A Theory on the Economic Impacts of Immigration

Harashima, Taiji

Kanazawa Seiryo University

5 May 2017

A Theory on the Economic Impacts of Immigration

Taiji HARASHIMA*

May 2017

Abstract

The standard view on the economic impact of immigration has been criticized for its inability to solve the "immigration policy puzzle." It also has a problem in that the "net" income of heterogeneous workers is equalized. These problems arise because the standard view generally depends on a production function in which the elasticity of substitution between heterogeneous workers is constant. This paper constructs an alternative production function in which the elasticity of substitution between heterogeneous workers is not constant and is instead based on a model of total factor productivity. The alternative view presented based on this production function indicates that an "open door" policy is not necessarily economically optimal for host countries under some conditions.

JEL Classification code: D24, E23, E24, F22, F62, F66, F68
Keywords: Immigration; Immigration policy; Production function

* Taiji HARASHIMA, Kanazawa Seiryo University, 10-1 Goshomachi-Ushi, Kanazawa-shi, Ishikawa, 920-8620, Japan.
Email: harashim@seiryo-u.ac.jp or t-harashima@mve.biglobe.ne.jp.

1 INTRODUCTION

A number of studies on the economic impact of immigration have been conducted (Altonji and Card, 1991; Borjas, 1994, 1999, 2003; Friedberg and Hunt, 1995; Card, 2005, 2009; Bodvarsson and Van den Berg, 2009; Ottaviano and Peri, 2012). The standard view on the economic impact of immigration is that it has positive economic effects as a whole, although it has some negative effects on wages for low-skilled workers. Borjas (1999) summarized this view by stating that the "immigration surplus" (income gains that accrue to natives) is zero if immigrants have the same skill distribution as natives, but it is positive if the skills of immigrants differ from those of natives. The surplus is maximized when immigrants are either exclusively skilled or exclusively unskilled.

However, Giordani and Ruta (2011) argued that the standard view based on the standard production function (Cobb-Douglas or CES) has an important problem, which they called the "immigration policy puzzle." Because immigration surpluses are generally positive, the optimal immigration policy is an "open door" policy. That is, a country should accept as many immigrants as possible, and it is not rational to limit the entry of immigrants. Giordani and Ruta (2011) argued that this is a puzzle, however, because many countries have historically strictly restricted immigration. Therefore, there must be some unknown important elements that are not being reflected in the standard view, and because of these unknown elements, many countries may have to strictly control immigration.

The standard production function has another problem if it is combined with the assumption that workers are differentiated only by acquirable skills. If acquiring skills is endogenized, that is, if the number of skills acquired by a worker is determined at equilibrium where the "net" income (income minus costs to acquire skills) is equalized across workers, the net income of high- and low-skilled workers becomes identical. In other words, skill levels are indifferent to net income. This feature will significantly affect the results of examinations on the economic impacts of immigration and thereby affect evaluations of immigration policy because it is an essential element in examinations of immigration in the standard view.

If all workers are homogenous, the standard production function is very useful. If workers are heterogeneous (e.g., high- or low-skilled), however, it may not necessarily be useful and, in fact, may be problematic in some cases. In this paper, I present an alternative production function that is more useful when workers are heterogeneous because it does not have the nature of constant elasticity of substitution (CES) in production among heterogeneous workers, although it still does have the nature of CES in production between labor and capital. The function is constructed based on the model of total factor productivity (TFP) shown in Harashima (2009, 2011, 2012) and incorporates TFP elements that cannot be acquired at the same cost by everybody. The alternative framework for examining immigration based on this production function leads to the following conclusions: there is no guarantee that immigration surpluses will be positive if immigrants are less productive workers, and there are several important factors that support restricting the number of immigrants. Therefore, an "open door" policy is not necessarily always optimal economically for a host country.

2 PROBLEMS WITH THE STANDARD VIEW

2.1 The standard view of the economic impacts of immigration

The standard view generally relies on the standard production function and implicitly assumes that the only difference between high- and low-skilled workers is their skills, which can be acquired at the same costs equally by any worker. A typical CES production function is

$$Y = A\left[(1-\lambda-\mu)K^q + \lambda L_{HS}^q + \mu L_{LS}^q\right]^{\frac{1}{q}}, \tag{1}$$

where Y is outputs, A is technology, K is capital inputs, L_{HS} is labor inputs by high-skilled (HS) workers, L_{LS} is labor inputs by low-skilled (LS) workers, and λ, μ, and q are parameters ($0 < \mu < \lambda < 1$, $\lambda + \mu < 1$ and $q \leq 1$). The skill difference is represented by the difference between the values of λ and μ. The value of λ is larger than that of μ because HS workers are assumed to have more skills than LS workers. If $q \rightarrow 0$, the production function degenerates to a Cobb-Douglas production function. Suppose that capital moves perfectly elastically. Let w_{HS} and w_{LS} be the wages for HS and LS workers, respectively. By equation (1),

$$w_{HS} = \frac{\partial Y}{\partial L_{HS}} = \frac{r\lambda}{(1-\lambda-\mu)}\left(\frac{K}{L_{HS}}\right)^{1-q} \tag{2}$$

and

$$w_{LS} = \frac{\partial Y}{\partial L_{LS}} = \frac{r\mu}{(1-\lambda-\mu)}\left(\frac{K}{L_{LS}}\right)^{1-q} \tag{3}$$

where $r = \dfrac{\partial Y}{\partial K}$.

Borjas (1999) summarized the standard view on the effect of immigration as follows. In the case of perfectly elastic capital, the impact of immigration on wages depends on how the skill distribution of immigrants compares to that of natives. If immigrants are relatively low skilled, the wage for low-skilled workers declines and that for high-skilled workers rises, and the reverse is true if the immigrants are relatively high skilled. The immigration surplus is zero if immigrants have the same skill distribution as natives, but it is positive if immigrants differ from natives. The surplus is maximized when the immigrants are either exclusively skilled or exclusively unskilled. These predictions imply that immigration has positive economic effects as a whole because immigration surpluses are generally positive.

2.2 Problems

Giordani and Ruta (2011) presented what they called the "immigration policy puzzle" related to the standard view. Because the standard view generally indicates that the optimal immigration policy is an "open door" policy. However, in actuality, many countries have strictly restricted immigration and therefore the standard view does not seem to be fully successful in describing actual immigration phenomena.

In addition, by equations (2) and (3), $w_{HS} > 0$, $w_{LS} > 0$, and

$$\frac{w_{HS}}{w_{LS}} = \frac{\lambda}{\mu}\left(\frac{L_{LS}}{L_{HS}}\right)^{1-q}.$$

If $\left(\dfrac{\lambda}{\mu}\right)^{\frac{1}{1-q}} < \dfrac{L_{HS}}{L_{LS}}$, thereby

$$w_{HS} < w_{LS};$$

that is, the wage for HS workers is lower than the wage for LS workers. Because $0 < \mu < \lambda < 1$ and $q \leq 1$, $1 < \left(\frac{\lambda}{\mu}\right)^{\frac{1}{1-q}}$ and thereby, if L_{HS} is sufficiently larger than L_{LS}, $w_{HS} < w_{LS}$. If λ and μ are nearly equal, then $w_{HS} < w_{LS}$ even if L_{HS} is only a little larger than L_{LS}. Hence, in these cases, HS workers will want to work as LS workers because $w_{HS} < w_{LS}$ (i.e., skills that HS workers obtain do not result in higher wages). By arbitrage, there will be an equilibrium ratio $\frac{L_{HS}}{L_{LS}}$ such that $\left(\frac{\lambda}{\mu}\right)^{\frac{1}{1-q}} = \frac{L_{HS}}{L_{LS}}$, and at this equilibrium, $w_{HS} = w_{LS}$. If L_{HS} is sufficiently larger than L_{LS}, therefore, the wages of HS and LS workers are always equal. This is particularly so if λ and μ are not largely different.

Nevertheless, acquiring skills entails costs. Suppose that a worker pays back a loan that was borrowed to cover the costs to acquire skills by c_S every year. The equilibrium ratio $\frac{L_{HS}}{L_{LS}}$ will be achieved when $w_{HS} - c_S = w_{LS}$. In this way, the net incomes (i.e., income minus the costs to acquire skills) of HS and LS workers are equal, and skill level is indifferent to net income.

These problems are related to two fundamental questions. First, there is a question about the CES between HS and LS workers. Cobb-Douglas and CES production functions have the nature of CES in production, and the CES indicates that any factor input is indispensable for production. Therefore, as a factor input becomes scarcer, its value and price soar. This relationship is very reasonable between labor and capital inputs, but it may not be true between HS and LS workers because HS workers can be employed as LS workers if LS workers become scarcer. Note that labor cannot be employed as capital, but it can substitute for capital inputs. Therefore, it is doubtful that the same nature of CES exists between HS and LS workers as that between labor and capital inputs.

Second, there is a question about the assumption that the only difference between HS and LS workers is skills that can be acquired by any worker equally at the same cost, that is, that a worker can freely choose whether to be high skilled or low skilled. In this case, the ratio $\frac{L_{HS}}{L_{LS}}$ is therefore an endogenous variable. The importance of this assumption is easily understood if we instead assume that any worker is born as either a HS or LS worker and cannot change from one to the other; that is, the ratio $\frac{L_{HS}}{L_{LS}}$ is exogenously given and fixed. In this case, Cobb-Douglas and CES production functions will predict that the phenomenon $w_{HS} < w_{LS}$ would have been widely and frequently observed across countries and time periods. However, the phenomenon $w_{HS} < w_{LS}$ is regarded to be very unnatural and has actually been rarely (probably never) observed in market-oriented economies. This result indicates that there are factors that intrinsically differentiate workers' abilities other than skills that are equally acquirable at the same cost. If such factors exist, it is problematic to use the standard production function to examine heterogeneous workers.

2.3 Other important factors that differentiate workers' wages

The question arises then, is there a factor(s) that intrinsically differentiates workers' abilities? Studies on TFP imply that such a factor does exist. Neo-classical Ramsey growth models naturally predict that the TFPs of economies will eventually converge, but many endogenous growth models (e.g., those based on human capital accumulation) do not necessarily support this convergence hypothesis (e.g., Romer, 1986, 1987). Prescott (1998) has shown that arguments based on human capital (e.g., acquired skills) are unconvincing. The conclusions of empirical studies on the convergence hypothesis are mixed and inconclusive (e.g., Abramovitz, 1986; Baumol, 1986; Barro, 1991; Mankiw et al., 1992; Bernard and Durlauf, 1995; Michelacci and Zaffaroni, 2000; Cheung and Garcia-Pascual, 2004). Prescott (1998) concluded that TFPs differ across economies and time for reasons other than differences in the publicly available stock of technical knowledge and that a theory of TFP is needed to solve this problem. Harashima (2009, 2011, 2012) showed a model of TFP based on average workers' innovative intelligences. An important point of this model is that the wages of workers are differentiated not only by skills but also by workers' innovative intelligences. In addition, innovative intelligences will not be acquired easily, at least not equally by workers at the same costs. If we take the difference of innovative intelligence among workers into consideration, therefore, the standard production function may not be appropriate for analyses of economic impacts of immigration.

3 AN ALTERNATIVE FRAMEWORK FOR EXAMINING IMMIGRATION

3.1 TFP

3.1.1 A model of TFP

3.1.1.1 The experience curve effect

Harashima (2009, 2011, 2012) constructed a model of TFP based on workers' innovative intelligences and the experience curve effect (see the Appendix). The experience curve effect states that the more often a task is performed, the lower the cost of doing it. The primary idea of the experience curve effect (the "learning curve effect" in earlier literature) dates back to Wright (1936), Hirsch (1952), Alchian (1963), and Rapping (1965). Since then, the concept and models of the experience curve effect have been widely used in many fields including business management, strategy, and organizational studies. The experience curve effect is usually expressed as

$$C_N = C_1 N^{-(1-\alpha)} , \tag{4}$$

where C_1 is the cost of the first unit of output of a task, C_N is the cost of the Nth unit of output, N is the cumulative volume of output and is interpreted as experience of a worker engaging in the task, and α is a constant parameter ($0 < \alpha < 1$).

3.1.1.2 The effective factor inputs

Based on workers' innovative intelligences and the experience curve effect, the effective technology input per unit capital ($\tilde{A}$) in production is expressed by

$$\tilde{A} = \upsilon_A W_A = \omega_A \left(\frac{A}{K} \right)^{\alpha} , \tag{5}$$

where A is technology, K is capital inputs, W_A is the effective amount of technology input per unit capital when a worker uses a unique combination of varieties of technologies in $\frac{A}{K}$. In addition, υ_A and ω_A are positive constant parameters, and α is a positive parameter that is the same as that used in equation (4). ω_A indicates the degree of worker's innovative intelligence with regard to technology input (see the Appendix). Equation (5) is the same as equation (A8) in the Appendix.

Next, the effective labor input ($\tilde{L}$) in production is expressed by

$$\tilde{L} = \upsilon_L W_L = \omega_L L^{\alpha} , \tag{6}$$

where L is labor input, and W_L is the total amount of workers' effective provision of labor input that is lowered by the inefficiency resulting from fragmented and incomplete information. In addition, υ_L and ω_L are positive constant parameters. ω_L indicates the degree of worker's innovative intelligence with regard to labor input (see the Appendix). Equation (6) is the same as equation (A13) in the Appendix.

Finally, the effective capital input used by a worker on average ($\tilde{K}$) for $1 \le L$ is expressed by

$$\tilde{K} = \bar{\sigma} K , \tag{7}$$

where $\bar{\sigma}$ $\left(0 < \bar{\sigma} < 1\right)$ is a positive constant. Equation (7) is the same as equation (A16) in the Appendix. $\bar{\sigma}$ represents a worker's accessibility limit to capital with regard to location. The average value of $\bar{\sigma}$ in an economy will depend on the availability of physical transportation facilities. Location constraints, however, are not limited to physical transportation facilities. For example, law enforcement, regulations, the financial system, and other factors will also influence accessibility. The value of $\bar{\sigma}$ reflects the combined effects of all of these factors. Establishing efficient financial and other institutions (e.g., government) is very important for $\bar{\sigma}$ as is investing in physical capital (e.g., in transportation).

3.1.1.3 The approximate effective production function and TFP

By equations (5), (6), and (7), an approximate effective production function (AEPF) can be constructed such that

$$\begin{aligned} Y &= \tilde{A}\tilde{K}\tilde{L} \\ &= \omega_A \left(\frac{A}{K}\right)^{\alpha} \bar{\sigma} K \omega_L L^{\alpha} \\ &= \bar{\sigma}\omega_A \omega_L A^{\alpha} K^{1-\alpha} L^{\alpha} . \end{aligned} \tag{8}$$

Equation (8) is the same as equation (A20) in the Appendix. Let ω be $\omega_K \omega_L$. Because ω_K and ω_L are both positive, ω is positive and indicates the degree of worker's innovative intelligence (see the Appendix). AEPF is therefore

$$Y = \bar{\sigma}\omega A^{\alpha} K^{1-\alpha} L^{\alpha} ,$$

and thereby TFP (P_{TF}) is

$$P_{TF} = \bar{\sigma}\omega A^{\alpha} .$$

Higher $\bar{\sigma}$ and ω mean higher degrees of accessibility and worker's innovative intelligence, respectively, and both lead to higher levels of TFP and production.

3.1.2 Industrial area, size of the economy, and population density

3.1.2.1 Industrial area

Suppose for simplicity that people live and engage in economic activities only in "industrial areas," which are areas where economic activities are sufficiently concentrated and therefore generally excludes deserts, deep forests, mountains, and other inaccessible areas. In addition, suppose that the density of capital per unit area is identical in all industrial areas, with an upper bound of $\bar{\sigma}$. An increase in the total sum of *K* indicates an increase in the density of *K* in the industrial area; thus, the portion of *K* used by a worker also increases at the same rate as *K*. On the other hand, an increase in the total sum of *L* does not indicate any change in the density of *K* in the industrial area, and the portion of *K* used by a worker does not change.

3.1.2.2 Economy size

The economy "size" is defined in this paper as the square measure of each economy's industrial area; that is, it is not the production size but rather the spatial size. For simplicity, it is assumed that the population density in the industrialized area is identical across economies, and thereby the size of an economy is directly proportionate to its population (or the number of workers). Hence, the size of an economy indicates not only the spatial size but also the population size.

Let *S* (> 0) be the size of the economy. As shown above, *S* is defined independently of endogenous variables *Y* and *K* by an exogenous variable such as the spatial (or population) size of an economy's industrialized areas. $\tilde{A}$, $\tilde{K}$, and $\tilde{L}$ in equation (8) have to be modified to include this spatial (population) size of the economy. Suppose that *Y*, *K*, *L*, and *S* of economy *X* are Y_X, K_X, L_X, and S_X, respectively, and *A* is internationally common. The effective capital input needs to be changed from $\tilde{K}_X$ to $\frac{\tilde{K}_X}{S_X}$, the effective technology input from $\tilde{A}_X$ to $S_X^{\alpha}\tilde{A}_X$, and the effective labor input from $\tilde{L}_X$ to $S_X^{1-\alpha}\tilde{L}_X$ (see Harashima, 2009, and the Appendix). Hence, AEPF modified by the size of economy is

$$\begin{aligned} Y_X &= S_X^{\alpha}\tilde{A}_X \frac{\tilde{K}_X}{S_X} S_X^{1-\alpha}\tilde{L}_X \\ &= \tilde{A}_X \tilde{K}_X \tilde{L}_X . \end{aligned} \tag{9}$$

Equation (9) is exactly the same as equation (8), i.e., $Y = \tilde{A}\tilde{K}\tilde{L}$. This means that the spatial (population) size of economy generally does not matter in terms of the nature of AEPF, and AEPF can be applied commonly to large and small economies.[1] Note that equation (9) is the same as equation (A21) in the Appendix.

3.1.2.3 Population density

The theory behind the model of TFP in Section 3.1.1 (and see the Appendix) indicates that the population density in an industrial area will converge at a certain optimal value through competition and arbitration among firms in the long run. The reason for this nature is that a sparser population has many effects. It will lower the degree of division of labor and therefore reduce the inefficiency due to fragmentation of information, but the advantages of division of labor will also be reduced at the same time. In addition, a less dense population will increase various costs for

[1] It must be noted, however, that aggregation is still impossible as is true with other Cobb-Douglas production functions unless $\frac{K}{L}$ is identical. Although *S* does not affect the relation among *Y*, *K*, and *L*, aggregation demands an additional more restrictive constraint on the relation among *Y*, *K*, and *L* such that $Y_1 + Y_2 = f(K_1 + K_2, L_1 + L_2)$, where Y_i, K_i, and L_i indicate *Y*, *K*, and *L* for economy *i*. Here, it is not the spatial size (*S*) but the size of *Y* that matters.

firms, including costs for transportation, communication, and other utilities. Therefore, the population density will not become too low in the industrial area and will converge on a certain optimal value. Convergence to the optimal population density indicates that the size of the industrial area in an economy will become optimum in the long run through changes in the population size.

In this paper, the "long run" means the period during which the population density in an industrial area can be seen on average as being optimal and constant. The "short run" means the period during which the population density is not yet sufficiently adjusted to reach its optimal state after deviating from it. Note that, in the standard view, the "long run" often means the period during which capital stocks are being adjusted corresponding to a change in labor inputs. Hence, the meanings of the long and short runs are completely different in the standard view and in this paper.

3.2 An alternative model for examining immigration

Instead of HS and LS workers, HI and LI workers are assumed. HI and LI workers are identical to each other except for their values of ω: $\omega_{HI} > \omega_{LI}$, where ω_{HI} and ω_{LI} are ω of HI and LI workers, respectively and they are constant. ω is the same as shown in Section 3.1.1.3 and indicates the degree of worker's innovative intelligence. Hence, HI workers mean the workers with higher innovative intelligence and LI workers mean the workers with lower innovative intelligence. Note that the reason why workers have different values of ω is beyond the scope of economics and is the subject of study in other fields. In this paper, I only examine what will happen to wages and incomes when a HI or LI worker immigrates given the number of HI and LI workers already in a host country.

3.2.1 The model

Let L_{HI} and L_{LI} be the numbers of HI and LI workers in a host country, respectively. Let L_S be a unit of the size of the economy, and initially $L_S = L_{HI} + L_{LI}$. Let also $S_{HI} = \frac{L_{HI}}{L_S}$ and $S_{LI} = \frac{L_{LI}}{L_S}$; thus, initially $S_{HI} = \frac{L_{HI}}{L_{HI} + L_{LI}}$ and $S_{LI} = \frac{L_{LI}}{L_{HI} + L_{LI}}$. S_{HI} and S_{LI} can be interpreted as the sizes of the economies of HI and LI workers within the host country, respectively. $\bar{\sigma}$ is assumed to be independent of ω and constant, but this assumption is removed in Section 6.2. Capital inputs are also assumed to move perfectly elastically. Based on the model of TFP shown in Section 3.1, the production function can be described by modifying equation (8) as

$$Y = \bar{\sigma} A K^{1-\alpha} \left(\omega_{HI} L_{HI}^{\alpha} S_{HI}^{1-\alpha} + \omega_{LI} L_{LI}^{\alpha} S_{LI}^{1-\alpha} \right), \tag{10}$$

where α (> 0) is a constant parameter and the same as used in equation (4); that is, α represents the experience curve effect. Equation (10) can be interpreted such that the economy of the host country is the combined economies of HI and LI workers. $\tilde{A}$ and $\tilde{K}$ are common for both HI and LI workers' economies; therefore, there is no need to change equation (8) because of the size difference. $\tilde{L}$ is different, however, between the economies of HI and LI workers because workers are heterogeneous. Hence, the $\tilde{L}$ of the economy of HI workers and that of LI workers should be adjusted by $S_{HI}^{1-\alpha}$ and $S_{LI}^{1-\alpha}$, respectively. As a result, equation (8) is modified to equation (10).

Equation (10) indicates that as L_S is smaller, Y is larger. A smaller L_S indicates lower population density in the industrial area. However, as shown in Section 3.1.2.3, the population density will converge at a certain optimal value through competition and arbitration among firms in the long run. Let $\overline{L}_S$ be a unit of the size of the economy when the population density is optimal (thereby $\overline{L}_s$ is constant). Hence, by equation (10), the production function in the long run is described by

$$Y = \overline{\sigma} A K^{1-\alpha} \overline{L}_S^{\alpha-1} \left(\omega_{HI} L_{HI} + \omega_{LI} L_{LI} \right), \tag{11}$$

where the population density is assumed to be optimal initially; thereby, $\overline{L}_S = \overline{L}_{HI} + \overline{L}_{LI}$, where $\overline{L}_{HI}$ and $\overline{L}_{LI}$ are the initial values of L_{HI} and L_{LI}.

Equations (10) and (11) are not standard production functions (i.e., they are not Cobb-Douglas or CES production functions). Therefore, as will be shown in the following sections, they avoid the previously noted problems caused by using the standard production function. Note, however, that if workers are not heterogeneous (i.e., if $\omega_{HI} = \omega_{LI} = \overline{\omega}$), then equation (10) degenerates to

$$Y = \overline{\omega} \overline{\sigma} A K^{1-\alpha} L^{\alpha}, \tag{12}$$

which is a Cobb-Douglas production function, and thereby α indicates the labor share. This means that, in the alternative production function, the elasticity of substitution between labor and capital inputs is still constant, as it is in the Cobb-Douglas and CES production functions, but it is not constant between HI and LI workers.

3.2.2 Wages and the real interest rate in the short run

When L_{HI} or L_{LI} unexpectedly increases, the industrial area will not be able to quickly increase to return to its optimal size, and the population density will therefore increase in the short run. This means that a unit of the size increases; in this case, the value of L_S increases. Hence, equation $L_S = L_{HI} + L_{LI}$ holds even after L_{HI} or L_{LI} increases, and $L_S = L_{HI} + L_{LI} \neq \overline{L}_S$ in the short run.

Therefore,

$$\frac{\partial S_{HI}}{\partial L_{HI}} = \frac{d\left(\frac{L_{HI}}{L_S}\right)}{dL_{HI}} = \frac{d\left(\frac{L_{HI}}{L_{HI} + L_{LI}}\right)}{dL_{HI}} = \left(L_{HI} + L_{LI}\right)^{-1} - L_{HI}\left(L_{HI} + L_{LI}\right)^{-2},$$

$$\frac{\partial S_{HI}}{\partial L_{LI}} = \frac{d\left(\frac{L_{HI}}{L_S}\right)}{dL_{LI}} = \frac{d\left(\frac{L_{HI}}{L_{HI} + L_{LI}}\right)}{dL_{LI}} = -L_{HI}\left(L_{HI} + L_{LI}\right)^{-2},$$

$$\frac{\partial S_{LI}}{\partial L_{HI}} = \frac{d\left(\frac{L_{LI}}{L_S}\right)}{dL_{HI}} = \frac{d\left(\frac{L_{LI}}{L_{HI} + L_{LI}}\right)}{dL_{HI}} = -L_{LI}\left(L_{HI} + L_{LI}\right)^{-2},$$

and

$$\frac{\partial S_{LI}}{\partial L_{LI}} = \frac{d\left(\frac{L_{LI}}{L_S}\right)}{dL_{LI}} = \frac{d\left(\frac{L_{LI}}{L_{HI} + L_{LI}}\right)}{dL_{LI}} = \left(L_{HI} + L_{LI}\right)^{-1} - L_{LI}\left(L_{HI} + L_{LI}\right)^{-2}.$$

Hence, the wages for HI workers (w_{HI}) and LI workers (w_{LI}) are

$$w_{HI} = \frac{\partial Y}{\partial L_{HI}} = \bar{\sigma} A K^{1-\alpha} \left(L_{HI} + L_{LI} \right)^{\alpha - 1} \left[\omega_{HI} - (1-\alpha)(L_{HI} + L_{LI})^{-1} \left(\omega_{HI} L_{HI} + \omega_{LI} L_{LI} \right) \right] \quad (13)$$

and

$$w_{LI} = \frac{\partial Y}{\partial L_{LI}} = \bar{\sigma} A K^{1-\alpha} \left(L_{HI} + L_{LI} \right)^{\alpha - 1} \left[\omega_{LI} - (1-\alpha)(L_{HI} + L_{LI})^{-1} \left(\omega_{HI} L_{HI} + \omega_{LI} L_{LI} \right) \right]. \quad (14)$$

Hence,

$$\lim_{L_{LI} \to \infty} \frac{w_{HI}}{w_{LI}} = \frac{\omega_{HI}}{\omega_{LI}} > 1$$

and

$$\lim_{L_{LI} \to 0} \frac{w_{HI}}{w_{LI}} = 1 + \frac{\omega_{HI} - \omega_{LI}}{\omega_{LI} - (1-\alpha)\omega_{HI}} .$$

Therefore, if $\omega_{LI} > (1-\alpha)\omega_{HI}$,

$$\lim_{L_{LI} \to 0} \frac{w_{HI}}{w_{LI}} > 1 ,$$

but if $\omega_{LI} < (1-\alpha)\omega_{HI}$,

$$\lim_{L_{LI} \to 0} \frac{w_{HI}}{w_{LI}} < 1 .$$

Here, because

$$\lim_{L_{HI} \to 0} \left[(1-\alpha) \left(\frac{\omega_{HI} L_{HI} + \omega_{LI} L_{LI}}{L_{HI} + L_{LI}} \right) \right] = (1-\alpha)\omega_{LI} \quad (15)$$

and

$$\lim_{L_{HI} \to \infty} \left[(1-\alpha) \left(\frac{\omega_{HI} L_{HI} + \omega_{LI} L_{LI}}{L_{HI} + L_{LI}} \right) \right] = (1-\alpha)\omega_{HI} ; \quad (16)$$

then, by equations (13), (15), and (16),

$$w_{HI} > 0 .$$

In addition, by equations (14), (15), and (16), if S_{LI} is not very small (conversely, S_{HI} is not close to unity) or $\omega_{LI} > (1-\alpha)\omega_{HI}$, then

$$w_{LI} > 0.$$

However, if S_{LI} is very small (conversely, S_{HI} is close to unity) and $\omega_{LI} < (1-\alpha)\omega_{HI}$,

$$w_{LI} < 0.$$

That is, if S_{LI} is very small and $\omega_{LI} < (1-\alpha)\omega_{HI}$, then not only $\frac{w_{HI}}{w_{LI}} < 1$ but also $w_{LI} < 0$. Conversely, because $w_{LI} < 0$, then $\frac{w_{HI}}{w_{LI}} < 1$ for $w_{HI} > 0$.

In sum, when S_{LI} is not very small, then $w_{HI} > w_{LI} > 0$ is always true, and even when S_{LI} is very small, if ω_{LI} is not too much lower than ω_{HI} such that $\omega_{LI} > (1-\alpha)\omega_{HI}$, then $w_{HI} > w_{LI} > 0$ is also true. However, if S_{LI} is very small and if ω_{LI} is far lower than ω_{HI} such that $\omega_{LI} < (1-\alpha)\omega_{HI}$, then the anomaly $w_{HI} > 0 > w_{LI}$ will occur. This anomaly disappears in the long run (i.e., when the optimal population density returns) as will be shown in Section 3.2.3, but it will be a problem in the short run because an LI immigrant cannot enter the labor market of the host country under this condition. The outcomes of this case are examined in detail in Sections 4 and 6.

Finally, the real interest rate is

$$r = \frac{\partial Y}{\partial K} = (1-\alpha)\overline{\sigma} A K^{-\alpha} (L_{HI} + L_{LI})^{\alpha-1} (\omega_{HI} L_{HI} + \omega_{LI} L_{LI})$$

and by equation (10),

$$r = (1-\alpha) Y K^{-1}. \quad (17)$$

Note that r is kept constant because capital moves perfectly elastically across economies.

3.2.3 Wages and the real interest rate in the long run

In the long run, $L_S = \overline{L}_S = \overline{L}_{HI} + \overline{L}_{LI}$ holds, and thereby $L_S \neq L_{HI} + L_{LI}$ if L_{HI} or L_{LI} changes from the initial value, even though $L_S = L_{HI} + L_{LI}$ holds in the short run. Thereby, $\frac{dL_S}{dL_{HI}} = \frac{d\overline{L}_S}{dL_{HI}} = 0$ and $\frac{dL_S}{dL_{LI}} = \frac{d\overline{L}_S}{dL_{LI}} = 0$, and

$$\frac{\partial S_{HI}}{\partial L_{HI}} = \frac{d\left(\frac{L_{HI}}{L_S}\right)}{dL_{HI}} = \frac{d\left(\frac{L_{HI}}{\overline{L}_{HI} + \overline{L}_{LI}}\right)}{dL_{HI}} = \overline{L}^{-1},$$

$$\frac{\partial S_{HI}}{\partial L_{LI}} = \frac{d\left(\frac{L_{HI}}{L_S}\right)}{dL_{LI}} = \frac{d\left(\frac{L_{HI}}{\overline{L}_{HI} + \overline{L}_{LI}}\right)}{dL_{LI}} = 0,$$

$$\frac{\partial S_{LI}}{\partial L_{HI}} = \frac{d\left(\frac{L_{LI}}{L_S}\right)}{dL_{HI}} = \frac{d\left(\frac{L_{LI}}{\overline{L}_{HI} + \overline{L}_{LI}}\right)}{dL_{HI}} = 0,$$

and

$$\frac{\partial S_{LI}}{\partial L_{LI}} = \frac{d\left(\frac{L_{LI}}{L_S}\right)}{dL_{LI}} = \frac{d\left(\frac{L_{LI}}{\overline{L}_{HI} + \overline{L}_{LI}}\right)}{dL_{LI}} = \overline{L}^{-1}.$$

Therefore, w_{HI} and w_{LI} in the long run are

$$w_{HI} = \frac{\partial Y}{\partial L_{HI}} = \overline{\sigma}\omega_{HI} AK^{1-\alpha}\overline{L}_S^{\alpha-1} \tag{18}$$

and

$$w_{LI} = \frac{\partial Y}{\partial L_{LI}} = \overline{\sigma}\omega_{LI} AK^{1-\alpha}\overline{L}_S^{\alpha-1}. \tag{19}$$

Hence, the following inequalities are always true:

$$w_{HI} > 0,$$

$$w_{LI} > 0,$$

and

$$\frac{w_{HI}}{w_{LI}} = \frac{\omega_{HI}}{\omega_{LI}} > 1.$$

That is, the ratio of w_{HI} to w_{LI} is constant and always equal to the ratio of ω_{HI} to ω_{LI}. Thereby, in the long run,

$$w_{HI} > w_{LI} > 0.$$

Note that the actual wages may be neither the short-run or long-run wages, but rather some mix of those because the speed of convergence of the population density to its optimal value is unclear for firms and workers.

The real interest rate in the long run is

$$r = \frac{\partial Y}{\partial K} = (1-\alpha)\overline{\sigma}AK^{-\alpha}\overline{L}_S^{\alpha-1}\left(\omega_{HI}L_{HI} + \omega_{LI}L_{LI}\right), \tag{20}$$

and thereby equation (17) still holds in the long run. In addition, r is also kept constant in the long run.

3.3 Incentives for immigration

As shown in Section 3.1.1.2, $\overline{\sigma}$ is influenced not only by physical transportation facilities but also by law enforcement, regulation, the financial system, and other related factors. Because efficiencies related to these factors clearly differ among countries, it is highly likely that the value of $\bar{\sigma}$ is heterogeneous among countries.

Suppose that there are only two countries (Countries 1 and 2), and $\bar{\sigma}$ of Country 1 is higher than that of Country 2; that is, $\bar{\sigma}_1 > \bar{\sigma}_2$, where $\bar{\sigma}_i$ is the $\bar{\sigma}$ of Country *i*. Both countries are at the long-run steady state, and therefore the size unit is constant. Hence, by equations (18) and (19), wages in both countries are

$$w_{HI,1} = \bar{\sigma}_1 \omega_{HI} A \left(\frac{K_1}{L_1} \right)^{1-\alpha},$$

$$w_{HI,2} = \bar{\sigma}_2 \omega_{HI} A \left(\frac{K_2}{L_2} \right)^{1-\alpha},$$

$$w_{LI,1} = \bar{\sigma}_1 \omega_{LI} A \left(\frac{K_1}{L_1} \right)^{1-\alpha},$$

and

$$w_{LI,2} = \bar{\sigma}_2 \omega_{LI} A \left(\frac{K_2}{L_2} \right)^{1-\alpha},$$

where $w_{HI,i}$, $w_{LI,i}$, K_i, and L_i are w_{HI}, w_{LI}, K, and $(L_{HI} + L_{LI})$ of Country *i*, and *A*, ω_{HI}, and ω_{LI} are common between the two countries. Hence,

$$\frac{w_{HI,1}}{w_{HI,2}} = \frac{w_{LI,1}}{w_{LI,2}} = \frac{\bar{\sigma}_1 \left(\frac{K_1}{L_1} \right)^{1-\alpha}}{\bar{\sigma}_2 \left(\frac{K_2}{L_2} \right)^{1-\alpha}}.$$

Here, $\bar{\sigma}_1 > \bar{\sigma}_2$ and obviously $\frac{K_1}{L_1} > \frac{K_2}{L_2}$, then

$$\frac{w_{HI,1}}{w_{HI,2}} = \frac{w_{LI,1}}{w_{LI,2}} > 1.$$

That is, the wages for both HI and LI workers in Country 1 are higher than those in Country 2. Therefore, there are economic incentives for both HI and LI workers in Country 2 to immigrate to Country 1.

4 IMPACTS OF LI IMMIGRANTS

4.1 Impacts on wages and incomes in the short run

4.1.1 The usual case

The "usual case" is the case when S_{LI} is not very small or when S_{LI} is very small and $\omega_{LI} > (1-\alpha)\omega_{HI}$. As is shown in Section 3.2.2, in this case, $w_{HI} > w_{LI} > 0$.

4.1.1.1 Changes in wages

By equations (13), (14), and (20),

$$\frac{dw_{HI}}{dL_{LI}} = \frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{LI}}\left[(1-\alpha)\bar{\sigma}A\left(\frac{K}{L_{HI}+L_{LI}}\right)^{-\alpha}\frac{L_{LI}}{L_{HI}+L_{LI}}(\omega_{HI}-\omega_{LI})\right] \tag{21}$$

and

$$\frac{dw_{LI}}{dL_{LI}} = \frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{LI}}\left[(1-\alpha)\bar{\sigma}A\left(\frac{K}{L_{HI}+L_{LI}}\right)^{-\alpha}\frac{L_{HI}}{L_{HI}+L_{LI}}(\omega_{LI}-\omega_{HI})\right]. \tag{22}$$

Here, by total differential of equation (20) under the condition that $dL_{HI} = 0$,

$$\frac{dK}{dL_{LI}} = \left(\omega_{LI} - \frac{1-\alpha}{L_{HI}+L_{LI}}\right)\frac{K}{\alpha}. \tag{23}$$

By equation (23),

$$\frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{LI}} = \left(\omega_{LI} - \frac{1}{L_{HI}+L_{LI}}\right)\frac{K}{\alpha}(L_{HI}+L_{LI})^{-1} - (L_{HI}+L_{LI})^{-2}. \tag{24}$$

Therefore, if $L_{HI}+L_{LI}$ is sufficiently large, $\omega_{LI} - \frac{1}{L_{HI}+L_{LI}} > 0$ and $(L_{HI}+L_{LI})^{-2} \cong 0$; thereby, by equation (24), $\frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{LI}} > 0$.

Because $\omega_{HI} > \omega_{LI}$, by equations (21) and (22), if $L_{HI}+L_{LI}$ is sufficiently large, then

$$\frac{dw_{HI}}{dL_{LI}} > 0$$

and

$$\frac{dw_{LI}}{dL_{LI}} < 0.$$

That is, LI immigrants usually increase the wage for HI workers and decrease the wage for LI workers in the short run. Particularly, if $L_{HI} = L_{LI}$, then $\frac{dw_{HI}}{dL_{LI}} + \frac{dw_{LI}}{dL_{LI}} = 0$ by equations (21) and (22). That is, the average wage of HI and LI workers is unchanged because there are both positive and negative effects of an increase in LI workers on w_{HI} and w_{LI}, and the magnitudes of the two opposing effects are different between w_{HI} and w_{LI}. Due to an increase in LI workers, K increases to keep r constant. The increase in K increases both w_{HI} and w_{LI}. On the other hand, the theory behind the model of TFP shown in Section 3.1 (also see the Appendix) indicates that an increase in the number of workers advances the degree of division of labor, but this advance increases inefficiency due to fragmentation of information. The positive effect of the increase in K on w_{HI} exceeds the negative effect of the increase in inefficiency on w_{HI} and thereby $\frac{dw_{HI}}{dL_{LI}} > 0$, but the positive effect on w_{LI} cannot exceed the negative effect on w_{LI} and thereby $\frac{dw_{LI}}{dL_{LI}} < 0$.

4.1.1.2 Increase in the national income of natives

The national income of natives (Y_D) when LI immigrants arrive is

$$Y_D = Y - Y_{M,LI} = Y - w_{LI} L_{M,LI}$$

where $Y_{M,LI}$ is the sum of LI immigrants' incomes and $L_{M,LI}$ is the number of LI immigrants. Here, $\frac{dY}{dL_{M,LI}} = \frac{dY}{dL_{LI}}$, and it is assumed that initially $L_{M,LI} = 0$. Hence,

$$\frac{dY_D}{dL_{M,LI}} = \frac{dY}{dL_{M,LI}} - w_{LI} - \frac{dw_{LI}}{dL_{M,LI}} L_{M,LI} = \frac{dY}{dL_{LI}} - w_{LI} . \tag{25}$$

By total differential of equation (10) under the condition that $dL_{HI} = 0$,

$$\frac{dY}{dL_{LI}} = r \frac{dK}{dL_{LI}} + w_{LI} , \tag{26}$$

and by total differential of equation (17) under the condition that $dL_{HI} = 0$,

$$\frac{dY}{dL_{LI}} = \frac{r}{1-\alpha} \frac{dK}{dL_{LI}} . \tag{27}$$

By equations (26) and (27),

$$\frac{dY}{dL_{LI}} = \frac{w_{LI}}{\alpha} . \tag{28}$$

Therefore, by equations (25) and (28),

$$\frac{dY_D}{dL_{M,LI}} = \left(\frac{1-\alpha}{\alpha} \right) w_{LI} > 0 . \tag{29}$$

That is, an immigration of LI workers increases the national income of natives in the usual case. Equation (29) indicates that, unlike in the standard view, the national income of natives increases regardless of S_{HI} and S_{LI}.

4.1.2 Exceptional case

The "exceptional case" is the case when S_{LI} is very small and at the same time $\omega_{LI} < (1-\alpha)\omega_{HI}$. As Section 3.2.2 shows, in this case, $w_{HI} > 0 > w_{LI}$.

4.1.2.1 Changes in wages

Because $0 > w_{LI}$, it is impossible for an LI immigrant to enter the labor market of the host country in the short run. This means that LI immigrants have to be fully supported financially by native residents in the short run.

More generally, suppose that there are N workers in the host country and ω_i is the ω of worker i (= 1, 2, 3, …, N). In addition, if $i < j$, then $\omega_i < \omega_j$. Therefore, the wage for worker i (w_i) is

$$w_i = \bar{\sigma} A K^{1-\alpha} N^{\alpha-1}\left[\omega_i - (1-\alpha)N^{-1}\sum_{i=1}^{N}\omega_i\right]. \tag{30}$$

Let $\tilde{\varpi}$ be the average ω in the host country; that is, $\tilde{\varpi} = N^{-1}\sum_{i=1}^{N}\omega_i$. Equation (30) therefore indicates that if

$$\omega_i < (1-\alpha)\tilde{\varpi},$$

worker i cannot enter the labor market in the short run. As shown in Section 3.2.1, α means the labor share, and it is well known that the labor share is about 0.7 in many countries. Hence, if an immigrant's ω is less than 30% of $\tilde{\varpi}$, the immigrant likely will need financial support in the short run. Note that in the long run, such an immigrant can enter the labor market.

The reason for $0 > w_{LI}$ is that an increase in the number of workers increases not only output but also inefficiency caused by fragmentation of information (see the Appendix). If the ω of a worker who newly enters the network of division of labor is too low, the negative effect on w_{LI} (i.e., the increase in inefficiency due to increased fragmentation of information) exceeds the positive effect on it (i.e., the increase in outputs), and thereby $\frac{\partial Y}{\partial L_{LI}}(= w_{LI}) < 0$.

An important point is that, unlike the standard view, no HI worker will want to work as an LI worker in the exceptional case because $w_{HI} > 0 > w_{LI}$.

4.1.2.2 Increase in the national income of natives

Suppose again that initially $L_{M,LI} = 0$. Because immigrants do not enter the labor market and receive financial aid from the host country, then

$$\frac{dY}{dL_{M,LI}} = 0.$$

Suppose that each immigrant receives financial aid of βw_{LI} $(0 < \beta < 1)$ from the host country. Hence,

$$\frac{dY_D}{dL_{M,LI}} = \frac{dY}{dL_{M,LI}} - \beta w_{LI} - \frac{dw_{LI}}{dL_{M,LI}} L_{M,LI} = -\beta w_{LI} < 0 \cdot$$

In the exceptional case, an LI immigrant decreases the national income of natives by $\beta w_{LI} L_{M,LI}$ in the short run.

4.2 *Impacts on wages and incomes in the long run*

4.2.1 Changes in wages

By equations (18) and (19), in the long run,

$$\frac{dw_{HI}}{dL_{LI}} = (1-\alpha)\bar{\sigma}\omega_{HI} A K^{-\alpha} \bar{L}_S^{\alpha-1} \frac{dK}{dL_{LI}} \tag{31}$$

and

$$\frac{dw_{LI}}{dL_{LI}} = (1-\alpha)\bar{\sigma}\omega_{LI} A K^{-\alpha} \bar{L}_S^{\alpha-1} \frac{dK}{dL_{LI}} \cdot \tag{32}$$

By total differential of equation (17) under the condition that $dL_{HI} = 0$,

$$\frac{dK}{dL_{LI}} = \frac{\omega_{LI}}{\omega_{HI} L_{HI} + \omega_{LI} L_{LI}} \alpha^{-1} K > 0 \cdot \tag{33}$$

Hence, by equations (31) and (32) and inequality (33),

$$\frac{dw_{HI}}{dL_{LI}} > 0$$

and

$$\frac{dw_{LI}}{dL_{LI}} > 0 \cdot$$

That is, the immigration of LI workers increases wages for both HI and LI workers in the long run. This result is natural because $\frac{w_{HI}}{w_{LI}} = \frac{\omega_{HI}}{\omega_{LI}}$ is kept in the long run as shown in Section 3.2.3.

4.2.2 Increase in the national income of natives

Equations (25), (26), (27), and (28) still hold in the long run. Thereby, inequality (29) also holds. In the long run, therefore, the immigration of LI workers increases the national income of natives. Moreover, unlike in the standard view, this increase is irrelevant to S_{HI} and S_{LI}.

5 IMPACTS OF HI IMMIGRANTS

5.1 *Impacts on wages and incomes in the short run*

5.1.1 Changes in wages

By equations (13), (14), and (20),

$$\frac{dw_{HI}}{dL_{HI}} = \frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{HI}}\left[(1-\alpha)\bar{\sigma}A\left(\frac{K}{L_{HI}+L_{LI}}\right)^{-\alpha}\frac{L_{LI}}{L_{HI}+L_{LI}}\left(\omega_{HI}-\omega_{LI}\right)\right] \tag{34}$$

and

$$\frac{dw_{LI}}{dL_{HI}} = \frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{HI}}\left[(1-\alpha)\bar{\sigma}A\left(\frac{K}{L_{HI}+L_{LI}}\right)^{-\alpha}\frac{L_{HI}}{L_{HI}+L_{LI}}\left(\omega_{LI}-\omega_{HI}\right)\right]. \tag{35}$$

Here, by total differential of equation (20) under the condition that $dL_{LI} = 0$,

$$\frac{dK}{dL_{HI}} = \left(\omega_{HI} - \frac{1-\alpha}{L_{HI}+L_{LI}}\right)\alpha^{-1}K . \tag{36}$$

By equation (36),

$$\frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{HI}} = K\left(L_{HI}+L_{LI}\right)^{-1}\alpha^{-1}\left(\omega_{HI} - \frac{1}{L_{HI}+L_{LI}}\right).$$

Therefore, if $\omega_{HI} > \frac{1}{L_{HI}+L_{LI}}$, then $\frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{HI}} > 0$. In other words, if $L_{HI}+L_{LI}$ is sufficiently large, $\frac{d\left(\frac{K}{L_{HI}+L_{LI}}\right)}{dL_{HI}} > 0$.

Because $\omega_{HI} > \omega_{LI}$, if $L_{HI}+L_{LI}$ is sufficiently large, then by equations (34) and (35),

$$\frac{dw_{HI}}{dL_{HI}} > 0$$

and

$$\frac{dw_{LI}}{dL_{HI}} < 0 .$$

That is, the immigration of HI workers increases the wage for HI workers and decreases the wage for LI workers in the short run.

5.1.2 Increase in the national income of natives

The national income of natives (Y_D) when HI workers immigrate is

$$Y_D = Y - Y_{M,HI} = Y - w_{HI} L_{M,HI} ,$$

where $Y_{M,HI}$ is the sum of HI immigrants' incomes and $L_{M,HI}$ is the number of HI immigrants. Here, $\frac{dY}{dL_{M,HI}} = \frac{dY}{dL_{HI}}$, and it is assumed that initially $L_{M,HI} = 0$. Hence,

$$\frac{dY_D}{dL_{M,HI}} = \frac{dY}{dL_{M,HI}} - w_{HI} - \frac{dw_{HI}}{dL_{M,HI}} L_{M,HI} = \frac{dY}{dL_{HI}} - w_{HI} . \tag{37}$$

By total differential of equation (10) under the condition that $dL_{LI} = 0$,

$$\frac{dY}{dL_{HI}} = r \frac{dK}{dL_{HI}} + w_{HI} , \tag{38}$$

and by total differential of equation (17) under the condition that $dL_{LI} = 0$,

$$\frac{dY}{dL_{HI}} = \frac{r}{1-\alpha} \frac{dK}{dL_{HI}} . \tag{39}$$

By equations (38) and (39),

$$\frac{dY}{dL_{HI}} = \frac{w_{HI}}{\alpha} . \tag{40}$$

Therefore, by equations (37) and (40),

$$\frac{dY_D}{dL_{M,HI}} = \left(\frac{1-\alpha}{\alpha} \right) w_{HI} > 0 . \tag{41}$$

That is, the immigration of HI workers increases the national income of natives in the short run. In addition, unlike in the standard view, this increase is irrelevant to S_{HI} and S_{LI}.

5.2 Impacts on wages and incomes in the long run

5.2.1 Changes in wages

In the long run, by equations (18) and (19),

$$\frac{dw_{HI}}{dL_{LI}} = (1-\alpha) \bar{\sigma} \omega_{HI} A K^{-\alpha} \bar{L}_S^{\alpha-1} \frac{dK}{dL_{LI}} \tag{42}$$

and

$$\frac{dw_{LI}}{dL_{LI}} = (1-\alpha)\overline{\sigma}\omega_{LI} AK^{-\alpha}\overline{L}_S^{\alpha-1}\frac{dK}{dL_{LI}}. \quad (43)$$

By total differential of equation (17) under the condition that $dL_{HI} = 0$,

$$\frac{dK}{dL_{HI}} = \frac{\omega_{HI}}{\omega_{HI} L_{HI} + \omega_{LI} L_{LI}}\alpha^{-1}K > 0. \quad (44)$$

Hence, by equations (42) and (43) and inequality (44),

$$\frac{dw_{HI}}{dL_{HI}} > 0$$

and

$$\frac{dw_{LI}}{dL_{HI}} > 0.$$

That is, the immigration of HI workers increases wages for both HI and LI workers in the long run.

5.2.2 Increase in the national income of natives

Equations (37), (38), (39), and (40) still hold in this case. Inequality (41) therefore also holds. The immigration of HI workers increases the national income of natives in the long run as well as in the short run. Moreover, unlike in the standard view, this feature is irrelevant to S_{HI} and S_{LI}.

6 DETERRENT FACTORS

Equation (12) and the examinations in Sections 3 and 4 indicate that there are several important factors that work against taking an open door policy. For example, there are circumstances in which LI immigrants reduce the national income of natives. In addition, even if the national income of natives is not reduced, some natives may suffer other significant damages, for example, widening inequality in wages. If we take these deterrent factors into consideration, the immigration policy puzzle described by Giordani and Ruta (2011) is no longer a puzzle.

6.1 The exceptional case

In the exceptional case (i.e., S_{LI} is very small and $\omega_{LI} < (1-\alpha)\omega_{HI}$, or more generally $\omega_i < (1-\alpha)\widetilde{\omega}$), $w_{HI} > 0 > w_{LI}$ and LI immigrants cannot enter the labor market in the short run. They need financial support (financial aid of $\beta w_{LI} L_{M,LI}$) from the natives (i.e., the government of the host country). However, the burden the natives have to bear is not limited to the transfer of $\beta w_{LI} L_{M,LI}$ to LI immigrants because the negative wage ($0 > w_{LI}$) may matter not only to LI immigrants but also to native LI workers. If the long-run wage determined by equation (19) is applied to native LI workers, the wage for the native LI workers will be kept the same as before the arrival of LI immigrants. If the wage for native LI workers changes in the short run according to equation (14), however, it will also become negative, similar to that for LI immigrants. Even if the long-run wage may be applied to native LI workers, it is likely that the negative wage for LI immigrants will put huge downward pressure on the wage for native LI workers. As a result, it seems likely that the wage for the native LI workers will be reduced to some extent if LI immigrants arrive. Although the national income of natives may not be reduced according to the reduction in the wage for native LI workers, a portion of the national income will be transferred from native LI workers to native HI workers and capital owners, and inequality in incomes among the native population will increase. The increase in inequality may be interpreted as another kind of burden that the natives have to bear. If there is a public welfare system such that low-income households receive financial aid from the government, the increase in inequality will further increase the fiscal burden of the government.

In sum, in the exceptional case, the national income of natives will decrease, the wage for the native LI workers will decrease, inequality will increase, and the fiscal burden of the government will increase. Hence, if the exceptional case is actualized, it seems likely that a country would strictly limit LI immigrants.

6.2 Endogenous $\bar{\sigma}$

6.2.1 $\bar{\sigma}$ as a function of the ratio of L_{HI} to L_{LI}

As discussed in Section 3, the value of $\bar{\sigma}$ is likely heterogeneous among countries. In addition, it is likely that if people have higher innovative intelligences, more efficient financial and other institutions will be established. Therefore, if the ratio of HI to LI workers in a country is higher, the country's $\bar{\sigma}$ will be higher. Taking this relationship between $\bar{\sigma}$ and innovative intelligences into consideration, it seems likely that $\bar{\sigma}$ is a function such that

$$\bar{\sigma} = \tilde{\sigma}(R) \tag{45}$$

and

$$\frac{d\sigma}{dR} > 0, \tag{46}$$

where $R = \frac{L_{HI}}{L_{LI}}$. Because $\frac{dR}{dL_{HI}} > 0$ and $\frac{dR}{dL_{LI}} < 0$, and because of inequality (46), then

$$\frac{\partial \bar{\sigma}}{\partial L_{HI}} = \frac{d\bar{\sigma}}{dL_{HI}} = \frac{d\bar{\sigma}}{dR}\frac{dR}{dL_{HI}} > 0$$

and

$$\frac{\partial \bar{\sigma}}{\partial L_{LI}} = \frac{d\bar{\sigma}}{dL_{LI}} = \frac{d\bar{\sigma}}{dR}\frac{dR}{dL_{LI}} < 0.$$

If $\bar{\sigma}$ is an endogenous variable as indicated by equation (45),

$$w_{HI} = \frac{\partial Y}{\partial L_{HI}} = AK^{1-\alpha}\frac{d\bar{\sigma}}{dL_{HI}}\left(\omega_{HI} L_{HI}^{\alpha} S_{HI}^{1-\alpha} + \omega_{LI} L_{LI}^{\alpha} S_{LI}^{1-\alpha}\right) + \overline{w}_{HI}$$

and

$$w_{LI} = \frac{\partial Y}{\partial L_{LI}} = AK^{1-\alpha} \frac{d\overline{\sigma}}{dL_{LI}} \left(\omega_{HI} L_{HI}^{\alpha} S_{HI}^{1-\alpha} + \omega_{LI} L_{LI}^{\alpha} S_{LI}^{1-\alpha} \right) + \overline{w}_{LI},$$

where $\overline{w}_{HI}$ and $\overline{w}_{LI}$ are, respectively, w_{HI} and w_{LI} when $\overline{\sigma}$ is constant and independent of R. Because $\frac{d\overline{\sigma}}{dL_{HI}} > 0$ and $\frac{d\overline{\sigma}}{dL_{LI}} < 0$, w_{HI} is larger than $\overline{w}_{HI}$ and w_{LI} is smaller than $\overline{w}_{LI}$.

6.2.2 Impacts of LI immigrants

By total differential of equation (20),

$$w_{LI} = \alpha Y K^{-1} \frac{dK}{dL_{LI}} - \overline{\sigma}^{-1} Y \frac{d\overline{\sigma}}{dL_{LI}}. \tag{47}$$

By equations (17) and (47),

$$\frac{dK}{dL_{LI}} = \frac{w_{LI} + \frac{Y}{\overline{\sigma}} \frac{d\overline{\sigma}}{dL_{LI}}}{r} \left(\frac{1-\alpha}{\alpha} \right). \tag{48}$$

By total differential of equation (10),

$$\frac{dY}{dL_{LI}} = \frac{Y}{\overline{\sigma}} \frac{d\overline{\sigma}}{dL_{LI}} + r \frac{dK}{dL_{LI}} + w_{LI}. \tag{49}$$

By equations (48) and (49),

$$\frac{dY}{dL_{LI}} = \alpha^{-1} \left(\frac{Y}{\overline{\sigma}} \frac{d\overline{\sigma}}{dL_{LI}} + w_{LI} \right). \tag{50}$$

Therefore, by equation (50), if $\frac{Y}{\overline{\sigma}} \frac{d\overline{\sigma}}{dL_{LI}} < -w_{LI}$, then

$$\frac{dY}{dL_{LI}} < 0.$$

That is, if $\frac{d\overline{\sigma}}{dL_{LI}}$ takes a sufficiently large negative value, LI immigrants make Y decrease. This result holds both in the short and long runs. Therefore, it is possible for the national income of natives to decrease through LI worker immigration even in the long run. This result indicates that an endogenous $\overline{\sigma}$ will work as an important deterrent against a country accepting LI immigrants.

As shown in Section 3.3, Country 1 is more productive than Country 2 because of its higher $\overline{\sigma}$. Accepting LI immigrant workers makes Country 1 less productive. This means that Country 1 changes to become relatively closer to Country 2 in terms of production.

6.2.3 Impacts of HI immigrants

By total differential of equation (20),

$$w_{HI} = \alpha Y K^{-1} \frac{dK}{dL_{HI}} - \bar{\sigma}^{-1} Y \frac{d\bar{\sigma}}{dL_{HI}} . \tag{51}$$

By equations (17) and (51),

$$\frac{dK}{dL_{HI}} = \frac{w_{HI} + \frac{Y}{\bar{\sigma}} \frac{d\bar{\sigma}}{dL_{HI}}}{r} \left(\frac{1-\alpha}{\alpha} \right). \tag{52}$$

By total differential of equation (10),

$$\frac{dY}{dL_{HI}} = \frac{Y}{\bar{\sigma}} \frac{d\bar{\sigma}}{dL_{HI}} + r \frac{dK}{dL_{HI}} + w_{HI} . \tag{53}$$

By equations (52) and (53),

$$\frac{dY}{dL_{HI}} = \alpha^{-1} \left(\frac{Y}{\bar{\sigma}} \frac{d\bar{\sigma}}{dL_{HI}} + w_{HI} \right). \tag{54}$$

Because $\frac{d\bar{\sigma}}{dL_{HI}} > 0$, then by equation (54),

$$\frac{dY}{dL_{HI}} > 0 .$$

That is, if HI immigrants arrive, Y always increases. This result holds both in the short and long runs. Because $\frac{dY_D}{dL_{M,HI}} = \frac{dY}{dL_{HI}} - w_{HI} > 0$ even when $\bar{\sigma}$ is constant, HI immigrants further increase the national income of natives in the case of an endogenous $\bar{\sigma}$. Hence, unlike LI immigrants, HI immigrants will be generally welcomed in many countries.

6.3 Expandability of the industrial area

In this paper, I have assumed that the industrial area can be expanded at no cost, although the expansion may take time. In actuality, expanding the industrial area will incur some costs. In addition, the costs will differ among economies, particularly depending on the scarcity of land where economic activities can be sufficiently concentrated. The costs for expansion will be higher in a country where appropriate land is scarcer. For example, the costs will be higher in a small country with a high population density than they will be in a large country with a low population density.

If there are large differences in the costs of expanding the industrial area among economies, the lengths of the short and long runs will differ among them. In a small country with a very high population density, the short run may be very long because of the very high costs to expand its industrial area. Similarly, the length may be shorter in a large country. Therefore, in a small country with a very high population density, the short-run effects of immigrants may persist for a very long period. Because LI immigrants will place significant downward pressures on wages for native LI workers in the "short" run, the significant downward pressures will prolong for a very "long" period. Therefore, this expandability factor may work as a very important deterrent against immigration in small and densely populated countries. Conversely, people in large countries with lower population densities may be more open to LI immigrants.

6.4 Impact on the rate of time preference

Many empirical studies indicate that the rate of time preference (RTP) is negatively correlated with income (e.g., Lawrance, 1991; Samwick, 1998; Ventura, 2003). Therefore, it is likely that the RTP of the representative household (RTP RH) is negatively correlated with R (see Harashima, 2012). Let θ be RTP RH, and θ is a function of R such that

$$\theta = \tilde{\theta}(R)$$

and

$$\frac{d\theta}{dR} < 0 .$$

Hence,

$$\frac{d\theta}{dL_{HI}} < 0 \tag{55}$$

and

$$\frac{d\theta}{dL_{LI}} > 0 . \tag{56}$$

An increase in θ decreases output, consumption, and capital at equilibrium because, as is well known, $\theta = r$ at equilibrium. Inequality (56) therefore indicates that LI immigrants will result in lower levels of output, consumption, and capital in the long run by raising θ. This factor may also work as an important deterrent against LI immigrants.

On the other hand, inequality (55) indicates that HI immigrants will result in higher levels of output, consumption, and capital, and thereby HI immigrants will be generally welcomed in many countries, similar to the case of endogenous $\bar{\sigma}$.

7 CONCLUDING REMARKS

The standard view on the economic impact of immigration is that immigration has positive economic effects as a whole, although it has some negative effects on wages for low-skilled workers. However, Giordani and Ruta (2011) argued that the standard view has an important problem (the "immigration policy puzzle"). In addition, the standard production function has another problem when it is used to examine immigration because skills are indifferent to net income. An alternative framework for examining immigration that does not rely on the standard production function is needed.

I present an alternative production function that will be more useful when workers are heterogeneous. It is constructed based on the model of TFP presented in Harashima (2009, 2011, 2012). The alternative framework based on this production function shows that there is no guarantee that immigration surpluses are positive if immigrants are less productive workers and there are several important deterrent factors against accepting less productive immigrants. These results indicate that an “open door” policy is not necessarily economically optimal for all host countries.

APPENDIX

A1 Workers' innovations

A1.1 Non-accumulative innovation

A1.1.1 Innovations need not be intrinsically accumulative

Innovations are usually considered to be intrinsically accumulative, and TFP reflects the total sum of innovations that have been created and accumulated in the long history of human beings. However, accumulativeness is not a necessary condition for innovation because, as discussed in the introduction, its core meaning is the act of introducing something new or the thing itself that has been newly introduced. Luecke and Katz (2003) argue that innovation is generally understood as the introduction of a new thing or method and the embodiment, combination, or synthesis of knowledge in original, relevant, valued new products, processes, or services. The essence of innovation is therefore not accumulativeness but newness.

Nevertheless, non-accumulative innovations have drawn little or no attention in economics because innovations that are not accumulated have been regarded as being without value from an economic point of view. Accumulated innovations are often thought of as knowledge or technology, and they are usually regarded as equivalent to TFP. An innovation that is not accumulated is not included as knowledge, technology, or TFP because these must be commonly accessible and non-accumulative innovations are not. From this perspective, non-accumulated innovations are considered to have no effect on production and therefore be meaningless. The neglect of non-accumulative innovation may also be partially attributed to the belief that innovations must be accumulated because they have the innate nature of spillover (i.e., transfer), which implies accumulation. If an innovation makes someone better off, rational people have incentive to obtain and utilize it; thus, the innovation spills over. To spill over, the innovation must be recorded and transferrable in advance, that is, accumulated as a common piece of knowledge or technology. Conversely, innovations must be accumulated if they are consistent with the incentives of rational people.

However, the above rationales do not necessarily hold, for the following reason. A non-accumulative innovation is without value to people who did not create it, and the above rationales are convincing if only those people are considered. There is, however, no *a priori* reason that a non-accumulative innovation is valueless to the person who created it because that person can utilize it personally for production even if others cannot. Therefore, even if an innovation is not accumulated and does not become common knowledge, it still can contribute to production. A non-accumulative innovation may even be an important production element for the person who created it. In addition, if the costs to acquire an innovation created by other persons are higher than its benefits, the innovation will not spill over. Therefore, the concept that some innovations do not spill over and are not accumulated is not inconsistent with rational people's incentives for using innovations. Clearly the accumulativeness of innovation is not a simple issue and requires more careful consideration.

A1.1.2 Innovations that are not accumulated

Innovations will be used personally even if they are not recognized and recorded. In addition, some innovations may be deliberately kept personal. Hence, an innovation will not be accumulated if nobody is aware of the innovation's novelty, nobody records or reports the innovation, or the person who created the innovation keeps it secret. The above conditions will be satisfied in the following situations. An innovation will not be recognized or recorded if the innovation is minor or if the innovation can be applied only to an unrepeatable incident. In addition, an incentive to keep an innovation secret will be strong if the person who creates the innovation cannot gain enough benefits by making it public. Thus an innovation will not be recorded if the costs of making the innovation public are higher than its expected benefits.

A1.1.2.1 Minor innovations

A person who creates an innovation may be unaware of having created it if its contribution to improving productivity is minor. The person may also notice the increased productivity but not seek to identify the reason for the improvement because such an investigation may seem too costly. Finally, even if the mechanism of the innovation is noticed and specified, the person who created it may not record it if it is deemed to be minor. It is therefore clearly possible that minor innovations are not noticed, identified, or recorded.

Even if an innovation is unnoticed or unrecorded, it still can be used for production by the person who created it, whether consciously or unconsciously, while the person continues doing that job. Unnoticed innovations will vanish when that person quits doing the job. If innovations are recognized but unrecorded, it is possible that at least some of them could be handed down to other workers. Because these are isolated and "personal" occurrences within a small closed group, they would not constitute a piece of accumulated knowledge common to all human beings.

A1.1.2.2 Innovations for unrepeatable incidents

Even if an innovation is not minor, it will not be recorded if it can be applied only to an unrepeatable situation. For example, a negotiation between a seller and a buyer will be basically unrepeatable. Similar negotiations may occur, but an identical one will not. There are also incidents that occur, for example, only on a specific machine installed at a particular location; these incidents are never reproduced at other machines installed at other locations. This type of isolated and non-reproducible incident can be interpreted as unrepeatable in a broad sense. In addition to these spatially unrepeatable incidents, each machine has unique characteristics even if it was designed to be exactly the same as other machines. There will not be sufficient incentive to record or widely disseminate an innovation that can be applied only to an unrepeatable situation or to a machine with unique characteristics.

A1.1.2.3 Costs of disseminating and acquiring information

There will be a strong incentive to keep an innovation secret if the innovation spills over freely without compensation to the innovator. However, even if a patent could be taken out to obtain appropriate compensation, the incentive to keep the innovation secret will still be strong if the cost of dissemination exceeds expected revenues. If an innovation was created for a minor incident, benefits gained from the innovation will usually be smaller than the cost of dissemination, and the incentive to keep the innovation personal will be strong. The costs for making an innovation public can be classified into two types: dissemination costs and acquisition costs. Dissemination costs are the costs paid to make an innovation public and to disseminate it, for example, patent application fees, advertising costs, marketing costs, and similar expenditures. Acquisition costs are the costs paid to acquire and utilize an innovation that some other person created, for example, search costs, transportation costs, and training costs. Patent royalties are included in acquisition costs only if the market value of the innovation exceeds the royalty plus other acquisition costs. Generally, dissemination costs are likely to be larger than acquisition costs, excluding patent royalties.

Let δ indicate dissemination costs, η indicate acquisition costs, and π indicate the market value of an innovation. As argued above, in general $\delta > \eta$ if $\delta > \pi$; therefore innovations are categorized into the following three ranges depending on the relative value of π compared with those of δ and η:

Range I: $\pi \geq \eta \geq \delta$ or $\pi \geq \delta \geq \eta$; patented accumulative innovations

Range II: $\delta > \pi \geq \eta$; uncompensated spillovers of accumulative innovations

Range III: $\delta > \eta > \pi$; non-accumulative innovations

If the market value of an innovation exceeds its dissemination and acquisition costs, the patent of the innovation will be sold and disseminated widely (Range I). If the market value of an innovation does not exceed its dissemination costs but exceeds its acquisition costs, the innovation will disseminate widely without compensation (i.e., uncompensated spillover; Range II). If the market value of an innovation does not exceed either cost, the innovation will not be disseminated and will be kept personal (i.e., non-accumulative innovation; Range III). Because it is highly likely that the number of minor innovations is far larger than the number of innovations that have high market values, the shape of innovation distribution slopes downward and to the right, and the distribution will have a long tail. This shape can be approximated simply by an exponential or Pareto distribution, but it is not necessary to assume a specific functional form of distribution. The important point is not the specific functional form of the distribution but its properties—if $\delta > \eta > \pi$, then non-accumulative innovations exist and there will be far more of them than of accumulative innovations.

A1.2 The origin of non-accumulative innovation

It seems clear that non-accumulative innovations exist, but who creates them? Researchers can certainly create them, but so can ordinary workers. Usually, workers are implicitly assumed to do only what they are ordered to do and nothing else. Workers in this sense can be substituted for capital. If the cost of using capital is lower than that of using workers, capital inputs will be chosen rather than labor inputs. Generally, such robot-like workers have been assumed as the labor input in typical production functions. Of course, workers are not robots. They are human beings that are fundamentally different from machines—only humans can fix unexpected problems by creating innovations.

A1.2.1 Unexpected problems require innovation

Actions taken to deal with expected incidents are determined by calculating the solutions to optimization problems that are built based on models constructed in advance. These calculations can be implemented by machines given a specific objective function, structural equations, parameter values, and necessary environmental information. However, this is not true if actions taken to deal with unexpected problems are required, because the models constructed in advance are guaranteed to be useful only for expected incidents, and they are not necessarily guaranteed to be applicable to unexpected incidents. When an unexpected problem occurs, workers in charge of the production first have to grasp the situation and then prioritize their actions. During these actions, the workers conduct two types of important intellectual activities: (1) discover unknown mechanisms that prevail in the surrounding environment and (2) invent new ways to manage the environment. That the problem is unexpected indicates that correct mechanisms for this particular situation are not known and need to be discovered, and on the basis of the newly discovered mechanisms, the structural equations and parameters in the model used for the plan of action should be revised. The revised model may indicate that there is no solution to resume efficient production, and new ways of managing the environment should be invented. Discovery and invention commonly involve the creation of something new, that is, innovation.

Machines deal with programmed tasks quite well, often much better than human beings. Conversely, machines cannot deal with non-programmed tasks. The performance of machines declines and often they stop working if unexpected problems occur because the machines do not have a program to deal with unexpected problems. When encountering unexpected problems, machines will immediately reach a dead end. They cannot solve unexpected problems by simply applying their pre-programmed optimization algorithms, and they cannot rewrite these algorithms to make them applicable to unexpected incidents. The revision or creation of models in the face of unexpected incidents can be implemented only by human beings.

A1.2.2 Workers' innovations to fix unexpected minor problems

Is it either necessary or expected to utilize workers' innovations for production? If workers are assumed to be robot-like beings, their abilities to solve unexpected problems will not be considered as part of production. However, it would be irrational for firms not to utilize workers' innovative abilities if the firms know that workers possess these abilities. An ordinary worker's ability to solve unexpected problems may be lower than that of educated and trained researchers, but the abilities of the former should be utilized fully for a firm to be rational. If anything, the workers' abilities to fix unexpected problems appear indispensable in production processes because many minor but unforeseeable incidents actually occur. It would be quite inefficient if a team of specialized highly educated and trained employees dealt with all unexpected incidents, no matter how minor, and workers had to wait for the team to arrive at the locations where a minor unexpected incident happened. If, however, an unexpected but minor problem is fixed by a worker at the location where the problem occurred, production can proceed more efficiently and smoothly. The well-known "Kaizen" method in Japanese manufacturing companies may be a way to more completely exploit such opportunities (e.g., Lee et al., 1999). Besides innovations by suppliers, "user innovation" by consumers and end users has drawn attention recently (e.g., Baldwin et al., 2006). It is quite reasonable and rational for firms to fully exploit any opportunity to improve productivity whether its source is an innovation created by a researcher, ordinary worker, or user.

Finally, a worker's ability to fix unexpected problems may seem to be part of the set of the worker's learned skills or techniques, but that ability is fundamentally different from learned skills or techniques because learning skills and techniques and creating skills and techniques are completely different activities.

A1.3 Imperfections make workers' innovations indispensable

Although it is rational for employers to fully exploit workers' innovations, in this section, I explain why workers' innovations are truly an indispensable element in production.

A1.3.1 Imperfect accumulated innovations

The current state of accumulated innovations is far from perfect, and, moreover, it always will be. Human beings will never know everything about the universe. Although we may be able to fully utilize known information, we still face many unexpected problems because the knowledge and technology we currently possess is imperfect. If accumulated innovations were perfect, machines that embody them would always work well in any situation. However, the accumulated innovations are not perfect, and thus machines malfunction occasionally or face other unexpected incidents. As stated previously, it is very efficient if workers' innovations are utilized to fix these minor but unexpected troubles. Imperfection of accumulated innovations therefore necessitates workers' innovations.

A1.3.2 Incomplete information caused by the division of labor

Labor input has the property of diminishing marginal product, which is usually explained by congestion or redundancy. However, this explanation is not necessarily convincing. The inefficiency caused by congestion or redundancy can be removed by division of labor. If labor is sufficiently divided, there will be no congestion or redundancy, and the labor input will not exhibit diminishing marginal product. This suggests that division of labor cannot remove all inefficiencies with regard to labor input. With division of labor, each worker experiences only a fraction of the whole production process. These divided and isolated workers can access only a fraction of information on the whole production process. It is also difficult for a worker to know information that many other workers at different production sites accessed. Because all of the labor inputs are correlated owing to division of labor, this feature of fragmented information is especially problematic when workers engage in intellectual activities. Correlation of the entire labor input indicates that all pieces of information on the whole production process need to be completely known to each worker to enable correct decision making. However, only a portion of the information on the whole production process is available to each worker; that is, each individual worker has incomplete information. When an unexpected problem occurs, workers with fragmented and incomplete information will make different, usually worse, decisions than those with complete information. As a result, overall productivity decreases.

For example, a CEO of a large company may know the overall plan of production but not the local and minor individual incidents that happen at each production site each day. In contrast, each worker at each production site may know little of the overall plan but a great deal about local and minor individual incidents that occur for each specific task each worker engages in at each production site. To be most efficient, even if many unexpected incidents happen, all of the workers and the CEO need to know all of the information on the entire process because all of the labor inputs are correlated owing to division of labor. However, it is nearly impossible for each worker to access all of the experiences of every other worker. Division of labor therefore leads to information fragmentation and obstructs any person from knowing all the information about the entire production process.

Each worker therefore must use incomplete information when encountering unexpected problems. Conjecturing the full detailed structure of the whole production process is an intellectual activity to discover unknown mechanisms. If a worker can discover more correct mechanisms even in the absence of complete information, the inefficiency is mitigated. However, completely mitigating the inefficiency will be impossible, and decisions based on less information will deviate from those made with full information. Sometimes actions that are relatively less urgent or important will be given priority, and efficiency will decline. As the division of labor increases, workers are less able to correctly estimate the full structure of the whole production process and less able to correctly prioritize actions to solve unexpected problems.

Division of labor cannot simultaneously solve inefficiency caused by congestion or redundancy and that caused by fragmented and incomplete information. Although a greater division of labor removes the former, it generates the latter. Inefficiency resulting from congestion and redundancy is probably much more serious than that caused by information fragmentation, and labor is divided almost completely despite the fact that information fragmentation harms productivity.

A1.3.3 Indispensable and economically important workers' innovation

Even if workers can innovate to fix unexpected minor troubles, the question remains whether these innovations are important economically. In general, most non-accumulative innovations are minor, which suggests that they may not be economically important. However, as discussed in Section A1.1, there will be far more minor innovations than major innovations. There are also usually far more ordinary workers than researchers and other highly trained or educated employees. In addition, the distributions of innovations for researchers and other highly trained employees and for ordinary workers are certainly different. Ordinary workers are likely to have a limited contribution to accumulative innovations (i.e., Ranges I and II) as compared to that of researchers and other highly trained employees, but the former will have a much larger contribution to non-accumulative innovations (Range III). As previously discussed, non-accumulative innovations are indispensable for production at each production site because of imperfect accumulative innovations and fragmented and incomplete information. Without worker-created non-accumulative innovations, the efficiency of production will decline considerably. This indispensability indicates that workers' innovations are economically important. The economic importance of workers' innovations is further examined in Section A3.

A2 The experience curve effect

A2.1 The experience curve effect and workers' innovations

Workers' innovations are indispensable, but how are they created? The experience curve effect gives a clue to this mechanism.

A2.1.1 The theory of the experience curve effect

The experience curve effect states that the more often a task is performed, the lower the cost of doing it. Workers who perform repetitive tasks exhibit an improvement in performance as the task is repeated a number of times. The primary idea of the experience curve effect (the "learning curve effect" in earlier literature) dates back to Wright (1936), Hirsch (1952), Alchian (1963), and Rapping (1965). The experience (or learning) curve effect has been applied in many fields, including business management, strategy, and organization studies.

The experience curve effect is usually expressed by the following functional form:

$$C_N = C_1 N^{-(1-\alpha)} \qquad \text{(A1)}$$

where C_1 is the cost of the first unit of output of a task, C_N is the cost of the *N*th unit of output, *N* is the cumulative volume of output and interpreted as experience of a worker engaging in the task, and α is a constant parameter ($0 < \alpha < 1$). $\frac{C_{2N}}{C_N}$ and $1-\alpha$ are often called the progress ratio and learning rate, respectively. This log-linear functional form is most commonly used probably because of its simplicity and good fit to data. Empirical studies have shown that α is usually between 0.6 and 0.9. Studies by Boston Consulting Group (BCG) in the 1970s showed that experience curve effects for various industries range from 10 to 25% cost reductions for every doubling of output (i.e., $0.58 \le \alpha \le 0.85$) (e.g., BCG, 1972). Dutton and Thomas (1984) present the distribution of progress ratios obtained from a sample of 108 manufacturing firms. The ratios mostly range from 0.7 to 0.9 (i.e., $0.48 \le \alpha \le 0.85$) and average 0.82 (i.e., $\alpha = 0.71$). OECD/IEA (2000) argues that industry-level progress ratios have a similar distribution as the firm-level ones shown in Dutton and Thomas (1984; see also, e.g., Hirsch, 1956; Womer and Patterson, 1983; Womer, 1984; Ayres and Martinas, 1992; Williams and Terzian, 1993).

The magnitude of α (or equivalently the progress ratio or learning rate) may be affected by various factors (e.g., Hirsch, 1956; Adler and Clark, 1991; Pisano et al., 2001; Argote et al., 2003; Sorenson, 2003; Wiersma, 2007). Nevertheless, the average α is usually observed to be almost 0.7 (i.e., a progress ratio of 0.8 and a learning rate of 0.3) as shown in BCG (1972), Dutton and Thomas (1984), and OECD/IEA (2000). It therefore seems reasonable to assume that α is 0.7 on average.

A2.1.2 Information conveyed by experience

An important element that an experience conveys is information. By accumulating experiences of doing a task, a worker increases the amount of information known about the task and makes it more complete. In this sense, *N,* which indicates experience in equation (A1), reflects the current amount of information a worker possesses about a task. Accumulated experiences will improve efficiency in implementing a task because the amount of information on the task increases. However, if other factors remain the same, the magnitude of improvement will diminish as *N* accumulates because the information on the task will approach saturation.

Let *I* be a set of the currently available maximum information on a task. Engaging in the task in a unit of period provides a subset of *I* to a worker. Engaging in more units of period (i.e., accumulating experience *N*) makes the information on the task the worker currently possesses ($\tilde{I}$) approach *I* (i.e., the difference between $\tilde{I}$ and *I* diminishes). A part of the subset of *I* the worker acquires in a unit of period will overlap the part of the subset of *I* the worker acquires in the next period. With more complete information, accordingly, efficiency will improve. Because $\tilde{I} \to I$ as $N \to \infty$, then the magnitude of improvement will asymptotically decrease as *N* increases. Nevertheless, this asymptotical decrease may not be a simple process. Some piece of information may be easily obtainable and some other piece may not be, and some portion of information may have a relatively large impact on efficiency and other portions have small effects. The functional form that describes the asymptotical decrease of the magnitude of improvement will depend on interaction between these effects. The log-linear functional form $C_N = C_1 N^{-(1-\alpha)}$ fits empirical data well and is simple, and thus it has been used mostly for the experience curve effect.

A2.1.3 Extending the concept of the experience curve effect

Because the essence of experience is that it conveys information, the experience curve effect can be extended to a wide variety of tasks. The tasks need not be limited to a worker's repeated actions, that is, tasks whose experiences are divided by periods. For example, consider that a human activity can be divided into many experiences, each of which is obtained by different workers. Each experience conveys a subset of information, and a part of the subset overlaps with subsets regarding other experiences. The experience curve effect will be applicable to this kind of task by interpreting *N* as a subset of all worker experiences, so a task in a period whose experiences are divided by workers will be also applicable to the experience curve effect in the same way that a task performed by a worker whose experiences are divided by periods is. Extending this logic suggests that tasks applied to the experience curve effect should not be limited to the ones whose experiences are divided only by periods or workers. As long as the task is a human intellectual activity and its experiences are divided by factors other than periods or workers, the task will also be applicable to the experience curve effect because it has the common nature that each divided experience conveys only a subset of all the information that affects the worker's intellectual activities. Nevertheless, the concept of the experience curve effect should not be expanded infinitely. It can be applied only to the tasks of workers, the performances of which differ depending on the amount of information the worker has.

A2.2 The experience curve effect in the technology input

A2.2.1 Dispersively embodied accumulative innovation in capital

To understand the mechanism for the creation of non-accumulative innovations, it is first necessary to examine how workers are in contact with capital inputs and the accumulative innovations embodied in them at each production site. Any single machine or tool cannot embody all the accumulated innovations in human history. Only a portion of accumulated innovations are embodied in each machine or capital input. Furthermore, different types of machines or tools embody different kinds of accumulative innovations. This relationship between accumulative innovation and capital suggests that accumulative innovations are varied, divisible, and dispersed among capital inputs. If there are negative effects of congestion and redundancy in the embodiment of accumulative innovation in capital, this division of accumulative innovation improves productivity. Embodying more types of accumulative innovations in a machine or tool may make it a more general purpose machine or tool. In implementing a specific task, however, a general purpose machine or tool will be less useful and efficient than a specialized one because congestion and redundancy of the accumulative innovations will occur and reduce efficiency.

Suppose that there is only one economy in the world and that all workers in the economy are identical. Let $Y = f(A, K, L)$ be a production function where *Y* is production, *A* is technology (accumulated innovations), *K* is capital input, and *L* is labor input. *A* can be interpreted as indicating the total amount of technology and, at the same time, the total number of varieties of technology in the economy. Let also τA be the portion of *A* embodied on average in a unit of capital where τ is a positive parameter. To incorporate the idea that the division of *A* mitigates congestion and redundancy and improves efficiency for production, the following assumption is introduced:

$$\frac{\partial Y}{\partial \tau} < 0, \qquad \text{(A2)}$$

which indicates that the smaller the value of τ (i.e., the smaller the magnitude of congestion and redundancy), the larger the production Y.

On the other hand, if τ is too small, there is the possibility that a piece of A is not embodied in any part of K. Without embodying any portion of A, K is no longer a machine or tool but merely a pile of useless materials. Avoiding this abnormal situation requires a condition that any K must embody at least some portion of A. If $\tau < \frac{1}{K}$, then the total amount of A used in the economy is $\tau AK < A$, and thus some portion of A is not embodied in any K, which indicates that the condition $\frac{1}{K} \leq \tau$ is necessary for avoiding the abnormal situation and that $\tau = \frac{1}{K}$ is the threshold value. As the rationale for the condition $\frac{1}{K} \leq \tau$ with the threshold value $\tau = \frac{1}{K}$, it is assumed here that the total differential dY with respect to A and τ is positive such that

$$dY = \frac{\partial Y}{\partial A} dA + \frac{\partial Y}{\partial \tau} d\tau > 0$$

for $\tau < \frac{1}{K}$, and thus

$$\frac{dY}{d\tau} = \frac{\partial Y}{\partial A}\frac{dA}{d\tau} + \frac{\partial Y}{\partial \tau} > 0 \qquad \text{(A3)}$$

for $\tau < \frac{1}{K}$, which means that if τ is smaller than the threshold value $\frac{1}{K}$, then the reverse effect of the amount of A on production is much larger than the effect of the division of A on production. If $\frac{1}{K} \leq \tau$, then any portion of A is embodied in some K, and thereby $\frac{dA}{d\tau} = 0$ and $\frac{dY}{d\tau} = \frac{\partial Y}{\partial \tau} < 0$.

Combining the characteristics of τ shown in inequalities (A2) and (A3) indicates that the optimal value of τ is $\frac{1}{K}$. As a result of the rational behavior of firms, the optimal dispersion of accumulative innovation in capital is obtained when $\tau = \frac{1}{K}$, and thus the portion of A embodied on average in a unit of capital is always

$$\frac{A}{K}$$

in the economy. A worker faces $\frac{A}{K}$ units of accumulative innovations at any time when the worker uses a unit of capital.[2] Because A indicates the total number of varieties of technology as well as the total amount of technology, dispersively embodied A in K indicates that a worker faces $\frac{1}{K}$ of

[2] In this paper, it is assumed that there is only one economy in the world. However, actually there are many smaller economies and a small economy may utilize only a small portion of A; i.e., the size of economy will matter to the optimal value of τ if there are many economies of various sizes. The problem of the size of economy as well as the problem of aggregation is discussed more in detail in Section A3.

varieties of *A* when the worker uses a unit of capital.

A2.2.2 Specialized or generalized machines or tools

Suppose that the amount of *A* is fixed; that is, no new variety of innovation is added. If *K* increases and *A* remains fixed, the proportion of *A* embodied in a unit of *K* becomes smaller because the proportion of *A* embodied in a unit of *K* is kept equal to $\frac{A}{K}$. A smaller $\frac{A}{K}$ means that machines or tools become more specialized because the purpose of a machine or tool embodying less *A* will be more limited. The types of machines or tools used will change even if *A* does not increase. If *K* increases in this case, machines and tools will become more specialized and vice versa. The variety and type of machines or tools, that is, how specialized or generalized they are, depend not only on *A* but also on *K*.

Note, however, that generalized does not necessarily mean advanced. On the contrary, general purpose machines or tools are more primitive, and conversely, special purpose ones are more advanced. To be general purpose, machines or tools must rely more on basic or core technologies, and many specialized functions will be downgraded.

A2.2.3 Effective technology input

As argued in Section A2.1, the experience curve effect can be applied to a task as long as the task is an intellectual creative activity and the experiences can be divided by some factor. The experience curve effect is applicable to the activity of creating non-accumulative innovations to supplement imperfect accumulative innovations because (1) the activity is an intellectual creative activity and (A2) the experiences can be divided by varieties of *A* in *K* a worker encounters. A worker encounters a portion of the accumulated innovations ($\frac{A}{K}$) when the worker uses a unit of capital. The portion of accumulated innovations conveys a subset of all the information on accumulated innovations and a part of the subset overlaps with those conveyed in other portions of accumulated innovations that other workers encounter.

A worker encounters a unique combination of varieties of accumulative innovations ($\frac{A}{K}$) per unit capital. Let N_A be a worker's average encounter frequency (i.e., the worker's experience) with each variety of accumulative innovations per unit capital in a period. As $\frac{A}{K}$ increases, the number of varieties per unit capital increases; thus, N_A will decrease because the probability of encountering each of the varieties in $\frac{A}{K}$ in a period decreases. The amount of $\frac{A}{K}$ therefore will be inversely proportional to a worker's experience on a variety per capital N_A such that

$$N_A = \beta_A \left(\frac{A}{K} \right)^{-1}$$

where β_A is a positive constant. Standardizing the worker's average encounter frequency β_A equal to unity, then

$$N_A = \left(\frac{A}{K} \right)^{-1} . \tag{A4}$$

Let C_{A,N_A} be the amount of inefficiency resulting from imperfect technology (which is equivalent to imperfect accumulative innovations) embodied in capital when a worker utilizes a variety of accumulative innovations in $\frac{A}{K}$ in a period. C_{A,N_A} does indicates not the inefficiency initially generated by imperfect technology but the one remaining after being mitigated by workers' innovations. Costs increase proportionally to increases in inefficiency; thus, C_{A,N_A} also indicates costs. Conversely, C_{A,N_A}^{-1} can be interpreted as a productivity in supplementing imperfect technology by creating non-accumulative innovations when a worker utilizes a variety of accumulative innovations in $\frac{A}{K}$ in a period. The creation of non-accumulative innovations will increase as the frequency of a worker encountering a variety of accumulative innovations in $\frac{A}{K}$ increases (i.e., the productivity in supplementing imperfect technology by creating non-accumulative innovations will increase as the number of experiences increases). Hence, the inefficiency C_{A,N_A} will decrease as the encounter frequency increases. The experience curve effect indicates that inefficiency C_{A,N_A} declines (i.e., productivity C_{A,N_A}^{-1} increases) as a worker's average encounter frequency on a variety per unit capital (N_A) increases (i.e., $\frac{A}{K}$ becomes smaller) such that

$$C_{A,N_A} = C_{A,1} N_A^{-(1-\alpha)} \quad , \tag{A5}$$

where $C_{A,1}$ is the inefficiency when $N_A = 1$. Note that α is the constant parameter ($0 < \alpha < 1$) used in equation (A1).

In addition, the "effective" amount of technology input per unit capital will increase as C_{A,N_A}^{-1} increases (i.e., C_{A,N_A} decreases) because the inefficiency is mitigated by an increased amount of workers' innovations. Thus, the effective amount of technology input per unit capital when a worker uses a variety of accumulative innovations in $\frac{A}{K}$ will be directly proportional to C_{A,N_A}^{-1} (i.e., inversely proportional to C_{A,N_A}) such that

$$W_A \left(\frac{A}{K} \right)^{-1} = \frac{\gamma_A}{C_{A,N_A}} \ , \tag{A6}$$

where W_A is the effective amount of technology input per unit capital when a worker utilizes a unique combination of varieties of accumulative innovations in $\frac{A}{K}$, and γ_A is a positive constant (i.e., $\gamma_A \left(\frac{A}{K} \right)$ indicates the amount of technology input per unit capital when a worker utilizes a unique combination of varieties of accumulative innovations in $\frac{A}{K}$ in a period when $C_{A,N_A} = 1$). Substituting equations (A4) and (A5) into equation (A6) gives

$$W_A = \frac{\gamma_A}{C_{A,N_A}}\left(\frac{A}{K}\right) = \frac{\gamma_A}{C_{A,1}N_A^{-(1-\alpha)}}\left(\frac{A}{K}\right) = \frac{\gamma_A}{C_{A,1}\left(\frac{A}{K}\right)^{1-\alpha}}\left(\frac{A}{K}\right) = \frac{\gamma_A}{C_{A,1}}\left(\frac{A}{K}\right)^{\alpha}. \tag{A7}$$

As discussed in Section A2.2.1, the amount of technology embodied in a unit capital is $\frac{A}{K}$. Because technology is imperfect, however, that level of technology input cannot be effectively realized. At the same time, the inefficiency resulting from the imperfections is mitigated by non-accumulative innovations created by ordinary workers even though it is not completely removed. Equation (A7) indicates that the magnitude of mitigation depends on $\frac{A}{K}$, and that, with the mitigation, technology input per unit capital is effectively not equal to $\frac{A}{K}$ but directly proportionate to $W_A = \frac{\gamma_A}{C_{A,1}}\left(\frac{A}{K}\right)^{\alpha}$. By equation (A7), therefore, the effective technology input per unit capital ($\tilde{A}$) is

$$\tilde{A} = \upsilon_A W_A = \omega_A \left(\frac{A}{K}\right)^{\alpha} \tag{A8}$$

where υ_A and ω_A are positive constant parameters and $\omega_A = \left(\frac{\upsilon_A \gamma_A}{C_{A,1}}\right)$.

A2.3 The experience curve effect in the labor input

The task of mitigating the inefficiency resulting from fragmented and incomplete information caused by the division of labor satisfies the condition for applying the experience curve effect (Section A2.1). As shown in Section A1.3, workers' innovations are related to this inefficiency. In addition, production processes are divided by workers as part of the division of labor. Each worker encounters only a portion of the whole production process, a portion of the process conveys only a portion of information on the whole production process, and the information overlaps partially with that on other processes that other workers encounter. Hence, the experience curve effect can be applied to this task. Because labor is divided fully at the global level, inefficiency mitigation activities are correlated at the global level.

Let N_L be the production processes a worker encounters (i.e., the experience of a worker); it indicates a proportion of all production processes in the economy (N), which is here normalized such that $N = 1$. A proportion of the production process conveys a subset of all the information on the production process, and a part of the subset overlaps with subsets of information on processes that other workers encounter. Remember, in this discussion, I am assuming that there is only one economy in the world and that all workers are identical. Thus, because the experience of a worker (N_L) is inversely proportionate to the number of workers, then

$$N_L = \frac{\beta_L}{L}$$

where L is the number of workers in the economy and β_L is a constant. $\beta_L (= N_L L)$ indicates the total of all production processes in the economy such that $\beta_L = N$. Because $N = 1$, then

$$N_L = \frac{1}{L}. \tag{A9}$$

Let C_{L,N_L} be the magnitude of inefficiency in a worker's labor input caused by fragmented and incomplete information when each worker's experience is N_L. Costs will increase proportionally with increases in inefficiency, and thus C_{L,N_L} also indicates costs. C_{L,N_L}^{-1} can be interpreted as a productivity in a worker's labor input, which increases as the amount of mitigation by the worker's innovations increases.

C_{L,N_L} decreases as the amount of individually available information (i.e., experience) increases. The increased amount of information enables a worker to discover more correct mechanisms of the production processes, and this discovery reduces the inefficiency in a worker's labor input. As mentioned previously, the experience curve effect can be applied to this inefficiency mitigation mechanism. The experience curve effect indicates that C_{L,N_L} declines as the experience of a worker (N_L) increases (i.e., the number of workers deceases) such that

$$C_{L,N_L} = C_{L,1} N_L^{-(1-\alpha)}, \tag{A10}$$

where $C_{L,1}$ is the inefficiency when $N_L = 1$ (i.e., $N_L = N$ and $L = 1$). Note again that α is the constant parameter ($0 < \alpha < 1$) used in equation (A1).

In addition, because the amount of a worker's effective provision of labor input increases as productivity (C_{L,N_L}^{-1}) increases (i.e., C_{L,N_L} decreases), then the amount of a worker's effective provision of labor input ($\frac{W_L}{L}$) is directly proportional to C_{L,N_L}^{-1} (i.e., inversely proportional to C_{L,N_L}) such that

$$\frac{W_L}{L} = \frac{\gamma_L}{C_{L,N_L}}, \tag{A11}$$

where W_L is the total amount of workers' effective provision of labor input that is lowered by the inefficiency resulting from fragmented and incomplete information, and γ_L is a constant (i.e., γ_L indicates the output per worker in a period when $C_{L,N_L} = 1$). Substituting equations (A9) and (A10) into equation (A11) gives

$$W_L = \frac{\gamma_L}{C_{L,N_L}} L = \frac{\gamma_L}{C_{L,1} N_L^{-(1-\alpha)}} L = \frac{\gamma_L}{C_{L,1} L^{1-\alpha}} L = \frac{\gamma_L}{C_{L,1}} L^{\alpha}. \tag{A12}$$

The inefficiency caused by fragmented and incomplete information constrains the effective labor provision by workers. As division of labor is widened (i.e., as *L* increases), the effective labor provision by workers is more constrained. Hence, the labor input that is effectively provided by workers is not simply proportional to *L*. Equation (A12) indicates that, instead of *L*, the labor input effectively provided by workers is directly proportional to $W_L = \frac{\gamma_L}{C_{L,1}} L^{\alpha}$; thus, the effective labor input $\tilde{L}$ is

$$\tilde{L} = \upsilon_L W_L = \omega_L L^{\alpha} , \tag{A13}$$

where υ_L and ω_L are positive constant parameters and $\omega_L = \frac{\upsilon_L \gamma_L}{C_{L,1}}$.

A2.4 The experience curve effect and the capital input

As with $\tilde{A}$ and $\tilde{L}$, an inefficiency with regard to the capital input K may exist, and this inefficiency may be solved by intellectual activities of workers. If such inefficiency exists, the effective capital input would not be equal to K. However, I was unable to find a factor that significantly necessitates a worker's intellectual activities to lessen inefficiencies in utilizing capital, in particular inefficiencies that result from imperfectness or incompleteness of information on capital. Therefore, I have assumed that capital input does not necessitate workers' innovations. However, capital input is constrained by another element that is basically irrelevant to workers' intellectual activities. It is impossible for each worker to use all capital inputs existing in the economy; each worker can access only a fraction of the total amount. This accessibility constraint sets bounds to the use of capital. Nevertheless, the accessibility is basically irrelevant in terms of worker innovation because accessibilities of workers in the world are not correlated with each other at the global level and thus it is not difficult for a worker to find a correct way to access capital inputs when an unexpected incident occurs. Therefore, information on accessibility is not incomplete, and it is enough for a worker to know only local information with regard to accessibility to capital. Therefore, there is little differentiation among workers in finding correct ways to access capital inputs, and as a consequence, there is little differentiation in the workers' experiences.

Machines or tools are not necessarily in constant operation during production; they are idle during some periods. A worker often uses various machines or tools in turn in a period, or equivalently several workers often use the same machine or tool in turn in a period. Let σK be the portion of K used by a worker on average where $\sigma(0 < \sigma \leq 1)$ is a positive parameter. Because the total sum of K used in the economy must not be smaller than K, $K \leq \sigma KL$, $\frac{1}{L} \leq \sigma$, and thereby $\frac{1}{L} \leq \sigma \leq 1$ for $1 \leq L$. It is highly likely that production increases if more K is used per worker, in which case, for a production function $Y = f(\sigma, A, K, L)$,

$$\frac{\partial Y}{\partial \sigma} > 0 . \tag{A14}$$

Condition (A14) and the constraint $\frac{1}{L} \leq \sigma \leq 1$ lead to a unique steady state value of σ such that $\sigma = 1$, which indicates that each worker uses all K existing in the economy. Clearly, that is impossible—accessibility to capital is not limitless. Even if a worker wants to use K installed at a distant location, it is usually meaningless to do so because it is too costly. Thus, it is highly likely that there is a boundary of accessibility with regard to location. A worker can use only a small portion of K installed in the small area around the worker. That is, the value of the parameter σ has an upper bound such that

$$\frac{1}{L} \leq \sigma \leq \bar{\sigma} , \tag{A15}$$

where $\bar{\sigma}$ $(0 < \bar{\sigma} < 1)$ is a positive constant. With the upper bound $\bar{\sigma}$, by conditions (A14) and (A15), the optimal portion of *K* used by a worker on average ($\tilde{K}$) for $1 \leq L$ is

$$\tilde{K} = \bar{\sigma} K . \tag{A16}$$

The parameter $\bar{\sigma}$ represents a worker's accessibility limit to capital with regard to location.[3] The average value of $\bar{\sigma}$ in the economy will depend on the availability of physical transportation facilities. Location constraints, however, are not limited to physical transportation facilities. For example, law enforcement, regulations, the financial system, and other factors will also influence accessibility. The value of $\bar{\sigma}$ reflects the combined effects of all of these factors. The values of $\bar{\sigma}$ with regard to workers who are obliged to work at a designated location using fixed machines in a factory (e.g., workers in manufacturing industries) may be nearly identical. However, values for workers in other jobs (e.g., in service industries) will be heterogeneous depending on conditions. Even in manufacturing industries, workers engage in a variety of activities (e.g., negotiating with financial institutions or marketing), so the values of $\bar{\sigma}$ will also be heterogeneous in manufacturing industries.

Suppose that the density of capital per unit area is identical in the industrial area in the economy with an upper bound of $\bar{\sigma}$.[4] An increase of the total sum of *K* indicates an increase of the density of *K* in the industrial area; thus, the portion of *K* used by a worker also increases at the same rate as *K*. On the other hand, an increase of the total sum of *L* does not indicate any change of the density of *K* in the industrial area, and the portion of *K* used by a worker does not change.

A3 Production function

A3.1 Effective production function

Suppose that production requires some strictly positive minimum amounts of technology ($\bar{A}$), capitals ($\bar{K}$), and labors ($\bar{L}$). In addition, suppose that $\bar{A}$, $\bar{K}$, and $\bar{L}$ each do not exhibit increasing marginal product; that is, $\frac{\partial^2 f(\bar{A},\bar{K},\bar{L})}{\partial A^2} \leq 0$, $\frac{\partial^2 f(\bar{A},\bar{K},\bar{L})}{\partial K^2} \leq 0$, and $\frac{\partial^2 f(\bar{A},\bar{K},\bar{L})}{\partial L^2} \leq 0$ for a production function $Y = f(\bar{A},\bar{K},\bar{L})$. If $\lim_{A\to\infty} \frac{\partial^2 f(\bar{A},\bar{K},\bar{L})}{\partial A^2} = 0$, $\lim_{K\to\infty} \frac{\partial^2 f(\bar{A},\bar{K},\bar{L})}{\partial K^2} = 0$, and $\lim_{L\to\infty} \frac{\partial^2 f(\bar{A},\bar{K},\bar{L})}{\partial L^2} = 0$, then for sufficiently large $\bar{A}$, $\bar{K}$, and $\bar{L}$, the production function is approximated by the production function in which any of $\bar{A}$, $\bar{K}$, and $\bar{L}$ exhibits constant marginal product such that

$$Y = \psi_1 (\bar{A} + \psi_2)(\bar{K} + \psi_3)(\bar{L} + \psi_4) + \psi_5 , \tag{A17}$$

where ψ_i (*i* = 1, 2, 3, 4, 5) are constants. Here, by the assumption that production requires some strictly positive minimum amounts of $\bar{A}$, $\bar{K}$, and $\bar{L}$, then $f(0,\bar{K},\bar{L}) = 0$, $f(\bar{A},0,\bar{L}) = 0$, and $f(\bar{A},\bar{K},0) = 0$. Among the approximated production functions (A17), the production function

[3] If there are many economies with various sizes, each economy's value of $\bar{\sigma}$ may be different. The effect of the size of economy on $\bar{\sigma}$ is discussed in Section A3.

[4] An industrial area is considered here to be an area that is appropriate for economic activities and excludes deserts, deep forests, mountains, and other inaccessible areas. This concept is important when we consider the size of economy, which will be examined in detail in Section A3.

that also satisfies this minimum requirement condition is

$$Y = \psi_1 \overline{A}\,\overline{KL} .$$

If ψ_1 is standardized such that $\psi_1 = 1$, then

$$Y = \overline{A}\,\overline{KL} . \tag{A18}$$

Production function (A18) appears intuitively understandable. Each of $\overline{L}$ workers uses $\overline{K}$ capital inputs per worker with $\overline{A}$ amount of technologies embedded in each $\overline{K}$.[5] However, production function (A18) cannot be realized as it is, because there are various constraints caused by various imperfections, as I argued in Section A2. The effective amounts of technology and labor inputs are not $\overline{A}$ and $\overline{L}$ but $\tilde{A}$ and $\tilde{L}$, and the portion of $\overline{K}$ usable for a worker on average is not $\overline{K}$ but $\tilde{K}$. Hence, the approximated production function is effectively

$$Y = \tilde{A}\tilde{K}\tilde{L} . \tag{A19}$$

Here, by equations (A8), (A13), and (A16),

$$\tilde{A}\tilde{K}\tilde{L} = \omega_A \left(\frac{A}{K} \right)^{\alpha} \bar{\sigma} K \omega_L L^{\alpha} = \bar{\sigma} \omega_A \omega_L A^{\alpha} K^{1-\alpha} L^{\alpha} . \tag{A20}$$

Rational firms utilize inputs fully so as to maximize Y, and by equations (A19) and (A20), the approximate effective production function (AEPF) can be represented as

$$Y = \bar{\sigma} \omega_A \omega_L A^{\alpha} K^{1-\alpha} L^{\alpha} .$$

A3.2 The approximate effective production function

AEPF has the following properties, which have been widely assumed for production functions and are consistent with data across economies and time periods: a Cobb-Douglas functional form, a labor share of about 70%, and strict Harrod neutrality. The function therefore also has diminishing marginal products of labor, capital, and technology.

A3.2.1 Cobb-Douglas functional form

The rationale and microfoundation of the Cobb-Douglas functional form have been long argued, but no consensus has been reached. For example, Jones (2005) argues that Cobb-Douglas production functions are induced if it is assumed that ideas are drawn from Pareto distributions. Growiec (2008), however, shows that Clayton-Pareto class of production functions that nest both the Cobb-Douglas functions and the CES are induced by assuming that each of the unit factor productivities is Pareto-distributed, dependence between these marginal distributions is captured by the Clayton copula, and that local production functions are CES. AEPF provides an alternative rationale and microfoundation of the Cobb-Douglas functional form. AEPF is the typical Cobb-Douglas production function, and the keys of its Cobb-Douglas functional form are workers' innovations and the experience curve effect.

A3.2.2 A 70% labor share

[5] Remember that all workers are assumed to be identical.

The parameter α indicates the labor share in the distribution of income. Data in most economies show that labor share is about 70% (Table 1). No persuasive rationale has been presented on why the labor share is usually about 70%, but AEPF can offer one. In AEPF, the value of α is derived from the experience curve effect, and the average value of α has been shown to be about 70% in many empirical studies on the experience curve effect (e.g., Hirsch, 1956; Womer and Patterson, 1983; Dutton and Thomas, 1984; Womer, 1984; Ayres and Martinas, 1992: Williams and Terzian, 1993; OECD/IEA, 2000), which implies that workers' average rate of reducing inefficiencies is bounded. This boundary probably exists because newly added information decreases as the number of experiences increases and also because the marginal efficiency in a worker's analyzing, utilizing, and managing information (i.e., in creating innovations) decreases as the amount of information increases.

A3.2.3 Strict Harrod neutrality and balanced growth

Because AEPF is a Cobb-Douglas production function, any of Harrod, Hicks, and Solow neutralities can be assumed as the type of technology change embodied in it. However, AEPF is Harrod neutral in the strict sense such that a unit of A is neither $A^{-\frac{\alpha}{1-\alpha}}$ (Solow neutral) nor $A^{-\alpha}$ (Hicks neutral) but A^{-1} because a unit of A is defined before the functional form of AEPF is induced using the experience curve effect. This strict Harrod neutrality is a necessary condition for a balanced growth path. In the balanced growth equilibrium, the capital intensity of the economy $\frac{K}{Y}$ is kept constant, and $\frac{Y}{L}$, $\frac{K}{L}$, and A grow at the same rate. Because AEPF is strictly Harrod neutral, it is possible for a growth model based on AEPF to achieve a balanced growth path.

At first glance, the essential factor behind the strict Harrod neutrality in AEPF appears to be that both $\tilde{A}$ and $\tilde{L}$ are subject to workers' intellectual activities and the experience curve effect. However, this view is somewhat superficial. In a deeper sense, there is a more essential factor. For strict Harrod neutrality to be achieved, it is necessary that both AEPF with constant L and AEPF with constant A be homogeneous of degree 1 with regard to (A and K) and (K and L), respectively. These conditions are satisfied in AEPF because $\tilde{A}$ is $\omega_A\left(\frac{A}{K}\right)^{\alpha}$, and $\tilde{A}$ therefore is not proportionate simply to A but to $\frac{A}{K}$. That is, strict Harrod neutrality requires various types of accumulative innovations in A to be dispersed in K, which means that A and K are closely related (like two sides of the same coin). Production (Y) increases at the same rate as A and K; thus, the capital intensity $\frac{Y}{K}$ is constant.

As shown in Section A2, the nature of dispersive accumulative innovations originates in the optimization of firms to minimize inefficiencies caused by congestion and redundancy of A (i.e., to maximize effects of the division of A). Because technology input is optimal when capital is as specialized as possible, then capital is actually as specialized as possible by the optimizing behaviors of firms, which implies that the very essence of the strict Harrod neutrality and the balanced growth path lies in the optimizing behaviors of rational firms.

A3.3 The size of the economy and aggregation

Because AEPF has the Cobb-Douglas functional form, it is impossible to simply disaggregate it unless any disaggregated capital labor ratio $\frac{K}{L}$ has the same value. AEPF offers an explanation for this difficulty of disaggregation (or equivalently aggregation). The effective labor input $\tilde{L}$

indicates that division of labor is a crucial factor for Cobb-Douglas production functions. Labor is divided at the global level, and even a division of labor in a small factory is a part of the global-level division of labor. Division of labor cannot be completed within a factory, but all divided labor inputs are correlated and not viable alone. Thereby, the global-level division of labor must be considered even if we construct a local production function. However, variables reflecting the global-level division of labor (e.g., the total number of workers in the world) are not included in local Cobb-Douglas production functions; that is, the effect of the global-level division of labor is ignored. The neglect of this effect matters more when local Cobb-Douglas production functions are aggregated to higher levels because the neglected correlations of labor inputs are not accounted for in the aggregation. Therefore, it is not possible to aggregate local Cobb-Douglas production functions by simply summing them up.

A similar problem may arise when a Cobb-Douglas production function is applied to multiple economies of different sizes. Large economies exhibit properties more similar to the global economy, and small economies exhibit properties that are less similar, which implies that a Cobb-Douglas production function cannot be applied equally to large and small economies. I have assumed that there is only one economy in the world, but if multiple economies are allowed, AEPF may have to incorporate the size of economy, for example, by including additional variables. However, the same AEPF can be applied to large and small economies without consideration of the size of economy because the size of an economy relates not only to $\tilde{L}$ but also to $\tilde{A}$ and $\tilde{K}$.

Let $S\,(0 < S \leq 1)$ be the size of economy, and $S = 1$ indicates the entire global economy. Here, S is defined independently of endogenous variables Y and K but by an exogenous variable such as the spatial size of an economy's industrialized areas. Given identical population density in industrialized areas across economies, S is directly proportionate to a given L. If this spatial (population) size of economy is considered, $\tilde{A}$, $\tilde{K}$, and $\tilde{L}$ need to be modified. Suppose an economy's Y, K, L, and S are Y_X, K_X, L_X, and S_X, respectively, and A is internationally common. First, the effective capital input $\bar{\sigma} K_X$ needs to be standardized by the spatial size parameter S_X. A worker's accessibility to capital does not depend simply on K_X anymore but on the spatial density of capital $\frac{K_X}{S_X}$; thus, the capital inputs a worker can access are not $\bar{\sigma} K_X$ but $\bar{\sigma} \frac{K_X}{S_X}$. Hence, the effective capital input is not $\tilde{K}_X$ but $\frac{\tilde{K}_X}{S_X}$. Similarly, the effective technology input $\tilde{A}_X$ needs to be standardized by the spatial size of economy S_X. The dispersive nature of A implies that, although any variety of A is available to any economy, a small economy will not utilize all varieties in A but will specialize in a portion of the varieties in A. The amount of varieties an economy utilizes will depend on its size. Larger economies utilize more varieties in A, and smaller economies use fewer. With this conjecture, equation (A4) ($N_A = \left(\frac{A}{K_X} \right)^{-1}$) needs to be adjusted by the size of economy S_X such that $N_A = \left(\frac{S_X A}{K_X} \right)^{-1}$; thus, by substituting $N_A = \left(\frac{S_X A}{K_X} \right)^{-1}$ into equation (A7), the effective technology input is not $\tilde{A}_X$ but $S_X^{\alpha} \tilde{A}_X$. Finally, the effective labor input is no longer $\tilde{L}_X$. As was mentioned above, S is directly proportionate to L given an identical population density. A larger S (L) superficially indicates a wider division of labor and more fragmented and incomplete information and vice versa. Thereby, an economy with a larger S (L) superficially looks more strongly affected by the inefficiency of fragmented and incomplete information than a smaller economy even though labor inputs in both large and small economies are equally divided at the global level. To remove this distortion, N_L in equation (A9) ($N_L = \frac{1}{L_X}$) must be artificially transformed to $N_L = \frac{S_X}{L_X}$ on the assumption that the size of the economy artificially becomes S_X^{-1} times as large (i.e., the same as the whole global economy). Hence, by substituting $N_L = \frac{S_X}{L_X}$ into equation (A12), $\tilde{L}_X$ is modified to $\frac{\tilde{L}_X}{S_X^{\alpha}}$. Nevertheless, the actual labor input of the economy is S_X times smaller; thus, $\frac{\tilde{L}_X}{S_X^{\alpha}}$ must be multiplied by S_X to be used as the amount of labor input of the economy. The effective labor input is thereby not $\tilde{L}_X$ but $S_X^{1-\alpha}\tilde{L}_X$.

Substituting $S_X^{\alpha}\tilde{A}_X$, $\frac{\tilde{K}_X}{S_X}$, and $S_X^{1-\alpha}\tilde{L}_X$ for $\tilde{A}$, $\tilde{K}$, and $\tilde{L}$, respectively, in equation (A19) as the effective technology, capital and labor inputs, AEPF adjusted for economy size is

$$Y_X = S_X^{\alpha}\tilde{A}_X \frac{\tilde{K}_X}{S_X} S_X^{1-\alpha}\tilde{L}_X = \tilde{A}_X \tilde{K}_X \tilde{L}_X . \tag{A21}$$

Equation (A21) is exactly the same as equation (A19). The spatial (population) size of the economy therefore does not matter, and AEPF can be applied equally to large and small economies. In addition, because equation (A21) holds for any size economy, simple comparisons of the values of the parameters $\bar{\sigma}$, ω_A, and ω_L between large and small economies are possible, which enables us to evaluate the effects of heterogeneous parameter values on production. If estimated parameter values are different between two economies when the same AEPF is used, these different values should be interpreted not as a result of distortions caused by size but as reflecting intrinsically different economic structures between the two economies.

It must be noted, however, that aggregation is still impossible as is true with other Cobb-Douglas production functions unless $\frac{K}{L}$ is identical. Although *S* does not matter to the relation among *Y*, *K*, and *L*, aggregation demands an additional more restrictive constraint on the relation among *Y*, *K*, and *L* such that $Y_1 + Y_2 = f\left(K_1 + K_2, L_1 + L_2\right)$, where Y_i, K_i, and L_i indicate *Y*, *K*, and *L* for economy *i*. It is not the spatial size (*S*) but the size of *Y* that matters.

References

Abramovitz, Moses. (1986) "Catching Up, Forging Ahead, and Falling Behind," *Journal of Economic History*, Vol. 46, No. 2, pp. 385-406.

Adler, Paul S. and Kim B. Clark. (1991) "Behind the Learning Curve: A Sketch of the Learning Process," *Management Science*, Vol. 37, No. 3, pp. 267-281.

Alchian, Armen A., (1963) "Reliability of Progress Curves in Airframe Production," *Econometrica*, Vol. 31, No. 4, pp. 679–693.

Altonji, Joseph G. and David Card (1991) "The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives," in John M. Abowd and Richard B. Freeman, eds., *Immigration, Trade and the Labor Market*, University of Chicago Press, Chicago.

Argote, Linda, Bill McEvily and Ray Reagans. (2003) "Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes," *Management Science*, Vol. 49, No. 4, pp. 571–582.

Ayres, Robert U. and Katalin Martinas. (1992), "Experience and Life Cycle: Some Analytical Implications", *Technovation*, Vol. 12, pp.. 465.

Baldwin, Carliss, Christoph Hienerth and Eric von Hippel. (2006) "How User Innovations Become Commercial Products: A Theoretical Investigation and Case Study," *Research Policy*, Vol. 35, No. 9, pp. 1291-1313.

Barro, Robert J. (1991) "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, Vol. 106, No. 2, pp. 407-43.

Baumol, William J. (1986) "Productivity Growth, Convergence, and Welfare: What the Long-run Data Show," *American Economic Review*, Vol. 76, No. 5, pp. 1072-85.

Bernard, Andrew B. and Steven N. Durlauf. (1995) "Convergence in International Output," *Journal of Applied Econometrics*, Vol. 10, No. 2, pp. 97-108.

Bodvarsson, Örn B. and Hendrik Van den Berg (2009) *The Economics of Immigration: Theory and Policy*, Springer Verlag, Berlin.

Borjas, George J. (1994) "The Economics of Immigration," *Journal of Economic Literature*, Vol. 32, No. 4, pp. 1667-1717.

Borjas, George J. (1999) "The Economic Analysis of Immigration," in *Handbook of Labor Economics*, Edition 1, Vol. 3, No. 3. (Ashenfelter, O. and D. Card, ed.), Elsevier, Amsterdam.

Borjas, George J. (2003) "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, Vol. 118, No. 1135-74.

Boston Consulting Group (BCG). (1972) *Perspectives on Experience*, Boston, Mass.

Card, David (2005) "Is the New Immigration Really So Bad?" *Economic Journal*, Vol. 115, pp.300-323.

Card, David (2009) "Immigration and Inequality," *American Economic Review*, Vol. 99, No. 2, pp. 1–21.

Cheung, Yin-Wong and Antonio Garcia-Pascual. (2004) "Testing for Output Convergence: A Re-examination," *Oxford Economic Papers*, Vol. 56, No. 1, pp. 45-63.

Dutton, John M. and Annie Thomas. (1984) "Treating Progress Functions as a Managerial Opportunity," *Academy of Management Review*, Vol. 9, No. 2, pp. 235-247.

Friedberg, Rachel and Jennifer Hunt (1995) "The Impact of Immigration on Host County Wages, Employment and Growth," *Journal of Economic Perspectives*, Vol. 9, No. 2, pp. 23-44.

Giordani, Paolo E. and Michele Ruta (2011) "The Immigration Policy Puzzle," *Review of International Economics*, Vol. 19, No. 5, pp. 922-935.

Growiec, Jakub (2008) "A New Class of Production Functions and an Argument Against Purely Labor-Augmenting Technical Change," *International Journal of Economic Theory*, Vol. 4, No. 4.

Harashima, Taiji (2009) "A Theory of Total Factor Productivity and the Convergence Hypothesis:

Workers' Innovations as an Essential Element," *MPRA* (*The Munich Personal RePEc Archive*) *Paper* No. 15508.
Harashima, Taiji (2011) "A Model of Total Factor Productivity Built on Hayek's View of Knowledge: What Really Went Wrong with Socialist Planned Economies?" *MPRA* (*The Munich Personal RePEc Archive*) *Paper* No. 29107.
Harashima, Taiji (2012) "A Theory of Intelligence and Total Factor Productivity: Value Added Reflects the Fruits of Fluid Intelligence," *MPRA* (*The Munich Personal RePEc Archive*) *Paper* No. 43151.
Hirsch, Werner Z. (1952) "Manufacturing Progress Functions," *The Review of Economics and Statistics*, Vol. 34, No. 2, pp. 143-155.
Hirsch, Werner Z. (1956) "Firm Progress Ratios," *Econometrica*, Vol. 24, pp. 136-143.
Jones, Charles I. (2005) "The Shape of Production Functions and the Direction of Technical Change," *The Quarterly Journal of Economics*, Vol. 120, No. 2, pp. 517-549.
Lawrance, Emily C. (1991) "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy*, Vol. 99, No. 1, pp. 54–77.
Lee, Samson S., John C. Dugger and Joseph C. Chen. (1999) "Kaizen: An Essential Tool for Inclusion in. Industrial Technology Curricula," *Journal of Industrial Technology*, Vol. 16, No. 1.
Luecke, Richard and Ralph Katz (2003). *Managing Creativity and Innovation*, Harvard Business School Press, Boston, MA.
Mankiw, N Gregory, David Romer and David N. Weil. (1992) "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, Vol. 107, No. 2, pp. 407-37
Michelacci, Claudio and Paolo Zaffaroni. (2000) "(Fractional) Beta Convergence," *Journal of Monetary Economics*, Vol. 45, No. 1, pp. 129-153.
OECD/IEA (2000) *Experience Curves for Energy Technology Policy*, The Organization for Economic Co-operation and Development (OECD), Paris, France.
Ottaviano, G., and G. Peri (2012) "Rethinking the Effects of Immigration on Wages," *Journal of the European Economic Association*, Vo. 10, No. 1, pp. 152–197.
Pisano, Gary P., Richard M.J. Bohmer and Amy C. Edmondson. (2001) "Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimal Invasive Cardiac Surgery," *Management Science*, Vol. 47, pp. 752-768.
Prescott, Edward C. (1998) "Needed: A Theory of Total Factor Productivity," *International Economic Review*, Vol. 39, No. 3, pp. 525-51.
Rapping, Leonard. (1965) "Learning and World War II Production Functions," *The Review of Economic Statistics*, Vol. 47, No. 1, pp. 81–86.
Romer, Paul Michael. (1986) "Increasing Returns and Long-run Growth," *Journal of Political Economy*, Vol. 94, No. 5, pp. 1002-37.
Romer, Paul Michael. (1987) "Growth Based on Increasing Returns Due to Specialization," *American Economic Review*, Vol. 77, No. 2, pp. 56-62.
Samwick, Andrew A. (1998) "Discount Rate Heterogeneity and Social Security Reform," *Journal of Development Economics*, Vol. 57, No. 1, pp. 117–146.
Sorenson, Olav. (2003) "Interdependence and Adaptability: Organizational Learning and the Longterm Effect of Integration," *Management Science*, Vol. 49, pp. 446-463.
Ventura, Luigi. (2003) "Direct Measure of Time-preference," *Economic and Social Review*, Vol. 34, No. 3, pp. 293–310.
Wiersma, Eelke. (2007) "Conditions That Shape the Learning Curve: Factors That Increase the Ability and Opportunity to Learn," *Management Science*, Vol. 53, No. 12, pp. 1903-1915.
Williams, R.H. and Terzian, G. (1993), "A Benefit/Cost Analysis of. Accelerated Development of Photovoltaic Technology," *PU/CEES Report No. 281*, Center for Energy and Environmental Studies, Princeton University, NJ.
Wright, T. P. (1936) "Factors Affecting the Cost of Airplanes," *Journal of the Aeronautical*

Sciences, Vol. 3, No. 4, pp. 122-128.
Womer, Norman Keith. (1984) "Estimating Learning Curves from Aggregate Monthly Data," *Management Science*, Vol. 30, No. 8, pp. 982-992.
Womer, Norman Keith and J. Wayne Patterson. (1983) "Estimation and Testing of Learning Curves," *Journal of Business & Economic Statistics*, Vol. 1, No. 4, pp. 265-272.