city and rural area. In the city activities are congregated thus there is externality in production. In rural area production is spread out and no externality exists. Given a distribution of workers with various human capital levels in an economy, the social optimal allocation gathers workers with higher human capital in each category in the city, which all competitive equilibria fail to achieve. As a result, any policies that keep workers with low human capital out of the city increase total output. We further demonstrate that in some cases it is necessary to impose direct and selective barrier on the rural-urban migration. However, such policy maintains the city premium for unskilled labor. Great incentives exist for illegal rural-urban labor flow.

*102B Calhoun Hall, Vanderbilt University, Nashville, TN. nien-huei.jiang@vanderbilt.edu

†30 Wohlers Hall, 1206 S. Sixth Street, University of Illinois, Champaign, IL 61820. ruizhao@uiuc.edu.

# 1 Introduction

Economic development and growth is almost always accompanied by increases of urban population. According to the estimates of the United Nation[?], less than 30 percent of the world population lived in the urban area in 1950, and by 2000, the proportion of the urban population rose to 47 percent. The Central and South American countries experienced the most rapid urbanization in the past 50 years. For example, the urban population in Mexico rose from 42.7 percent in 1950 to 74.4 percent in 2000, in Brazil from 36.5 percent to 81.2 percent. The flow of labor from rural areas to urban areas is the main contributor to the rapid growth of urban population. Such a large-scaled rural-urban labor migration and such a rapid urbanization as experienced by Mexico and Brazil may not be desirable. Accompanied with the rapid urbanization are many social and economic problems, such as high urban unemployment rate, urban sprawl, poverty and crime. Urban economics has for a long time dwelled on this topic.[1]

In this paper, we look at urbanization, rural-urban labor mobility, and development from a new perspective. We address this problem from the human capital allocation point of view. Unlike previous studies, in our economy, workers are heterogeneous. More specifically, workers are differentiated by the amount of human capital they have acquired. The questions we try to address are the following. Is it desirable to restrict the movement of rural labor to urban area or at least certain types of rural labor during economic development? If so, what policies can be used to achieve this

[1] The seminal work by Harris and Todaro (1970)[?] set up the framework of rural-urban migration. Bencivenga and Smith (1997)[?] studied a dynamic version of the model. Brueckner and Zenou (1999)[?] extended the model with land market. Laing, Park and Wang (2002) [?] fit a specific model for China.

goal? Under what circumstance will the direct restriction on labor mobility, such as the Household Registration System in China, be necessary?

The driving force behind the labor movement from rural areas to urban areas is the relative higher wages in cities. The higher labor productivity in cities could come from many sources. For instance, high labor productivity could be due to the fact that capital per work is higher in cities than rural areas. But this explanation again begs the question why cities are more attractive for capital, or in general production activities. It is commonly agreed that the observed congregation of productive activities in cities cannot be explained without the existence of positive externalities. This is the view we will take as well. More specifically, we concentrate on the positive externality of human capital following Lucas (1988)[**?**] . That is to say that a worker's productivity increases if the average human capital of his coworker is high. We further believe that such positive externality can only arise within cities, where workers are concentrated in a small areas thus have constant interaction with each other as modelled by Jovanovic and Rob (1989)[**?**]. Rauch (1993)[**?**] provides empirical evidence of the productivity gain from concentration of human capital in cities.

The exercise we do is as follows. We take the distribution of workers with various human capital levels within an economy as given. We then set up a model concerning resource allocation between two locations: a city and a rural area. Given that externality exists in the city, the competitive equilibrium allocation will not be socially optimal. More specifically, social planner will allocate workers with relatively high human capital to the city to take the advantage of the positive externality, while under competitive equilibrium, there is no guarantee that workers in the city will have higher human capital level than those in the rural area. In some competitive equilibria, the population in the city is larger than the social optimal level. Furthermore, under competitive equilibria, wage rates at the two locations are equalized, while under the social optimal allocation, shadow wage rates at the two locations are, in general, different. The difference between the social optimal allocation and competitive equilibria could be drastic if the distribution of human capital is bi-modal, i.e., there is a large concentration of workers with low human capital and some workers with high human capital and little in between.

The difference between the social optimal allocation and competitive equilibria warrants policy interventions. These government policies must be able to attract workers with high human capital to the city and to retain workers with low human capital in the rural area. Our analysis shows that government subsidy to the rural area financed by flat rate labor income tax increases the average human capital of workers in the city, thus increases total output. This is very effective in terms of getting closer to the social optimal allocation when the average human capital in an economy is sufficiently high and the distribution is concentrated in the middle. However, if the average human capital is rather low or if there are too many workers with extremely low human capital, the effect of such policy is limited. In these cases, taking advantage of externality in the city results in a huge difference in percentage of workers with high human capital between the city and the rural area. The extreme scarcity of workers with high human capital in the rural area raises their wage in the rural area above their wage in the city despite of the high productivity in the city. To retain workers with high human capital in the city, government needs to subsidize the urban area instead of the rural area. On the other hand, to retain the workers with low human capital in the rural area requires the government subsidizing the rural area. The solution to this conflict adopted by China is a directly restriction on mobility of worker with low human capital through the Household Registration System accompanied by subsidy to the urban area.

The rest of the paper is organized as follows. Section 2 sets up the baseline model. Section 3 characterizes the social optimal allocation. Section 4 characterizes the competitive equilibria. Section 5 provides some numerical examples. Section 6 discusses policy interventions and in particular the Household Registration System in China. Section 7 concludes.

## 2 Framework

There are continuum of measure one workers in the economy. Workers differ in terms of their human capital level. A worker of type $h$ has human capital level $h$. The measure of human capital is normalized between the interval of one and two, which means that a worker with the highest level of human capital is twice as productive as a worker with the lowest level of human capital if both of them take the same type of job. Although we have continuum types of workers, not all types of workers are perfect substitutes. Production requires two categories of labor: the skilled and the unskilled. Workers with human capital above $\bar{h}$ qualify as skilled labor and those with human capital below $\bar{h}$ are unskilled labor. Workers in the same category are perfect substitutes and workers in different categories are imperfect substitutes. In addition, skilled labor can fill the position as an unskilled one but not the other way around. The cutoff value $\bar{h}$ could be economy specific and time specific. It depends on the overall technology level of an economy at a particular time. For instance, a high school graduate in the United States 80 years ago is skilled labor. Today to qualify as skilled labor in the US one must finish college education, while in China and other developing country, a worker with high school education is still skilled labor.

Our model is static. We take the distribution of human capital in an economy as given. Function $g(h)$ and $G(h)$ denote the pdf and cdf, respectively. We assume that $G(h)$ is continues. This is a technical assumption that rules out the possibility of having a point mass on any human capital level.

There is only one good in this economy. The production of this good requires two types of labor inputs. Capital is abstract away in the baseline model. Adding capital will not change our main result as shown later. Let $L_s$ and $L_u$ denote skilled and unskilled labor, respectively. The production function $AF(L_u, L_s)$ is strictly increasing and strictly concave in both arguments and displays constant return to scale. Function $F$ also satisfies the Inada conditions. Parameter $A$ denotes the total factor productivity.

The production in this economy can take place at two locations: a city and a rural area. In the rural area, the productive activities spread out. No externality exists. The total factor productivity of a firm in the rural area is fixed and normalized to be one. The productive activities in the city, on the other hand, are congregated so there are productivity spill-overs. The productivity of any individual firm depends on the average human capital level of all workers that are allocated in the city. Let $h^a$ be the average human capital of city workers. We assume that the total factor productivity in the city is $A^c = \phi(h^a)$, where function $\phi(\cdot)$ satisfies the following assumption.

**Assumption 1** $\phi'(h) > 0$, $\phi(1) < 1$, *and there exists* $1 < h^* < 2$ *such that* $\phi(h^*) = 1$.

The above assumption implies that a concentration of production activities increases productivity only when the average human capital in the city reaches some level. This is the reduced form of a model, where negative externality due to congestion and positive externality due to productivity spill-over both exist.

Each worker is assumed to supply one unit of labor inelastically. With this assumption, a necessary condition for social optimal is output maximization. Throughout our paper, we abstract away from the consequence of income distribution to the total welfare of the economy. The social welfare is measured by the total output, not some utility based social welfare function. In another word, we concentrate on efficiency not fairness.

Let $\bar{L}_u$ and $\bar{L}_s$ denote the total efficient units of unskilled and skilled labor in an economy. Then $\bar{L}_u = \int_1^{\bar{h}} hdG$ and $\bar{L}_s = \int_{\bar{h}}^2 hdG$. Let $F_1(\cdot, \cdot)$ denote the partial derivative of $F$ with respect to its first argument and $F_2(\cdot, \cdot)$ the partial derivative with respect to the second argument. We make the following two assumptions on human capital distribution to rule out some trivial and uninteresting cases:

**Assumption 2** $F_1(\bar{L}_u, \bar{L}_s) < F_2(\bar{L}_u, \bar{L}_s)$

**Assumption 3** *There exists* $1 < \hat{h} < 2$, *such that* $\int_{\hat{h}}^2 hdG = h^*$.

The first assumption states that skilled labor are scarce resources in an economy. It is a waste to let skilled labor to do unskilled job unless there are other gains. The second assumption guarantees that, at the optimum, productions will always take place in the city.

# 3 Social Optimal Allocation

The social planner assigns each type of workers to a location and a task to maximize total output. The set of possible types of workers is $H = [1, 2]$. The set of possible locations is $I = \{c, r\}$ where $c$ stands for the city and $r$ the rural area. The set of possible tasks is $J = \{s, u\}$ where $s$ is a job that can be done only by skilled labor and $u$ a job that can be done both by skilled and unskilled labor. Let $\pi(h, i, j)$ denote the assignment rule. It specifies the probability of assigning a worker of type $h$ to location $i$ and task $j$. An admissible assignment rule $\pi : H \times I \times J \to [0, 1]$ should satisfies the following constraints:

(1) $\pi$ is piecewise continues in $h$,

(2) $\pi(h, i, s) = 0, \quad \text{for all } h < \bar{h}, \quad \text{for all } i = r, c,$

(3) $\sum_i \sum_j \pi(h, i, j) = 1, \quad \text{for all } h.$

The first constraint is a technical one. The second constraint simply states that unskilled labor cannot be assigned to do a task that requires skilled labor. The last constraint reflects the fact that $\pi(h, \cdot, \cdot)$ is a probability measure. Let $\Pi$ denote the set of all admissible allocation rules.

The social planner picks an admissible assignment rule to maximize total output.

$$\max_{\pi \in \Pi} F(L_u^r, L_s^r) + F(L_u^c, L_s^c)\phi(h^a)$$

subject to

$$L_j^i = \int_1^2 \pi(h, i, j) h dG, \text{ for all } i, j,$$
$$h^a = \frac{L_u^c + L_s^c}{\int_1^2 [\pi(h, c, u) + \pi(h, c, s)] dG},$$

where $L_u^r$ and $L_s^r$ are the efficient units of unskilled and skilled labor that are assigned to the rural area and $L_u^c$ and $L_s^c$ are those assigned to the city. The average human capital of workers assigned to the city is $h^a$.

The social planner's problem can be simplified drastically by the following observation. Let the total efficient units of unskilled and skilled labor assigned to both locations be fixed. If there is a type of worker assigned in the rural area who has higher human capital level than some workers in the same category but work in the city, swapping them increases total output. Lemma 1 builds on this intuition and shows that the optimal assignment rule is a simple rule, which can be summarized by two cutoff points. As a result, to solve the above maximization problem, without loss of generality, the social planner only needs to search over allocation rules that are simple.

**Lemma 1** *If $\pi^*$ solves the social planner's problem, then $\pi^*$ must take the following form:*

$$\pi^*(h,r,u) = \begin{cases} 1 & \text{if } 0 \leq h < h_1 \\ 0 & \text{otherwise} \end{cases}$$

$$\pi^*(h,r,s) = \begin{cases} 1 & \text{if } \max\{\bar{h}, h_1\} \leq h \leq h_2 \\ 0 & \text{otherwise.} \end{cases}$$

**Proof.** See Appendix. ■

**Lemma 2** *Let $h_1^*$ and $h_2^*$ be the optimal cutoff points. Under assumption 2, $h_1^* \leq \bar{h}$.*

**Proof.** See Appendix. ■

Lemma 2 says that if skilled labor is scarce, it is never optimal to assign skilled labor to do an unskilled task in the rural area, although it might still be optimal to assign skilled labor to do an unskilled task in the city. With the above two lemmas, to maximize output the social planner chooses two cutoff points $h_1$ and $h_2$ and a measure of skilled labor, $x$, in the city to carry out unskilled tasks.

$$\max_{h_1,h_2,x} F(L_u^r, L_s^r) + F(L_u^c, L_s^c)\phi(h^a)$$

subject to

$$L_u^r = \int_1^{h_1} hdG, \qquad L_s^r = \int_{\bar{h}}^{h_2} hdG,$$

$$L_u^c = \int_{h_1}^{\bar{h}} hdG + x, \qquad L_s^c = \int_{h_2}^{2} hdG - x,$$

$$h^a = \frac{L_u^c + L_s^c}{M^c}, \qquad M^c = \int_{h_1}^{\bar{h}} dG + \int_{h_2}^{2} dG,$$

$$1 \leq h_1 \leq \bar{h}, \qquad \bar{h} \leq h_2 \leq 2,$$

$$x \geq 0.$$

**Proposition 1** *A solution to the simplified social planner's problem exists.*

**Proof.** The problem is a maximization of a continues function over a compact set. ■

Because of the externality term, the objective function may not be globally concave. Still the optimal solution satisfies the following first order necessary conditions.[2]

$$F_1(L_u^r, L_s^r) \geq F_1(L_u^c, L_s^c)\phi(h^a) - F(L_u^c, L_s^c)\phi'(h^a)\frac{1}{M^c h_1}(h^a - h_1), \tag{1}$$

$$F_2(L_u^r, L_s^r) = F_2(L_u^c, L_s^c)\phi(h_a) - F(L_u^c, L_s^c)\phi'(h^a)\frac{1}{M^c h_2}(h^a - h_2), \tag{2}$$

$$x[F_2(L_u^c, L_s^c) - F_1(L_u^c, L_s^c)] = 0. \tag{3}$$

From the above first order conditions we can see that the social planner compares the productivity of a type of workers when they are allocated to the rural area with their productivity when they are allocated to the city plus their effect on everybody else's productivity in the city. Due to the concern of externality, the marginal products of one efficient unit of labor at the two locations differ from each other.

Let $MP_j^i$ denote the marginal productivity of one efficient unit of labor working on task $j$ at location $i$. We also refer the marginal product as shadow wage rate. Naturally,

$$MP_u^r = F_1(L_u^r, L_s^r), \qquad MP_s^r = F_2(L_u^r, L_s^r),$$

$$MP_u^c = F_1(L_u^c, L_s^c)\phi(h^a) \qquad MP_s^c = F_2(L_u^r, L_s^r)\phi(h^a)$$

Proposition 2 characterizes the marginal productivity differentials between two locations and two tasks. Most of the them come directly from the first order necessary conditions.

[2]The Inada conditions and assumption 1–3 together rule out all the corner solutions for $h_1$ and $h_2$, except $h_1 = \bar{h}$ which explains the inequality in equation (1).

**Proposition 2** *The optimal allocation satisfies the following conditions:*

1. $MP_u^c > MP_u^r$;

2. $MP_s^c > MP_s^r$ *if and only if* $h^a > h_2$;

3. $MP_s^c \geq MP_u^c$, *and the equality holds if* $x > 0$;

4. $MP_s^r > MP_u^r$;

5. $MP_s^r - MP_u^r > MP_s^c - MP_u^c$.

**Proof.** See Appendix. ■

The difference between the marginal products of the same type of labor between the city and the rural area is referred as the city premium. The difference between the marginal products of the skilled and the unskilled labor in the same location is referred as the skill premium. The proposition 2 says that the city premium of unskilled labor is always positive, while the city premium of skilled labor may be positive or negative. The unskilled labor in the city earns a premium because of their "high quality". Moreover, the skill premium in the rural area is always positive, while the skilled premium in the city could reduce to zero. Whenever the skill premium disappears, we have some skilled labor taking on unskilled tasks in the city. The skill premium in the rural area is always higher than that in the city, which means that the skilled-unskilled labor ratio in the city is always higher than that in the rural area. The city always has relative higher concentration of skill than the rural area.

For our future policy analysis it is important to distinguish the following two cases: Case A, where the marginal product of skilled labor in the city is no less than that in the rural area, and Case B, otherwise. The condition that separates Case A and Case B is whether the highest human capital level in the rural area is below or above the city average. If it is below the city average, we have Case A. In this case, the skilled labor in the city earns a premium, also because they are of "high quality". Case B occurs when the skilled labor in the rural area is very scarce compared with the city. Even though moving the worker with highest human capital in the rural area to the city increases the city average human capital level, the extreme high marginal product of the marginal type in the rural area makes the moving not worthwhile. As shown later by our computed examples, case B is more likely if the economy overall has little skilled labor or if there is a large concentration of unskilled labor with extremely low human capital.

# 4 Competitive Equilibrium

A competitive equilibrium in this economy is a set of wages $\{w_j^i\}_{j\in\{s,u\},i\in\{r,c\}}$, and an admissible allocation rule $\pi \in \Pi$, which satisfies the following requirements.

1. Given wage rate, the allocation rule maximizes the expected income for each type of workers.

2. Wage rates are determined by the marginal products of each category of labor at each location, if production takes place at that location. Otherwise, wage rates are zero at that location. More specifically, define

$$L_j^i = \int_1^2 \pi(h,i,j)hdG, \text{ for all } i,j,$$
$$h^a = \frac{L_u^c + L_s^c}{\int_1^2 [\pi(h,c,u) + \pi(h,c,s)]dG}.$$

   If $L_u^r > 0$ or $L_s^r > 0$ then $w_u^r = F_1(L_u^r, L_s^r)$ and $w_s^r = F_2(L_u^r, L_s^r)$. Otherwise $w_s^r = w_u^r = 0$. Similarly, if $L_u^c > 0$ or $L_s^c > 0$, then $w_u^c = F_1(L_u^c, L_s^c)\phi(h^a)$ and $w_u^c = F_2(L_u^r, L_s^r)\phi(h^a)$. Otherwise, $w_s^c = w_u^c = 0$.

3. If production takes place in the city, the total factor productivity in the city can not be less than 1.

The last requirement in the definition of competitive equilibrium makes sure that the city has potential advantage over the rural area so that it is meaningful to distinguish the two locations. It is obvious that a trivial competitive equilibrium always exists, where all production activities take place in the rural area. We consider this equilibrium uninteresting, thus it is omitted in our analysis. We only consider non-trivial equilibria where the city exits. We say that a competitive equilibrium is regular if production takes place both in the city and in the rural area. To have an irregular competitive equilibrium, the average human capital of the economy must exceed $h^*$.

**Proposition 3** *If a set of wages $\{w_j^i\}_{j\in\{s,u\},i\in\{r,c\}}$, and an allocation rule $\pi$ is a regular competitive equilibrium, then*

1. *there are no city premia, that is, $w_j^r = w_j^c$ for all $j$, and*

2. *the allocation rule satisfies the following restrictions:*

   (a) *the city does not have technological advantage, i.e. $\phi(h^a) = 1$,*

   (b) *the skilled-unskilled labor ratios are the same at both locations, i.e. $\frac{L_s^c}{L_u^c} = \frac{L_s^r}{L_u^r}$.*

**Proof.** The disappearance of city premia comes directly from individual maximization. It also implies that the skill premia in the city and the rural area are the same. When the production function displays constant return to scale, the equalization of skill premium dictates that the skilled-unskilled labor ratios at the two locations be the same.[3] Once the wage rates and the skilled-unskilled labor ratio are equalized across two locations, there can not be any differences in total factor productivity across two locations either, i.e., $\phi(h^a) = 1$. ■

Proposition 3 shows that at the competitive equilibrium wage rates are determined but not the allocation rule. There are continuum of allocation rules that satisfy the restrictions (a) and (b). However, no matter which competitive equilibrium we end up with, the allocation is not socially optimal. There is no city premium for unskilled labor to attract the ones with "high quality" and the city has no technological advantage over the rural area.

The indeterminacy of the competitive equilibrium also implies that we have all kinds of cities with various population sizes and compositions. We can have cities composed of workers with very high human capital together with those with very low human capital, or cities composed of workers with medium human capital mostly. But there is an upper bound on city population. We denote it by $M^*$. In addition, the competitive equilibrium with maximum city population maximizes the output among all competitive equilibria. Obviously, if the average human capital level in an economy exceeds $h^*$, the output maximizing competitive equilibrium is to move all workers into the city where the total factor productivity is above one. All the other regular competitive equilibria have all population producing with total factor productivity equal to one. If the average human capital level in an economy is less than $h^*$, then all competitive equilibria are regular and output of all equilibria is the same.

[3] From the CRS production function we can derive that the skilled and unskilled wage ratio is a monotone decreasing function of the skilled and unskilled labor ratio only, as shown in the proof of Lemma 2.

# 5 Distribution Matters

In this section we present some numerical examples to give a quantitative assessment on the difference between competitive equilibria and the social optimal allocation. The main point is that the endowment of human capital in an economy plays an important role. Not only the average level of human capital matters but also how human capital is distributed.

The specification of technology is as follows:

$$F(L_u, L_s) = [\alpha L_u^{\frac{\sigma-1}{\sigma}} + (1-\alpha) L_s^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}},$$

$$\phi(h^a) = B(h^a - 1)^{\gamma}.$$

Parameter $\alpha$ is a weighting parameter. The elasticity of substitution between two categories of labor is given by parameter $\sigma$. When $\sigma$ goes to infinity, the two categories of labor are perfect substitutes. When $\sigma$ equals zero, the two categories of labor are not substitutable at all. Parameter $\gamma$ gives a curvature to the externality term. When $\gamma$ is less than unity, the marginal contribution of average human capital is decreasing. The following values are used in our numerical examples.

| $\alpha$ | $\sigma$ | $B$ | $\gamma$ |
|---|---|---|---|
| 0.4 | 2 | 2 | 0.7 |

A worker's human capital in our analysis is matched with his education level. More specifically, an illiterate worker without any formal education has human capital level of one. The maximum education level a worker can achieve is eighteen years of formal schooling.[4] Those workers who have highest possible education attainment

[4] A typical eighteen years of formal schooling includes six years of elementary school, three years of junior high school, three years of senior high school, four years of college and two years of graduate education.

have human capital level of two. The skilled labor refers to workers with at least eleven years of formal schooling, which means that senior high school graduates or junior high school graduates with two more years of professional training are qualified. We think that this criteria is suitable for many developing countries. With this criteria, $\bar{h} = 1.61$.

Figure 1: Shape of Probability Distribution

For computational convenience, we treat years of schooling as a continuous variable. The probability density function $g$ is piecewise linear with break point at $\bar{h}$, i.e.,

$$g(h) = \begin{cases} a_1 + b_1 h & 0 \leq h \leq \bar{h} \\ a_2 + b_2 h & \bar{h} \leq h \leq 2 \end{cases} .$$

The slope parameters $b_1$ and $b_2$ define the shape of a distribution function. A positive slope parameter implies that the distribution of various types of workers within each category are more concentrated on the upper end of human capital spectrum. The overall human capital distribution in an economy has single mode, if $b_1$ is positive while $b_2$ is negative. If, instead, $b_1$ is negative, the distribution of human capital is bi-modal as shown by figure 1.

In the first set of experiments, we focus on the level of human capital endowment. The distributions of types in each category are assumed to be uniform, which means that $b_1$ and $b_2$ are both zero. Table 1 reports the percentage output gain as we move from competitive equilibria[5] to social optimal allocation, the percentage of population in the city at the optimum allocation, and the urban-rural differences in average education level, output per worker, marginal products of both skilled and unskilled labor.

As the average human capital level in an economy increases, the output difference between the social optimal allocation and competitive equilibria decreases, while the city premium of unskilled labor increases dramatically. So it is increasingly more difficult to move the economy from a competitive equilibrium to the social optimal allocation and at the same time the potential gains of any intervention declines as an economy accumulates more human capital.

Also notice that as the average human capital level in a economy increases, it is

[5] If there are more than one possible level of output, pick the highest.

Table 1: Studies on Average Human Capital Level

| Average Years of Schooling | Output Gain % | Optimal City Size | Urban-Rural Difference (Rural=100) | | | |
|---|---|---|---|---|---|---|
| | | | Schooling | Productivity | City Premium | |
| | | | | | Unskilled | Skilled |
| 6 | 12.9 | 41 | 127 | 197 | 145 | 93 |
| 7 | 15.8 | 48 | 130 | 242 | 181 | 92 |
| 8 | 11.3 | 57 | 133 | 259 | 209 | 95 |
| 9 | 8.6 | 66 | 135 | 261 | 232 | 99 |
| 10 | 4.7 | 75 | 137 | 262 | 256 | 102 |

optimal to increase city population. But when the average human capital level is low, the optimal level of urbanization is rather low. Furthermore, the optimal allocation in an economy with overall low human capital also maintains a negative city premium of skilled labor. It is only when skilled labor becomes less scarce that the city premium of skilled labor becomes positive.

In the next set of experiments, the percentage of population that are qualified as skilled labor are fixed. Different combinations of slope parameters $b_1$ and $b_2$ determine how various levels of human capital are distributed in the population. Our focus is on $b_1$, the slope parameter of unskilled labor. So the slope parameter of the skilled, $b_2$ is fixed to be -1. As shown by figure 1, a human capital distribution with positive $b_1$ has a single mode, while that with negative $b_1$ is bi-modal. Table 2 reports the same set of statistics as in Table 1, as we vary the scarcity of skilled labor and the shape of human capital distribution.

From Table 2, we find that the more of unskilled workers are concentrated at the lower end of its human capital spectrum, the larger is the difference between

Table 2: Studies on the Shape of Human Capital Distribution

| | | Optimal | Urban-Rural Difference (Rural=100) | | | |
|---|---|---|---|---|---|---|
| | Output | City | School- | Produc- | City Premium | |
| $b_1$ | Gain % | Size | ing | tivity | Unskilled | Skilled |
| Skilled Worker: 25% | | | | | | |
| 3 | 6.4 | 68 | 127 | 235 | 187 | 96 |
| 1 | 10.3 | 59 | 130 | 245 | 195 | 95 |
| -1 | 14.6 | 50 | 133 | 253 | 203 | 94 |
| -3 | 18.3 | 42 | 134 | 258 | 214 | 92 |
| Skilled Worker: 45% | | | | | | |
| 2 | 3.4 | 79 | 132 | 250 | 228 | 101 |
| 0 | 5.9 | 71 | 136 | 257 | 241 | 101 |
| -2 | 8.2 | 63 | 137 | 259 | 256 | 99 |

competitive equilibria and the social optimal allocation. To reach the social optimal allocation in an economy with bi-modal distribution, higher wage differentials between the city and the rural area have to be maintained. The above is true disregard of the population size of skilled workers. In addition, it could be optimal to have a large urban population even when the population size of skilled labor is small, as long as the distribution is single-modal. On the other hand, even when the population size of the skilled labor is high, a higher concentration of unskilled labor at the lower end of the human capital spectrum reduces optimal city population size and could tip the city premium of skilled labor to negative.

According to the census 2000,[6], 3.6% of Chinese papulation finished college education, 11.1% finished senior high school, 34.0% finished junior high school, 35.7% finished elementary school, and the rest 15.6% has little or no formal education. In the next set of experiments, we fit a distribution of human capital for China and conduct some sensitivity analysis with regards to the technology parameters. The sensitivity analysis shows that the output gain is most sensitive to the curvature of the externality. The city premium of unskilled labor is sensitive to all parameters, while the city premium of skilled labor is less so.(Table 3)

# 6 Household Registration System in China

As shown both analytically and numerically, competitive equilibria fail to reach the social optimum. This failure warrants government interventions. In this section we will discuss both indirect interventions of rural-urban migration through tax and subsidy as experimented by Mexico and direct barriers of labor mobility such as the

[6]See Appendix 2 for details.

Table 3: Sensitivity Analysis on Technology Parameters

| | | Optimal | Urban-Rural Difference (Rural=100) | | | |
|---|---|---|---|---|---|---|
| | Output | City | School- | Produc- | City Premium | |
| | Gain % | Size | ing | tivity | Unskilled | Skilled |
| Weighting parameter $\alpha$ | | | | | | |
| 0.2 | 15.7 | 46 | 127 | 403 | 240 | 97 |
| 0.4 | 13.1 | 54 | 127 | 218 | 165 | 93 |
| 0.6 | 12.3 | 56 | 127 | 176 | 146 | 89 |
| Elasticity of substitution $\sigma$ | | | | | | |
| 0.8 | 13.3 | 52 | 126 | 259 | 276 | 95 |
| 2.0 | 13.1 | 54 | 127 | 218 | 165 | 93 |
| 3.2 | 13.5 | 53 | 128 | 218 | 155 | 94 |
| Curvature parameter of externality $\gamma$[a] | | | | | | |
| 0.3 | 4.8 | 56 | 125 | 156 | 123 | 96 |
| 0.7 | 13.1 | 54 | 127 | 184 | 142 | 93 |
| 1.1 | 24.0 | 50 | 128 | 304 | 223 | 93 |

[a]Parameter $B$ is such that $\phi(1.3715) = 1$ in all cases

Household Registration System implemented in China.

The social optimum improves upon competitive equilibria by gathering workers with higher human capital in each category in the city. Through externality, the higher average human capital in the city makes workers in the city very productive. As a byproduct, the marginal products of each category of labor necessarily differ between the city and the rural area. Obviously, if government can tax or subsidize workers based on their human capital level, the social optimal allocation can be replicated exactly. However, it is not realistic for any government to run such a complicated tax system. Hence, we consider the following tax and subsidy regime: government can provide different public goods at the two different locations. All public goods are financed by flat rate income tax, with tax rate $\tau$.

Without loss of generality, we assume that there is no need to provide local public goods intrinsically. Public goods are provided to induce desirable human capital allocation only. At each location, no workers can be excluded from enjoying local public goods and everyone enjoys the same amount. Effectively, the provision of local public goods acts as a lump-sum transfer from government. The per capita provision of local public goods in the rural area is $T_r$ and that in the city is $T_c$. Under this tax and subsidy policy, an unskilled worker with human capital $h$ will stay in the rural area if and only if $(1-\tau)w_u^r h + T_r \geq (1-\tau)w_u^c h + T_c$, and a skilled worker with human capital $h$ will stay in the rural area if and only if $(1-\tau)\max\{w_s^r, w_u^r\}h + T_r \geq (1-\tau)\max\{w_s^c, w_u^c\}h + T_c$.

When the social optimal allocation is such that $w_u^c > w_u^r$ and $w_s^c > w_s^r$, the above individual optimization problem implies that only the unskilled workers with human capital more than $\frac{T_r - T_c}{(1-\tau)(w_u^c - w_u^r)}$ and the skilled workers with human capital more than $\frac{T_r - T_c}{(1-\tau)(w_s^c - w_s^r)}$ stay in the city. Consequently, if the social optimal allocation results in positive city premia of both skilled and unskilled labor, (Case A) then competitive equilibrium with government subsidy to the rural area, i.e., $T_r > 0$ and $T_c = 0$ reduces the undesirable rural-urban migration. The optimal subsidy is given by the following optimization problem:

$$\max_{T_r,\tau,x} F(L_u^r, L_s^r) + F(L_u^c, L_s^c)\phi(h^a)$$

subject to

$$\tau[F(L_u^r, L_s^r) + F(L_u^c, L_s^c)\phi(h^a)] = T_r(1 - M^c),$$

$$\begin{aligned}
h_1 &= \frac{T_r}{(1-\tau)(w_u^c - w_u^r)}, & h_2 &= \frac{T_r}{(1-\tau)(w_s^c - w_s^r)}, \\
L_u^r &= \int_1^{h_1} hdG, & L_s^r &= \int_{\bar{h}}^{h_2} hdG, \\
L_u^c &= \int_{h_1}^{\bar{h}} hdG + x, & L_s^c &= \int_{h_2}^{2} hdG - x, \\
h^a &= \frac{L_u^c + L_s^c}{M^c}, & M^c &= \int_{h_1}^{\bar{h}} dG + \int_{h_2}^{2} dG, \\
w_u^r &= F_1(L_u^r, L_s^r), & w_s^r &= F_2(L_u^r, L_s^r), \\
w_u^c &= F_1(L_u^c, L_s^c)\phi(h^a), & w_s^c &= F_2(L_u^c, L_s^c)\phi(h^a), \\
w_u^c &> w_c^r, & w_s^c &> w_s^r, \\
1 \geq h_1 &\geq \bar{h}, & \bar{h} &\geq h_2 \geq 2, \\
x &\geq 0.
\end{aligned}$$

Rural subsidy, however, is less effective, when the social optimal allocation requires that the city premium of unskilled labor remain positive but the city premium of skilled labor be negative (Case B). The rural subsidy only keeps workers with lower human capital from moving to the city, but fails to achieve a higher concentration of human capital in the city. The social optimal allocation in Case B requires that the depletion of skilled labor in the rural area be so extreme that to keep skilled labor in the city, government subsidy in the city is needed. Furthermore, the subsidy in the city has to be high enough that the workers with highest human capital weakly prefer staying in the city. But, with this level of city subsidy, every worker prefers moving to the city. Hence, additional policy instruments are needed to achieve the social optimal allocation. The additional instrument China adopted is a direct barrier to labor mobility implemented through Household Registration (Hukou) System.

The Household Registration System in China was formally implemented in January 1958. Under this system, every person in China has a Hukou, which records his/her name, age, sex, birthplace, occupation, and other basic information. However, Hukou is not merely a record keeping device, more importantly, it establishes the legal residence of its owner. Only the legal resident of an area has full access to the local education and employment opportunities, social and health insurance and any other government subsidies which are specific to that particular area. As a result, inability to change a person's Hukou is the major constraint on labor mobility in China. Especially, this constraint is binding for the majority of rural residents who wish to move to urban areas. Under the Household Registration System, a new born inherent his/her Hukou from his/her parents. The government guidelines put less restriction for a person with high education to change his/her Hukou. For example, a rural Hukou holder almost automatically acquires urban residency once he/she finishes college education. The recent adoption of temporary residential permit in major cities is employment based, and studies[7] show that a more educated rural labor is more likely to obtain long-term employment in cities. In summary, the HRS is a selective barrier of rural-urban migration, which discourages mobility of rural labor with low education level.

[7] For example, Liu and Chan [?]

Historically, urban residents are subsidized through controlled food price, better health care and education service and extended safety-nets provided by government. Today, with libration of many markets, the in-kind subsidy and price control are partially replaced by direct income subsidy. Still the better education, health care, and other public services are the major attractions of the urban area.

The heavy subsidy in the urban area together with selective barrier to rural-urban migration allow China to increase the concentration of human capital in urban areas as shown in Table 4. However, the unbalanced development policy and restriction of labor mobility increase the inequality between rural areas and urban areas and areas with high urbanization and areas with low urbanization. The increased inequality puts tremendous pressure on the enforcement of the Household Registration System. Even though the HRS is here to stay in the near future, the question is how long can it hold off the influx of rural labor into urban areas.

# 7 Conclusion and Extension

This paper intends to rationalize the direct government restrictions on rural-urban migration, especially barriers that prevent those with low human capital from moving to the city. These policies are rationalized, in our analysis, due to the human capital externality existed in cities and the scarcity of skilled labor. The resulting concentration of skilled labor in the city improves upon competitive equilibrium with free labor mobility at the cost of the increased rural-urban income inequality.

Table 4: Rural-Urban Differential in Selected Area of China

| Region | GDP per capita | Urban population % | Income difference (rural=1) | Average Education (year) | | Transfer as % of total income | |
|---|---|---|---|---|---|---|---|
| | | | | Rural | Urban | Rural | Urban |
| China | 6547 | 32.1 | 2.7 | 6.3 | 8.6 | 4.5 | 21.3 |
| Shanghai | 30805 | 79.2 | 2.0 | 6.6 | 9.3 | 4.3 | 31.0 |
| Beijing | 19846 | 66.9 | 2.2 | 7.5 | 10.2 | 5.0 | 26.7 |
| Tianjin | 15976 | 66.0 | 2.2 | 6.8 | 9.3 | 2.6 | 30.7 |
| Guangdong | 11728 | 41.6 | 2.5 | 6.4 | 8.4 | 6.3 | 13.7 |
| Liaoning | 10086 | 51.7 | 2.0 | 6.8 | 9.0 | 2.8 | 26.3 |
| Gansu | 3668 | 21.4 | 3.3 | 5.2 | 8.8 | 5.3 | 15.6 |
| Guizhou | 2475 | 22.3 | 3.6 | 4.8 | 7.6 | 5.5 | 19.7 |

*Source: China Statistical Yearbook [**?**]*

# A Proofs

## A.1 Lemma 1

Given any allocation rule $\pi \in \Pi$, there exist $h_1, h_2$, and $h_3$ such that

$$\int_1^2 \pi(h,r,u)hdG = \int_1^{h_1} hdG, \qquad \int_1^2 \pi(h,r,s)hdG = \int_{\max\{\bar{h},h_1\}}^{h_2} hdG$$

and $\displaystyle\int_1^2 \pi(h,c,s)hdG = \int_{h_3}^2 hdG$

Define a new allocation rule $\hat{\pi}$ as follows

$$\hat{\pi}(h,r,u) = \begin{cases} 1 & \text{if } 1 \le h \le h_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$\hat{\pi}(h,r,s) = \begin{cases} 1 & \text{if } \max\{\bar{h},h_1\} \le h \le h_2 \\ 0 & \text{otherwise,} \end{cases}$$

$$\hat{\pi}(h,c,s) = \begin{cases} 1 & \text{if } h_3 \le h \le 2 \\ 0 & \text{otherwise,} \end{cases}$$

$$\hat{\pi}(h,c,u) = 1 - \hat{\pi}(h,r,u) - \hat{\pi}(h,r,s) - \hat{\pi}(h,c,s).$$

We want to show that the new allocation rule generates a (weakly) larger total output than the original rule. Since, by construction, the effective units of both categories of labor at both locations are the same under the original allocation rule as those under the new allocation rule, the only channel through which the new rule can improve up the original rule is increasing the average human capital in the city, $h^a$. Based on the definition of $h^a$, it suffices to show that $\int_1^2 [\hat{\pi}(h,c,u) + \hat{\pi}(h,c,s)]dG - \int_1^2 [\pi(h,c,u) + \pi(h,c,s)]dG \le 0$.

$$\begin{aligned}
& \int_1^2 [\hat{\pi}(h,c,u) + \hat{\pi}(h,c,s)]dG - \int_1^2 [\pi(h,c,u) + \pi(h,c,s)]dG \\
= & \int_1^2 [\pi(h,r,u) + \pi(h,r,s)]dG - \int_1^2 [\hat{\pi}(h,r,u) + \hat{\pi}(h,r,s)]dG \\
= & [\int_1^2 \pi(h,r,u)dG - \int_1^{h_1} dG] + [\int_{\bar{h}}^2 \pi(h,r,s)dG - \int_{\max\{\bar{h},h_1\}}^{h_2} dG] \\
= & \{\int_{h_1}^2 \pi(h,r,u)dG - \int_1^{h_1} [1 - \pi(h,r,u)]dG\} \\
& \quad +\{\int_{h_2}^2 \pi(h,r,s)dG - \int_{\bar{h}}^{h_2} [1 - \pi(h,r,s)]dG\} \\
& or \\
= & \{\int_{h_2}^2 [\pi(h,r,u) + \pi(h,r,s)]dG - \int_1^{h_2} [1 - \pi(h,r,u) - \pi(h,r,s)]dG\}
\end{aligned}$$

In either cases the value is less than zero, which comes from the fact that

$$\begin{aligned} h_1 \int_1^{h_1} [1-\pi(h,r,u)]dG &\geq \int_1^{h_1} [1-\pi(h,r,u)]hdG \\ = \int_{h_1}^2 \pi(h,r,u)hdG &\geq h_1 \int_{h_1}^2 \pi(h,r,u)dG, \end{aligned}$$

and similarly

$$h_2 \int_{\bar{h}}^{h_2} [1-\pi(h,r,s)]dG \geq h_2 \int_{h_2}^2 \pi(h,r,s)dG,$$

and

$$h_2 \int_1^{h_2} [1-\pi(h,r,u)-\pi(h,r,s)] \geq h_2 \int_{h_2}^2 [\pi(h,r,s)+\pi(h,r,c)]dG.$$

## A.2 Lemma 2

The production function $AF(L_u, L_s)$ displays constant return to scale, then the ratio of the marginal product of skilled labor and unskilled labor is a strictly decreasing function of the skilled-unskilled labor ratio. More specifically,

$$\frac{MP_s}{MP_u} \equiv f(\frac{L_s}{L_u})$$

where

$$f(x) = \frac{F_2(1,x)}{F(1,x) - xF_2(1,x)}$$

and

$$f'(x) = \frac{F_{22}(1,x)F(1,x)}{(F(1,x) - xF_2(1,x))^2} < 0$$

Suppose that at the optimal allocation, $h_1 > \bar{h}$. Then the effective units of both categories of labor inputs in the rural area are $L_u^r = \bar{L}_u + \int_{\bar{h}}^{h_1} hdG$ and $L_s^r = \int_{h_1}^{h_2} hdG < \bar{L}_s$, respectively. By assumption 2 and property of CRS function, under this allocation rule, the marginal product of skilled labor in higher than the marginal product of unskilled labor in the rural area. Then reallocating the skilled labor in the rural area that currently carries out unskilled tasks to skilled tasks increases output. A contradiction.

### A.3 Proposition 2

Result (4) comes directly from Lemma 2. Result (3) and (2) comes directly from equation (3) and (2), respectively. To derive result (1) and (5) we need to look at two cases. If $h_1 < \bar{h}$, the equation (1) holds with equality. It is obvious that $h^a > h_1$, hence $MP_u^c > MP_u^r$. From equation (1) and (2) we have

$$(MP_s^r - MP_u^r) - (MP_s^c - MP_u^c) = F(L_u^c, L_s^c)\phi(h^a)\frac{1}{M^c}[\frac{h^a}{h_1} - \frac{h^a}{h^2}] > 0$$

If $h_1 = \bar{h}$, then $x > 0$ and $h^a > h_2$. By result (2), (3), and (4) we have $MP_u^c = MP_s^c > MP_s^r > MP_u^r$.

## B Population Statistics in China

Table 5 reports the urbanization and education attainment of Chinese population based on the five Census conducted in 1953, 1964, 1982, 1990, and 2000, respectively.[8]

[8] Source: China Population Statistics Yearbook [?]

Table 5: Urbanization and Education Attainment in China

| Census | I | II | III | IV | V |
|---|---|---|---|---|---|
| Year | 1953 | 1964 | 1982 | 1990 | 2000 |
| Total population (millions) | 594.35 | 694.58 | 1008.18 | 1133.68 | 1265.83 |
| Urban population % | 13.0 | 18.3 | 20.9 | 26.4 | 36.2 |
| Education attainment % | | | | | |
| No or less than 6 years | n.a. | 65.3 | 39.5 | 30.1 | 15.6 |
| Primary school | n.a. | 28.3 | 35.2 | 37.1 | 35.7 |
| Junior high | n.a. | 4.7 | 17.9 | 23.3 | 34.0 |
| Senior high | n.a. | 1.3 | 6.8 | 8.0 | 11.1 |
| College and above | n.a. | 0.4 | 0.6 | 1.4 | 3.6 |
| Years of formal schooling (average) | n.a. | 2.35 | 4.64 | 5.52 | 7.11 |