Journal of International Economics 90 (2013) 162–176

Contents lists available at SciVerse ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

# Brain drain or brain gain? Technology diffusion and learning on-the-job

Thomas Sampson

*Centre for Economic Performance, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom*

ARTICLE INFO

*Article history:*
Received 23 May 2011
Received in revised form 29 October 2012
Accepted 30 October 2012
Available online 7 November 2012

*JEL classification:*
F2
J24
M53
O33

*Keywords:*
Technology diffusion
Managerial knowledge
Learning on-the-job
FDI
Brain drain

ABSTRACT

This paper studies technology transfer when technology is embodied in human capital and learning requires on-the-job communication between managers and workers. Patterns of technology diffusion depend on where high knowledge managers work and how much time they allocate to training workers. Managers appropriate the surplus training creates and in the open economy managers face a cross-country trade-off between labor costs and the value of knowledge transfer. Complementarity between country level efficiency and managerial knowledge makes learning more valuable in the North meaning that high knowledge managers choose to work in the North and globalization precipitates a brain drain of high knowledge Southern agents to the North. The brain drain reduces learning opportunities in the South and exacerbates cross-country technology differences.


## 1. Introduction

Persistent total factor productivity differences account for more than half of cross-country income variation (Caselli, 2005). This observation has stimulated a large literature seeking to understand international variation in technology adoption. In recent years much work has focussed on the role played by managerial know-how in determining productivity. Across both firms and countries there is substantial heterogeneity in management practices and better quality management is strongly correlated with higher productivity (Bloom and Van Reenen, 2010). Burstein and Monge-Naranjo (2009) estimate average welfare gains to developing countries from importing foreign managers of 3.5%. Further welfare gains may be realized if domestic agents learn from imported managers (Monge-Naranjo, 2011; Dasgupta, 2012). The potential benefits from knowledge transfers are demonstrated by Bloom et al. (2011) who use a field experiment with large Indian textile firms to show that management training dramatically improves both a firm's management and its profitability. However, the factors that determine the equilibrium allocation of managers across countries and cross-border transfers of managerial knowledge are still not fully understood.

This paper studies the effect of allowing free movement of managers across countries on international differences in learning and the distribution of knowledge. I develop a model in which firms produce both output and on-the-job knowledge transfers between managers and workers. Since managers internalize the benefits of knowledge transfers, I show that firm location is determined by a trade-off between the cost of labor and the value of knowledge. When the North is more productive than the South, labor is cheaper in the South, but knowledge is more valuable in the North. Consequently, the high skill managers from whom workers learn the most set up firms in the North, while less able managers produce in the South. This sorting of managers across countries precipitates a brain drain of the best Southern managers to the North, reduces learning in the South and increases the knowledge gap between North and South.[1]

When technological knowledge is embodied in human capital and learning requires on-the-job communication between agents,[2] then knowledge is excludable and knowledge diffusion depends on the welfare maximizing choices made by a limited supply of high knowledge agents.[3] The paper demonstrates that even in the absence of barriers to

[1] See Gibson and McKenzie (2011) for an overview of empirical findings related to brain drain.

[2] Arrow (1969) stresses the importance of inter-personal communication in facilitating technology diffusion.

[3] In contrast, much of the technology adoption literature treats technologies as disembodied ideas that are firm or country specific and studies either the barriers faced by firms seeking to invest in new technologies (Aghion et al., 2005; Howitt and Mayer-Foulkes, 2005; Parente and Prescott, 1999; Comin and Hobijn, 2009), or optimal technology transfers by firms with monopoly ownership of ideas (Helpman, 1984; Helpman et al., 2004). In both cases, the supply of high knowledge agents does not constrain technology diffusion.

*E-mail address:* t.a.sampson@lse.ac.uk.

0022-1996/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jinteco.2012.10.004

knowledge diffusion caused by market imperfections, low adoption capacity or legal constraints, and even if lower labor costs appear to make the South an attractive destination, complementarity between the value of knowledge and country specific efficiency can lead to a brain drain from South to North. Moreover, differences in managerial knowledge across firms and countries are persistent despite the existence of perfectly competitive, globally integrated markets and the absence of learning externalities. The paper captures these ideas in a model of on-the-job learning, but it is likely to be a robust feature of any embodied technology model in which cross-country heterogeneities imply that Northern agents value advanced technologies more than Southern agents.

To model knowledge diffusion I use a dynamic general equilibrium model that endogenizes both managerial location and the level of within firm knowledge transfers. I start from a span-of-control model of the firm based on Lucas (1978) and Rosen (1982) in which production requires both labor and a managerial input. There is heterogeneity across agents in entrepreneurial knowledge and high knowledge agents select into management. I allow for on-the-job knowledge transfer between managers and workers. However, learning-by-doing is not automatic. Learning and output are jointly produced within the firm, but the relative quantities produced depend on how the manager allocates her time between production and training.[4] Anticipating higher future income, workers who receive training accept lower wages and equilibrium employment contracts are such that, from the manager's perspective, she receives a price per unit of training provided. Consequently, managers fully internalize the value of knowledge transfer and allocate time such that the marginal cost of training in terms of foregone production, equals the marginal benefit in terms of lower labor costs. Higher knowledge managers allocate more time to training because, although the volumes of both output and learning produced per unit of managerial time are proportional to the manager's knowledge, the price of training is increasing in managerial knowledge.

To study cross-country knowledge transfers I consider a world with two countries: North and South. Agents in both countries draw from the same initial knowledge distribution, but the North has higher efficiency, which raises the productivity of all Northern agents in the production of both output and training. This efficiency difference can be interpreted as the North having superior institutions, infrastructure or education. In the open economy there is free trade and managers can hire workers from either country, but workers are immobile across countries. I assume that when international production teams are formed labor productivity depends on efficiency in the workers' country and managerial productivity depends on efficiency in the manager's country. Since managers cannot hire workers from both countries simultaneously, the locations of learning and output production must coincide and knowledge flows depend on where managers choose to hire workers.

Managers' choice of location depends on a trade-off between lower labor costs in the South and cross-country differences in the price of training. In equilibrium the training price is proportional to the increase in income obtained by workers who are successfully trained and become managers. The relative price of training across countries depends on the interaction between: (i) a profit effect —— holding managerial knowledge constant Northern managers make higher profits because the North has higher efficiency, and; (ii) a wage effect —— Northern workers receive a higher wage in the absence of training. Importantly, there is a complementarity between country efficiency and managerial knowledge in the profit effect. At low knowledge levels the wage effect dominates and the price of training is higher in the South, but for high knowledge managers the profit effect dominates and the price of training is higher in the North. In equilibrium there is a threshold such that all managers whose knowledge exceeds the threshold hire workers from the North. Consequently, globalization precipitates a brain drain from the South as the best Southern managers hire workers and offer training in the North. In the open economy equilibrium knowledge transfer is concentrated in the North and, instead of leading to convergence between knowledge levels in the South and North, international knowledge flows act to magnify the knowledge gap.

Although the baseline model assumes that each country has an economy wide efficiency level which determines the productivity of both output and training production, I also show that cross-country variation in sector specific efficiency levels is sufficient to generate the profit effect complementarity. Therefore, provided the North has higher efficiency in either output production or training production globalization leads to a brain drain from South to North. Extending the baseline model by lowering the elasticity of managerial productivity with respect to country efficiency weakens the profit effect complementarity between country efficiency and managerial knowledge. This can reverse the sorting of managers across countries, but only if the elasticity is sufficiently low and the relative efficiency of the South is sufficiently high.

This paper is related to the work of Monge-Naranjo (2011) and Dasgupta (2012) who study the formation of international production teams when workers can learn from managers. In Monge-Naranjo (2011) workers learn both from their manager and from economy wide knowledge spillovers. There are no efficiency differences across countries, but, prior to globalization, Northern managers have higher knowledge than Southern managers. A key finding is that the knowledge of Southern managers always converges to Northern levels when there are no knowledge spillovers. In this paper I show that, because of complementarity between managerial knowledge and country efficiency, efficiency differences imply that the most knowledgeable managers always train Northern workers. Therefore, convergence may not occur even in the absence of learning externalities.

Dasgupta (2012) incorporates learning into a dynamic version of the Antràs et al. (2006) model of offshoring and calibrates large potential welfare gains to low income countries from inflows of foreign entrepreneurial knowledge. However, Dasgupta (2012) also abstracts from cross-country efficiency differences. Consequently, globalization leads to a factor price equalization equilibrium in which all managers are indifferent between operating in North and South, some of the best Northern managers match with Southern workers and there is no brain drain from the South. Without heterogeneity in efficiency I obtain similar predictions to Dasgupta (2012).

Closer in intent to this paper is the work of Beaudry and Francois (2010) who construct a model of learning on-the-job in which agents from a country with a high discount rate or mortality rate value learning less. This reduces the training surplus making the country less attractive to foreign managers and can lead to an equilibrium with no knowledge transfer. However, unlike in this paper, all managers are homogenous and the possibility of an equilibrium without knowledge transfer follows from the existence of a traditional sector in which learning cannot occur, not from endogenous training decisions.

The paper also contributes to the debate over whether foreign direct investment (FDI) raises host country productivity through knowledge transfers to workers in multinational enterprises (MNEs). The potential of within firm knowledge transfers, in combination with labor turnover, to act as a channel through which foreign knowledge diffuses into the domestic economy is widely recognized.[5] Existing empirical work is consistent with the idea that workers learn from foreign managers, but suggests that spillovers to other firms and workers are small in developing countries. Balsvik (2011) finds that

[4] The joint production of learning and output is also found in Acemoglu (1997) and Acemoglu and Pischke (1998) in which training and output production are separated temporally, but remain linked because of labor market imperfections that tie workers to firms.

[5] See Glass and Saggi (2002) and Fosfuri et al. (2001) for models that study the FDI decision in oligopolistic markets when the foreign firm knows that domestic workers will learn its technology and may defect to its competitors. For discussion of the role played by spin-offs from hi-tech firms in knowledge diffusion in a closed economy see Chatterjee and Rossi-Hansberg (2012), Franco and Filson (2006) and Cabral and Wang (2008).

in Norway workers with experience working for a MNE contribute 20% more to plant productivity than comparable workers without MNE experience. Poole (forthcoming) finds that Brazilian firms which employ a higher share of workers with previous experience at MNEs pay higher wages, even after controlling for worker, firm and time fixed effects. However, the effect is small. A 10% increase in the share of workers with MNE experience increases wages by 0.6%. In a more limited sample Görg and Strobl (2005) find that Ghanian manufacturing firms have higher productivity when their owner previously worked for a MNE in the same industry, but only when the owner has below median education. In addition, they find no evidence that the owner having worked for a MNE in a different industry, or having received formal training from a foreign firm, affect productivity.[6] This paper suggests an explanation for the lack of substantial spillovers from FDI via the labor turnover channel in developing countries. Even if workers can learn from higher knowledge foreign managers, in equilibrium the highest skilled managers, who provide the most training, choose to set up firms in developed countries where their training has the greatest value and the managers that hire workers from developing countries provide little or no training.

The remainder of the paper is organized as follows. In Section 2 I lay out the model in a closed economy setting and then solve for the closed economy equilibrium. Section 3 extends the model to a two country world, characterizes the open economy equilibrium and discusses the welfare implications of global integration. Then in Section 4 I analyze how three extensions to the baseline model effect the sorting of managers across countries. First, I introduce sector specific efficiency levels. Next, I vary the strength of the complementarity between country efficiency and managerial knowledge. Finally, I modify the training technology to include cross-worker heterogeneity in learning ability. Section 5 concludes the paper.

## 2. Closed economy

### 2.1. Model set-up

Consider an economy populated by a mass $R$ of heterogeneous agents. Time is discrete and each agent faces a constant probability $\zeta$ of death per period. In addition, $\zeta R$ agents are born each period ensuring that the population remains constant over time. Each agent is endowed at birth with a skill level $\theta$ that represents her managerial knowledge. I will use skill and managerial knowledge interchangeably when referring to $\theta$. At birth $\theta$ is drawn from a distribution with cumulative distribution function $F$. I assume $F$ is continuously differentiable and has continuous support on $\left[\underline{\theta},\bar{\theta}\right]$, but I do not place any functional form restrictions on the shape of $F$. Each agent's skill evolves during her lifetime as she receives on-the-job training. The training technology is described in detail below.

Agents are risk neutral and seek to maximize:

$$V_t(\theta) = \sum_{s=t}^{\infty}(1-\zeta)^{s-t}\delta^{s-t}C_s, \tag{1}$$

where $\delta\in(0,1]$ is the discount factor and $C_s$ denotes consumption of the single output good, which I take as the numeraire. There is no storage technology so agents consume their entire income each period. Let $1-p\equiv(1-\zeta)\delta$ be agents' effective discount factor.

Output is produced using a span-of-control technology based on Lucas (1978). Production uses labor and a managerial input. Each agent can choose either to start a firm and become an owner-manager or to be a worker. Although there is heterogeneity across agents in managerial knowledge, all agents who select into wage labor supply $Z$ efficiency units of labor per period. $Z$ is an economy wide efficiency term that determines not only labor productivity, but also the productivity with which the managerial input and training are produced. $Z$ should be interpreted as capturing the effect of institutions, infrastructure and the education system on the efficiency with which all productive activities in an economy are performed. More developed economies have higher $Z$ and cross-country differences in $Z$ will play a central role in the open economy model in Section 3.

The managerial input is produced using managerial knowledge and a manager's time. Each manager is endowed with one unit of time per period, which she can split between production and training. A manager with skill $\theta$, who devotes a fraction $x$ of her time to production in an economy with efficiency $Z$, supplies $xZ\theta$ units of managerial input.[7] A firm that uses $X$ units of managerial input and hires $L$ workers produces output:

$$\begin{aligned} Y &= \left(\frac{X}{\alpha}\right)^{\alpha}\left(\frac{ZL}{1-\alpha}\right)^{1-\alpha}, \quad \alpha\in(0,1), \\ &= Z\left(\frac{x\theta}{\alpha}\right)^{\alpha}\left(\frac{L}{1-\alpha}\right)^{1-\alpha}. \end{aligned} \tag{2}$$

The production process requires agents with different skills to interact within a firm. These interactions can lead, through demonstration, instruction or observation, to intra-firm knowledge transfers. To introduce learning into the model I assume that a worker can acquire new skills if she receives on-the-job training from her manager. Successful training requires a worker to both receive instruction from her manager and to observe her manager in the act of production. In particular, I assume that a manager who devotes a fraction $h$ of her time to training produces $H=hx\tau Z$ units of training, where $\tau>0$ is a parameter that measures the efficiency of training. This training technology is a hybrid of an education technology that requires the investment of costly resources and a learning-by-doing technology in which workers automatically learn new skills through participating in the production process. Training is costly because it uses a manager's time and time used in training cannot be used in production. However, unlike models in which training can be provided by a specialized education sector, training also requires production experience as an input meaning it must occur on-the-job. Substituting the time constraint $x+h=1$ into the expression for $H$ gives:

$$H = h(1-h)\tau Z. \tag{3}$$

Using the terminology of Becker (1964) all training is general, not specific, and a worker with skill $\theta(t)$ in period $t$ who receives $q\leq 1$ units of training from a manager with skill $\theta'>\theta(t)$ has probability $q$ of learning her manager's skill level. Therefore, the worker's skill $\theta(t+1)$ in period $t+1$ is given by:

$$\theta(t+1) = \begin{cases} \theta' & \text{with probability q}, \\ \theta(t) & \text{otherwise}. \end{cases} \tag{4}$$

In general, it is not clear how the volume of training produced $H$ should depend on managerial skill $\theta$. In this paper the training technology is such that holding time spent training $h$ constant the volume of training produced is independent of $\theta$. However, allowing the volume of training produced to increase in managerial skill, by assuming $H=hx\tau Z\theta$, does not qualitatively affect the model's implications. The key feature of the training technology is that more

[6] See also Malchow-Møller et al. (forthcoming) and Markusen and Trofimenko (2009) for evidence from Denmark and Colombia, respectively. For case studies where labor turnover from a foreign entrant kick-starts the development of a domestic industry see Hausmann and Rodrik (2003).

[7] Rosen (1982) justifies a similar specification on the grounds that production requires a manager to spend time supervising her workers.

knowledgeable managers train workers to a higher skill level, implying that the value of each unit of training is increasing in managerial skill.

Training does not require any input from workers, meaning that workers supply $Z$ efficiency units of labor to the firm regardless of whether or not they receive training. However, because training is on-the-job, workers can only receive training from their own manager. This has two important implications: firstly, managers can never receive training, and; secondly, the location of production and training are jointly determined. The joint location of production and training will play a central role in the open economy model.

There is perfect competition in both the output market and the labor market. Employment contracts can be written for one period only and each agent chooses in every period whether they want to select into management or wage labor. Employment contracts are defined by the quadruple $\left(\theta', \tilde{w}, q, \theta\right)$ representing the manager's skill level $\theta'$, the wage paid $\tilde{w}$, the amount of training received by the worker $q$ and the worker's skill level $\theta$. Workers take the set of employment contracts in non-zero supply as given and choose the contract that maximizes Eq. (1) subject to the training technology Eq. (4). Firms take the labor supply function as given and offer the set of employment contracts that maximizes profits subject to Eqs. (2), (3) and (4).[8] Agents choose to be either workers or managers based on which occupation maximizes their expected lifetime utility given by Eq. (1). In equilibrium the market for workers with each skill level clears. Note that there are no learning externalities in the model, instead agents fully internalize the costs and benefits of training.

Finally, I will assume that the model's parameters satisfy the following restrictions:

$$\tau < \frac{4\zeta}{1-\zeta}\frac{1}{Z}, \tag{A1}$$

$$\tau > \frac{p}{1-p}\frac{1}{Z}. \tag{A2}$$

I will discuss the role played by Assumptions (A1) and (A2) when solving for the steady state equilibrium in the next sub-section, but note that since $p \geq \zeta$ Assumption (A1) implies $\tau < \frac{4p}{1-p}\frac{1}{Z}$.

### 2.2. Closed economy equilibrium

The first step in solving the model is to characterize the set of employment contracts that are written in equilibrium. Achieving this characterization permits a considerable simplification of the firm's profit maximization problem which, in turn, makes the remainder of the model tractable. Let $\Psi_t(\theta)$ denote the set of employment contract triples $\left(\theta', \tilde{w}, q\right)$ that firms offer to workers with skill $\theta$ and let $W$ and $M$ superscripts denote workers and managers, respectively. Given the specification of the training technology, the value function in Eq. (1) can be expressed as:

$$V_t(\theta) = max\left\{V_t^W(\theta), V_t^M(\theta)\right\}, \tag{5}$$

$$V_t^M(\theta) = \pi_t(\theta) + (1-p)V_{t+1}(\theta), \tag{6}$$

$$V_t^W(\theta) = \max_{\left(\theta', \tilde{w}, q\right) \in \Psi_t(\theta)} \left\{\tilde{w} + (1-p)\left[qV_{t+1}\left(\theta'\right) + (1-q)V_{t+1}(\theta)\right]\right\}, \tag{7}$$

where $\pi_t(\theta)$ is the period $t$ profit function of a firm owned by a manager with skill $\theta$. To obtain Eq. (6) I have used that a manager's skill level remains constant over time, while for Eq. (7) I have used that a worker who receives $q$ units of training from a manager with skill $\theta'$ has probability $q$ of learning the manager's skill level and probability $1-q$ of learning nothing.

I will solve for the steady state equilibrium of the economy. In steady state all value functions, the profit function, the set of employment contracts that are written and the mapping from skill levels to occupations are time invariant. Making use of this time invariance, Eq. (6) can be solved giving:

$$V^M(\theta) = \frac{\pi(\theta)}{p}. \tag{8}$$

Let $\psi(\theta) \subseteq \Psi(\theta)$ be the set of employment contract triples that maximize the worker's value function $V^W(\theta)$. To characterize the set of steady state employment contracts I will start by showing $\psi(\theta)$ is independent of $\theta$, meaning that employment contracts do not depend on workers' managerial knowledge and all workers are indifferent between all contracts that are written in equilibrium. This result follows from the assumptions that: (i) all workers supply the same quantity $Z$ of efficiency units of labor, and; (ii) the probability a worker is successfully trained is independent of her initial managerial knowledge.[9] Together, these assumptions imply that all agents who select into wage labor are symmetric as workers and, consequently, are offered the same set of employment contracts. Thus, $\Psi(\theta)$ is independent of $\theta$. Moreover, Eq. (7) implies that the expected lifetime utility of a worker who accepts an employment contract is independent of her initial managerial knowledge. Therefore, all workers are willing to accept the same set of employment contracts ensuring that $\psi(\theta)$ does not depend on $\theta$. Lemma 1 summarizes this result. A formal proof is given in Appendix A.

**Lemma 1.** *In steady state, employment contracts do not depend on workers' managerial knowledge.*

An immediate corollary of Lemma 1 is that $V^W$ is independent of $\theta$. Since all agents who select into wage labor are equally productive and, in steady state, the mapping from skill levels to occupations is stationary, the expected lifetime utility of a worker is independent of her managerial knowledge. Using this result, together with Eqs. (7) and (8) shows that for all contracts $\left(\theta', \tilde{w}, q\right)$ which occur in equilibrium the wage $\tilde{w}$ must satisfy:

$$\tilde{w} = [1-(1-p)(1-q)]V^W - (1-p)q\frac{\pi\left(\theta'\right)}{p}. \tag{9}$$

Note that when no training is given the wage is independent of the manager's skill level. Let $w$ be the "no training wage" that is paid to a worker who does not receive training. From Eq. (9), when $q = 0$ we have:

$$V^W = \frac{w}{p}, \tag{10}$$

and substituting this expression back into Eq. (9) implies that the wage $\tilde{w}\left(q, \theta'\right)$ paid by any contract that is observed in equilibrium is given by:

$$\tilde{w}\left(q, \theta'\right) = w - c\left(\theta'\right)q, \quad \text{where} \quad c\left(\theta'\right) = \frac{1-p}{p}\left[\pi\left(\theta'\right) - w\right]. \tag{11}$$

Trainees receive a lower wage today in expectation of higher income tomorrow. In particular, the wage paid is linearly decreasing

[8] For ease of exposition, I assume that whenever a firm is indifferent between two employment contracts it offers both contracts.

[9] In Section 4.3 below I analyze an extension of the model that relaxes this assumption.

in the amount of training given, with a slope proportional to the difference between the profit flow the trainee will obtain if the training is successful and the no training wage. $c(\theta')$ is the price of training —— the amount a worker will pay for one unit of training from a manager with skill $\theta'$. If the training price is greater than the no training wage then the wage paid $\tilde{w}\left(q,\theta'\right)$ can in principle be negative. To allow for this possibility I assume that agents are not credit constrained. Since a worker who receives one unit of training is certain to learn her manager's skill, no worker will pay for more than one unit of training. Subject to this constraint, both workers and managers are indifferent as to how the manager allocates the training she produces across workers within her firm. Consequently, the distribution of training within the firm is indeterminate, but this indeterminacy does not extend to any firm level or aggregate variables. The simple characterization of equilibrium wage determination given in Eq. (11) plays a central role in making the model tractable. Note that Eq. (11) holds regardless of the functional forms of the output production technology Eq. (2) or the training production technology Eq. (3).[10]

Eq. (10) shows that a worker's expected lifetime utility is independent of whether or not she receives training, meaning that managers capture the entire surplus training creates. Consequently, a manager's income depends on both output production and training production and to maximize profits the firm must take account of the value of learning. The model implies that managers appropriate the training surplus ex-ante through paying lower wages. However, an alternative set-up in which employment contracts obliged a successful trainee to become a manager in the firm which trained her, thereby allowing the firm to appropriate the training surplus ex-post, would have identical implications for managerial decision making.

Substituting Eqs. (8) and (10) into Eq. (5) it is clear that agents select into management if and only if $\pi(\theta)\geq w$. The next step is to solve for $\pi(\theta)$. Using Eqs. (2), (3), (11) and $X=xZ\theta$ the firm's profit maximization problem is:

$$\max_{h\geq 0, L\geq 0} Z\left(\frac{(1-h)\theta}{\alpha}\right)^{\alpha}\left(\frac{L}{1-\alpha}\right)^{1-\alpha}-wL+h(1-h)\tau Zc(\theta). \tag{12}$$

This is a concave problem with solution[11]:

$$L^*(\theta)=\frac{1-\alpha}{\alpha}\left[1-h^*(\theta)\right]\theta\left(\frac{w}{Z}\right)^{-\frac{1}{\alpha}}, \tag{13}$$

$$h^*(\theta)=\begin{cases}0 & \text{if} \quad \dfrac{c(\theta)}{\theta}<\dfrac{1}{\tau}\left(\dfrac{w}{Z}\right)^{\frac{\alpha-1}{\alpha}},\\ \frac{1}{2}\left[1-\dfrac{\theta}{\tau c(\theta)}\left(\dfrac{w}{Z}\right)^{\dfrac{\alpha-1}{\alpha}}\right] & \text{otherwise}.\end{cases} \tag{14}$$

A manager provides training to her workers if and only if the price of training relative to her skill level (the training price per unit of skill) exceeds a threshold that is strictly decreasing in the ratio of the no training wage to efficiency $\frac{w}{Z}$. I will call this ratio the efficiency wage since it represents the wage per efficiency unit of labor when workers receive no training. Whenever a manager provides training, the fraction of her time allocated to training is strictly increasing in both the training price per unit of skill and the efficiency wage and is bounded above by one half. A higher training price per unit of skill, or a higher efficiency wage, increases the profitability of training relative to output production causing the manager to allocate more of her time to training. However, spending less time on output production reduces the contribution of learning-by-doing to skill transfer and this creates diminishing returns to managerial time allocated to training. As a result a manager never allocates more than half her time to training.[12] Note also that, holding managerial knowledge constant, employment is decreasing in time allocated to training. More time used for training means less managerial input employed in production and, consequently, a smaller optimal workforce.

Substituting the expressions for labor demand Eq. (13) and training supply Eq. (14) into Eq. (12) the firm's profit function is:

$$\pi(\theta)=\begin{cases}\theta Z^{\frac{1}{\alpha}}w^{\frac{\alpha-1}{\alpha}} & \text{if } \dfrac{c(\theta)}{\theta}<\dfrac{1}{\tau}\left(\dfrac{w}{Z}\right)^{\dfrac{\alpha-1}{\alpha}},\\ \dfrac{1}{4\tau Zc(\theta)}\left[\theta Z^{\frac{1}{\alpha}}w^{\frac{\alpha-1}{\alpha}}+\tau Zc(\theta)\right]^2 & \text{otherwise}.\end{cases} \tag{15}$$

We can now use the definition of $c(\theta)$ in Eq. (11) together with Eq. (15) to solve for the training price per unit of skill.

$$\frac{c(\theta)}{\theta}=\begin{cases}\frac{1-p}{p}\left[Z^{\frac{1}{\alpha}}w^{\frac{\alpha-1}{\alpha}}-\frac{w}{\theta}\right] & \text{if} \quad \theta<\theta_2,\\ \dfrac{\theta Z^{\frac{1}{\alpha}}w^{\frac{\alpha-1}{\alpha}}-2w+2Z^{\frac{1}{\alpha}}w^{\frac{\alpha-1}{\alpha}}\left[\left(\frac{w}{Z}\right)^{\frac{2}{\alpha}}-\theta\left(\frac{w}{Z}\right)^{\frac{1}{\alpha}}+\frac{p}{1-p}\dfrac{\theta^2}{\tau Z}\right]^{\frac{1}{2}}}{\dfrac{4p\theta}{1-p}-\tau Z\theta} & \text{otherwise},\end{cases}$$

where $\theta_2\equiv\left(\frac{w}{Z}\right)^{\frac{1}{\alpha}}\frac{\tau Z(1-p)}{\tau Z(1-p)-p}$. Assumption (A1) ensures that the solution for $c(\theta)$ is well-defined when $\theta>\theta_2$.

The training price per unit of skill, $\frac{c(\theta)}{\theta}$, is increasing in $\theta$,[13] and $\frac{c(\theta)}{\theta}<\frac{1}{\tau}\left(\frac{w}{Z}\right)^{\frac{\alpha-1}{\alpha}}\Leftrightarrow\theta<\theta_2$. That is, $\theta_2$ is a threshold skill level such that only managers with skill above $\theta_2$ provide training. At low skill levels allocating time to training workers is unprofitable, but as managerial skill increases the price of training increases more quickly than managers' productivity in production. Consequently, the relative value of time allocated to training rises. Assumption (A2) ensures the efficiency of training is sufficiently high that there is a finite, positive skill threshold at which managers start to provide training.[14] When $\theta>\theta_2$ time allocated to training $h^*(\theta)$ is increasing in $\theta$. Note that to obtain this prediction it is sufficient to assume more skilled managers train workers to a higher skill level and the volume of training produced Eq. (3) is independent of managerial skill. Modifying the training technology Eq. (3) to allow more skilled managers to produce a greater quantity of training for a given time allocation would only strengthen the incentive for higher skilled managers to allocate more time to training. The prediction that $h^*(\theta)$ is increasing in $\theta$ is consistent with empirical work on the incidence of training at the firm level. For example, Almeida and Aterido (2010) use firm level data from 99 countries to show that the probability a firm provides formal training to its employees is higher when the firm's manager has a tertiary education or when the firm is more productive (where employment, R&D investment, exporting and foreign ownership are used as proxies for productivity).

Using Eq. (15) we also have that:

$$\pi(\theta)>w\Leftrightarrow\theta>\theta_1\equiv\left(\frac{w}{Z}\right)^{\frac{1}{\alpha}}.$$

[10] Monge-Naranjo (2011) obtains an analogous result in a setting where skill acquisition results from investment decisions by workers rather than managers' allocation of time to training.

[11] Since no worker will pay for more than one unit of training a manager can sell at most $L^*(\theta)$ units of training. To avoid a taxonomy of cases I will assume $2(1-p)(1-\alpha)>\alpha[\tau Z(1-p)-p]$, which, together with the definition of $\theta_2$ below, ensures that $H^*(\theta)<L^*(\theta)$ meaning the constraint is non-binding.

[12] The fact that the upper bound equals one half is a consequence of the quadratic form of the training technology Eq. (3). However, the implication that managers will not specialize in training holds whenever production experience is an essential input to training.

[13] To see this differentiate the expression for $\frac{c(\theta)}{\theta}$ above with respect to $\theta$.

[14] If Assumption (A2) does not hold then $\frac{c(\theta)}{\theta}<\frac{1}{\tau}\left(\frac{w}{Z}\right)^{\frac{\alpha-1}{\alpha}}$ $\forall\ \theta$ and there is no training.

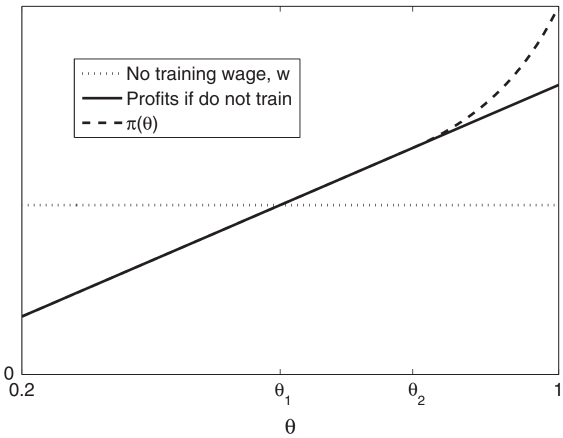


**Fig. 1.** Occupation selection in the closed economy.

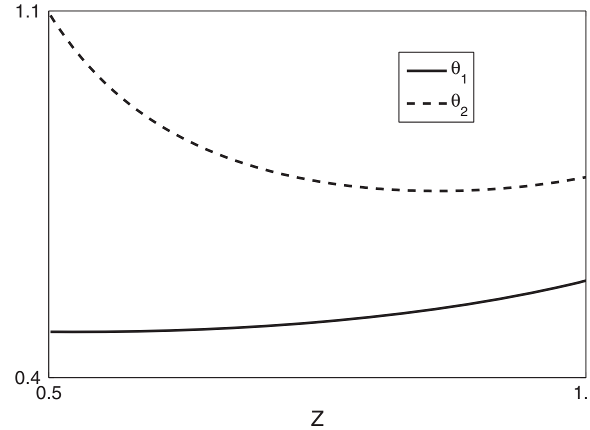


**Fig. 2.** Skill thresholds in the closed economy.

Note that $\theta_1$ is strictly increasing in the efficiency wage. It is now possible to fully characterize agents' occupational selection for a given value of the no training wage.

**Proposition 1.** *In steady state there exist threshold skill levels $\theta_1<\theta_2$ such that: (i) agents with skill below $\theta_1$ become workers; (ii) agents with skill in $(\theta_1, \theta_2)$ become managers, but do not provide any training, and; (iii) agents with skill above $\theta_2$ become managers and give training. Both thresholds are increasing in the efficiency wage $\frac{w}{Z}$.*

Fig. 1 illustrates occupation selection in the closed economy. It shows the no training wage, $w$, the profit function for a firm that does not offer training and the profit function for a firm that provides the optimal amount of training, $\pi(\theta)$, all plotted as a function of managerial skill, $\theta$.[15]

All that remains to complete the solution of the closed economy model is to find the no training wage. To do this we must first solve for the steady state skill distribution, $G(\theta)$. From Proposition 1 it follows that $G(\theta)$ must satisfy:

$$G(\theta_1) = (1-\zeta)\left[G(\theta_1) - \int_{\theta_2}^{\bar{\theta}} H^*(\theta)g(\theta)d\theta\right] + \zeta F(\theta_1),$$
$$g(\theta) = \begin{cases} (1-\zeta)g(\theta) + \zeta f(\theta) & \text{if } \theta_1 < \theta < \theta_2, \\ (1-\zeta)g(\theta) + \zeta f(\theta) + (1-\zeta)H^*(\theta)g(\theta) & \text{if } \theta_2 < \theta. \end{cases}$$

To understand the expression for $G(\theta_1)$ note that each period a mass $R\int_{\theta_2}^{\bar{\theta}} H^*(\theta)g(\theta)d\theta$ of workers is successfully trained, a fraction $\zeta$ of workers dies and $R\zeta F(\theta_1)$ new workers are born. In addition, since in steady state all workers are symmetric, we only need to keep track of the fraction of agents with skill below $\theta_1$. Similar logic can be used to obtain the remaining two expressions. Solving these equations gives:

$$G(\theta_1) = F(\theta_1) - \frac{1-\zeta}{\zeta}\int_{\theta_2}^{\bar{\theta}} \frac{\zeta H^*(\theta)}{\zeta-(1-\zeta)H^*(\theta)} f(\theta)d\theta,$$
$$g(\theta) = \begin{cases} f(\theta) & \text{if } \theta_1 < \theta < \theta_2, \\ \dfrac{\zeta f(\theta)}{\zeta-(1-\zeta)H^*(\theta)} & \text{if } \theta_2 < \theta. \end{cases} \tag{16}$$

Remembering that $h^*(\theta)$ is bounded above by one half, Assumption (A1) is sufficient to ensure that the solution for $G(\theta)$ is well-defined. Training implies that the steady state skill distribution $G$ has first order stochastic dominance over the skill distribution at birth $F$. Mass is shifted from the lower tail below $\theta_1$ to the upper tail above $\theta_2$.

Obtaining labor demand from Eq. (13) the labor market clearing condition, which determines the no training wage, is:

$$G(\theta_1) = \frac{1-\alpha}{\alpha}\frac{1}{\theta_1}\left[\int_{\theta_1}^{\theta_2}\theta g(\theta)d\theta + \int_{\theta_2}^{\bar{\theta}}\left[1-h^*(\theta)\right]\theta g(\theta)d\theta\right]. \tag{17}$$

Without assuming a functional form for $F$ we cannot solve for the no training wage explicitly. However, it can be shown that the labor market clearing condition defines a unique efficiency wage which is increasing in $Z$.

**Proposition 2.** *The steady state no training wage $w$ is uniquely determined. In any steady state with training the efficiency wage $\frac{w}{Z}$ is strictly increasing in Hicks-neutral efficiency $Z$. Otherwise, $\frac{w}{Z}$ is independent of $Z$.*

The proof of Proposition 2 is in Appendix A. For a given efficiency wage, higher $Z$ makes management more profitable which raises the value of training. Consequently, the training price per unit of skill increases causing managers to allocate more time to training and successfully train more workers. The rise in training causes labor demand per manager, given in Eq. (13), to fall, but, aggregate labor demand grows because of the higher number of managers. To bring the labor market back into equilibrium the efficiency wage must increase.

When $\theta_2 \geq \bar{\theta}$ there is no training in the steady state equilibrium and the efficiency wage is independent of $Z$. Since $\theta_2 = \theta_1\frac{\tau Z(1-p)}{\tau Z(1-p)-p}$, whenever $\bar{\theta}$ is finite there will be no training provided $Z$ is sufficiently small. Thus, there is no learning in low efficiency economies. If $\theta_2 > \bar{\theta}$ then $\theta_2$ is strictly decreasing in $Z$, but in general the effect of higher efficiency on $\theta_2$ is ambiguous. Higher $Z$ has a direct negative effect on $\theta_2$, but the induced efficiency wage increase pushes in the opposite direction. Fig. 2 plots $\theta_1$ and $\theta_2$ against $Z$ for a case where the relationship between $Z$ and $\theta_2$ is U-shaped.[16] This completes the characterization of the closed economy equilibrium.

[15] The prediction that some managers do not provide training is a consequence of the quadratic form of the training technology Eq. (3) which implies the marginal product of time allocated to training is bounded as $h \to 0$. If instead we assume $H = \frac{h^\gamma}{\gamma}\tau Z$ with $0<\gamma<1$ then all managers with $\theta>\theta_1$ give training. With this training technology we still have that time allocated to training is strictly increasing in both the efficiency wage and $\frac{c(\theta)}{\theta}$ and, under appropriate parameter restrictions, that $\frac{c(\theta)}{\theta}$ is strictly increasing in $\theta$. However, with an isoelastic training technology the model is less tractable in general equilibrium. In particular, the methodology used to prove the existence of a unique autarky equilibrium in Proposition 2 below cannot be applied.

[16] Fig. 2 is drawn with $\alpha = \frac{1}{3}, \tau = 0.9, \zeta = 0.2, \delta = 1$ and $F$ a truncated Pareto distribution on [0.2, 1] with shape parameter $k = 1$. Figs. 5 and 6 below also use these parameter values.

## 3. Open economy

### 3.1. Open economy equilibrium

To understand how globalization affects knowledge transfers suppose now that there are two countries: North and South. I am interested in how efficiency differences between North and South affect firm location and the equilibrium skill distribution in each country. I will use $N$ and $S$ superscripts to denote Northern and Southern variables, respectively. The two countries are identical in all respects except that the North has a higher efficiency, $Z^N > Z^S$, and their populations may differ. There is free trade in the output good. Workers are immobile across countries, but agents who select into management may choose to form a production team with workers in either country —— although not with workers from both countries simultaneously. To simplify the presentation I will mostly focus on the case where the South is small relative to the North, but I will also show that the main results continue to hold when both North and South are large economies.[17]

I assume that when a manager forms an international production team labor productivity is determined by efficiency in the workers' country and managerial productivity is determined by efficiency in the manager's country. Thus, if a manager from country $j$ with skill $\theta$ forms a production team with $L$ workers from country $k$ and devotes a fraction $1-h$ of her time to production then she provides $(1-h)Z^j\theta$ units of managerial input and the workers supply $Z^kL$ efficiency units of labor. There are two ways to motivate this assumption. First, consider a world in which a manager who forms an international production team is based in her home country and uses information and communications technology to interact with and monitor her workers remotely. Then it is natural to assume that each factor's productivity will depend on the efficiency in the country where that factor is located. Alternatively, suppose a manager who hires foreign workers relocates to the same country as her employees. In this case we can think of $Z^k$ as a measure of the quality of education of agents from country $k$. Both labor productivity and managerial productivity are proportional to education, but whereas individual managerial skill is transferable through training, agents are endowed with a non-transferable country specific education level prior to entering the workforce. The mathematical formulation of the model is consistent with either of these interpretations. For ease of exposition I will refer to the country in which a manager hires workers as her firm's host country or location.

The key difference from the closed economy model is that in addition to optimizing over employment and her time allocation, a manager must decide whether to hire workers from the North or the South. The firm's profit maximization problem can be broken into two stages. First, maximize profits holding location fixed. Second, choose the location with higher profits. Let $\pi^{jk}(\theta)$ be the profits made by a firm with manager from country $j$ and workers from country $k$ and let $\pi^j(\theta) = max\{\pi^{jN}(\theta), \pi^{jS}(\theta)\}$. As in the closed economy equilibrium wage contracts are given by Eq. (11), but the no training wage and the price of training are now host country dependent. Therefore, holding its location fixed, the firm faces the profit maximization problem:

$$\max_{h\geq 0, L\geq 0} \left(Z^j\right)^{\alpha}\left(Z^k\right)^{1-\alpha}\left(\frac{(1-h)\theta}{\alpha}\right)^{\alpha}\left(\frac{L}{1-\alpha}\right)^{1-\alpha} - w^kL + h(1-h)\tau Z^j c^k(\theta), \qquad (12')$$

with solution:

$$L^{jk*}(\theta) = \frac{1-\alpha}{\alpha}\left[1-h^{k*}(\theta)\right]\theta\frac{Z^j}{Z^k}\left(\frac{w^k}{Z^k}\right)^{-\frac{1}{\alpha}}, \qquad (13')$$

$$h^{k*}(\theta) = \begin{cases} 0 & \text{if } \frac{c^k(\theta)}{\theta} < \frac{1}{\tau}\left(\frac{w^k}{Z^k}\right)^{\frac{\alpha-1}{\alpha}}, \\ \frac{1}{2}\left[1-\frac{\theta}{\tau c^k(\theta)}\left(\frac{w^k}{Z^k}\right)^{\frac{\alpha-1}{\alpha}}\right] & \text{otherwise}. \end{cases} \qquad (14')$$

As in the closed economy firms only give training if the training price exceeds a threshold value, but now both the training price and the threshold depend on where the firm locates. From Eq. (13′) we see that, holding managerial skill and firm location fixed, managers from the high efficiency North will hire more workers than managers from the South. In addition, Eq. (14′) shows that the fraction of her time a manager allocates to training is independent of her country of origin. Using Eqs. (13′) and (14′) we can write $\pi^{jk}(\theta) = Z^j\theta\tilde{\pi}^k(\theta)$ where:

$$\tilde{\pi}^k(\theta) = \left[1-h^{k*}(\theta)\right]\left(\frac{w^k}{Z^k}\right)^{\frac{\alpha-1}{\alpha}} + \tau h^{k*}(\theta)\left[1-h^{k*}(\theta)\right]\frac{c^k(\theta)}{\theta}. \qquad (18)$$

It immediately follows that choice of firm location does not depend on which country a manager comes from. All managers with skill $\theta$ will locate in the country where $\tilde{\pi}^k(\theta)$ is higher. This result greatly simplifies analysis of the open economy equilibrium. It is a consequence of the fact that the equilibrium quantities of output and training produced by a profit maximizing firm are both proportional to efficiency in the manager's home country.

A second implication of Eq. (18) is that firms which do not give training always locate in the country with the lower efficiency wage. This is intuitive —— absent training managers simply seek out the lowest cost labor available. However, managers that do give training must take into account not only the cost of labor, but also the price workers will pay for training. Let $\tilde{\pi}(\theta) = max\left\{\tilde{\pi}^N(\theta), \tilde{\pi}^S(\theta)\right\}$. Then, recalling Eq. (11), we can write the price of training as:[18]

$$c^k(\theta) = \frac{1-p}{p}Z^k\left[\theta\tilde{\pi}(\theta) - \frac{w^k}{Z^k}\right]. \qquad (19)$$

Suppose $\frac{w^N}{Z^N} < \frac{w^S}{Z^S}$, then (19) and $Z^N > Z^S$ together imply $c^N(\theta) > c^S(\theta)$ whenever the price of training is positive in either country. If the North has both a lower efficiency wage and a higher price of training then all managers will hire workers from the North. In this case the Southern labor market will not clear.[19] Therefore, the efficiency wage cannot be lower in the North and we have the following result.

**Proposition 3.** *In the open economy steady state the efficiency wage is weakly higher in the North than in the South:* $\frac{w^N}{Z^N} \geq \frac{w^S}{Z^S}$. *Whenever the efficiency wage is strictly higher in the North all managers who do not give training hire workers from the South.*

The (weakly) higher efficiency wage in the North is necessary to ensure that at some skill levels managers face a trade-off between cheaper labor in the South and more valuable training in the North. This trade-off, which exists because on-the-job learning necessitates the joint production of output and training, is central to the open economy model. From Eq. (19) we see that the training price is proportional to the difference between the profits a worker will make if successfully trained and the no training wage. The no training wage is higher in the North generating a wage effect which favors training in the South. However, Northern managers make higher

[17] Since $Z^N > Z^S$ the open economy versions of Assumptions (A1) and (A2) are $\tau < \frac{4\zeta}{1-\zeta}\frac{1}{Z^N}$ and $\tau > \frac{p}{1-p}\frac{1}{Z^S}$, respectively. I assume these restrictions hold.

[18] Note that provided the volume of training produced $H$ is proportional to $Z^j$, Eq. (19) holds regardless of how $H$ depends on the time allocated to training.

[19] To rule out the possibility of an equilibrium in which all agents in the South become managers note that to obtain a non-zero labor supply in the North we must have: $\underline{\theta}\tilde{\pi}\left(\underline{\theta}\right) < \frac{w^N}{Z^N} \Rightarrow \underline{\theta}\tilde{\pi}\left(\underline{\theta}\right) < \frac{w^S}{Z^S}$. This ensures a non-zero labor supply in the South also.

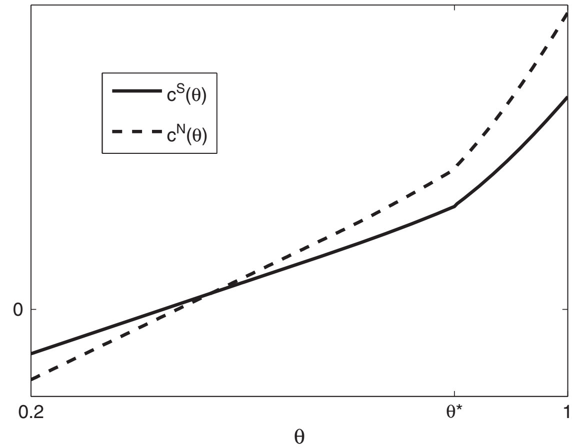


**Fig. 3.** Price of training in the open economy.

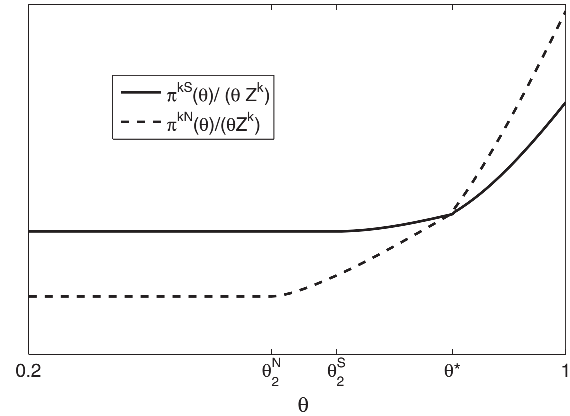


**Fig. 4.** Location choice in the open economy.

profits and this profit effect provides an incentive for training in the North. The size of the wage effect is independent of managerial skill, but from differentiating Eq. (19) we have that:

$$\frac{\partial^2 c^k(\theta)}{\partial Z^k \partial\theta} = \frac{1-p}{p}\frac{\partial}{\partial\theta}[\theta\tilde{\pi}(\theta)] > 0.$$

Thus, the training price in country $k$ is supermodular in $Z^k$ and $\theta$ and this complementarity between managerial skill and host country efficiency implies that the profit effect becomes stronger as $\theta$ increases. At low skill levels the wage effect dominates and the training price is higher in the South, but at high skill levels the profit effect dominates and the training price is higher in the North (see Fig. 3).[20]

At skill levels such that the training price is higher in the South, managers who give training must prefer to locate in the South because Southern workers are both cheaper and pay more for training. However, at skill levels for which the price of training is higher in the North, managers face a trade-off between labor costs and the value of training. Since relative labor costs across countries are independent of a manager's skill level, but the relative price of training in the North is increasing in $\theta$ the training price effect dominates the labor cost effect only at sufficiently high skill levels. Fig. 4 shows $\tilde{\pi}^S(\theta)$ and $\tilde{\pi}^N(\theta)$. Low skill managers do not give training and locate in the South where the efficiency wage is lower. As managerial knowledge increases so does the price of training and there exist threshold skill levels $\theta_2^k$, $k=S, N$ such that managers who locate in country $k$ give training whenever $\theta>\theta_2^k$. These thresholds vary across host countries, but are independent of which country the manager is from. Among managers whose skill exceeds $\theta_2^N$ profitability rises faster in the North than in the South and there exists some $\theta^*$ such that $\theta>\theta^* \Rightarrow \tilde{\pi}^N(\theta) > \tilde{\pi}^S(\theta)$ meaning managers locate in the North whenever their skill exceeds $\theta^*$. A formal proof of this result, which relies on showing that $\tilde{\pi}^S(\theta)$ and $\tilde{\pi}^N(\theta)$ satisfy a single crossing property, can be found in Appendix A as part of the proof of Proposition 4. Fig. 4 is drawn with $\theta_2^N<\theta_S^2$ and $\theta_2^S<\theta^*$, but these restrictions will not always hold.

As in the closed economy agents select into management when profits exceed the no training wage. This generates thresholds $\theta_1^j$, $j=S, N$ such that an agent in country $j$ becomes a manager if and only if her skill level $\theta>\theta_1^j$. In the proof of Proposition 4 I show that $\theta_1^S\le\theta_1^N$ with strict inequality whenever the efficiency wage is strictly higher in the North. Therefore, the skill threshold for becoming a manager is higher in the North than in the South. Proposition 4, which holds regardless of the relative sizes of North and South, summarizes equilibrium sorting into occupations and locations in the open economy. The proof is in Appendix A.

**Proposition 4.** *In the open economy steady state there exist threshold skill levels $\theta_1^S$, $\theta_1^N$, $\theta_2^S$, $\theta^*$ with $\theta_1^S\le\theta_1^N$, $\theta_1^S<\theta_2^S$ and $\theta_1^N<\theta^*$ such that: (i) Southern agents with skill below $\theta_1^S$ and Northern agents with skill below $\theta_1^N$ become workers while all other agents become managers; (ii) managers with skill below $\min\{\theta_2^S,\theta^*\}$ do not give any training; (iii) managers with skill in $(\theta_2^S,\theta^*)$ hire workers from the South and give training, and; (iv) managers with skill above $\theta^*$ hire workers from the North and give training.*

This is the main result of the paper. It tells us that in any steady state in which training occurs[21] there is a brain drain from the South as the highest skill managers set up firms in the North. Consequently: (i) managers who set up firms in the South spend less time training workers than managers in the North; (ii) Southern agents who are trained learn less than Northern trainees, and; (iii) in steady state, the North has a higher proportion of the most highly skilled agents. In fact, if $\theta^*\le\theta_2^S$, which is guaranteed when the efficiency wage is equal across countries, there will be no training in the South. In addition to the brain drain of high skill managers from South to North, Proposition 4 implies a flow of less skilled managers with $\theta\in(\theta_1^N,\theta^*)$ from North to South. These managers allocate a lower proportion of their time to training than the more highly skilled managers who operate firms in the North and, consequently, their location decision is driven by the desire to access low cost labor in the South. The existence of two way cross-border managerial flows, with managers from different segments of the skill distribution moving in opposite directions, is a key empirically testable prediction of the model. It implies the most skilled Indian agents will manage US firms, while less skilled US managers will run Indian firms.

Labor market clearing conditions pin down the no training wages $w^S$ and $w^N$, which in turn define the skill thresholds. With the skill distribution $F$ unrestricted, the labor market clearing conditions are insufficiently tractable to permit a general characterization of the dependence of $w^S$ and $w^N$ on the model's parameters. Therefore, to make further progress I will now restrict attention to the case where South is a small economy.

Under this assumption integration with the South leaves the no training wage in the North unchanged. Also, to ensure labor market

[20] Figs. 3 and 4 show the case where $\frac{w^N}{Z^N} > \frac{w^S}{Z^S}$. If $\frac{w^N}{Z^N} = \frac{w^S}{Z^S}$ then $\frac{c^N(\theta)}{c^S(\theta)} = \frac{Z^N}{Z^S}\forall\theta$ and $\tilde{\pi}^N(\theta) = \tilde{\pi}^S(\theta)\forall\theta\le\theta_2^N = \theta^*$.

[21] There will be no training in either country if $\bar{\theta}<\min\left\{\theta_2^S,\theta^*\right\}$. However, if $\frac{w^S}{Z^S}<\frac{w^N}{Z^N}$ then $\theta^*<\bar{\theta}$ must hold to ensure non-zero labor demand in the North.

clearing in the South the efficiency wage must be the same in both countries.[22] Consequently, in the open economy:

$$\frac{w^S}{Z^S} = \frac{w^N}{Z^N} = \frac{w^{N,A}}{Z^N},$$

where the A superscript is used to denote an autarky value. In both countries agents with skill in $(\theta_1^{N,A},\theta_2^{N,A})$ become managers, do not give training and are indifferent between locating in the North and the South. Agents with skill above $\theta_2^{N,A}$ become managers, give training and hire workers in the North. From Proposition 2 we know that in a closed economy $\frac{d}{dZ}\left(\frac{w}{Z}\right) > 0$ whenever some managers undertake training in equilibrium. Therefore, provided $\theta^* = \theta_2^{N,A} < \bar{\theta}$ integrating with the North raises the no training wage in the South.[23] Consequently, $\theta_1^S > \theta_1^{S,A}$ and globalization causes occupational downgrading in the South with the least skilled Southern managers moving into wage labor.[24] Moreover, it follows from $\frac{w^S}{Z^S} = \frac{w^N}{Z^N}$ that in the open economy equilibrium $w^S$ is increasing in both $Z^S$ and $Z^N$. Combining these observations gives the following result.

**Proposition 5.** *If South is a small economy then in the open economy steady state the efficiency wage is constant across countries and no training takes place in the South. Provided a positive mass of managers give training in the North, the open economy efficiency wage is higher than South's autarky efficiency wage.*

The changes in the no training wage and the skill thresholds in the South following integration with the North are instantaneous, meaning that at the individual level adjustment to the new steady state is immediate. However, whenever training occurred in the South prior to integration the skill distribution in the South displays transition dynamics as agents that received training before integration die off. Since there is no training in the South following globalization the Southern skill distribution $G^S(\theta)$ converges over time to $F(\theta)$. Consequently, the autarky steady state skill distribution first order stochastically dominates the open economy steady state skill distribution. Globalization leads to deskilling in the South as knowledge transfer concentrates in the North where the value of training is higher.

It is useful to compare the predictions of the open economy model with those obtained by Monge-Naranjo (2011) and Dasgupta (2012). In both these papers countries have the same efficiency, but differ in their initial knowledge distribution. Monge-Naranjo (2011) finds that the knowledge of Southern managers may fail to converge to Northern levels only if there are externalities in the learning technology. By contrast, this paper shows that, if efficiency differences cause the highest skill managers to sort into the North, convergence will not occur even when the value of knowledge transfers is fully internalized. In Dasgupta (2012) globalization leads to a factor price equalization equilibrium in which all managers are indifferent between North and South. Abstracting from cross-country efficiency differences in the model above leads to the same prediction. To see this point consider an alternative version of the model in which North and South have the same efficiency, but Northern agents draw their skill endowment from a better distribution. Let $Z^N = Z^S = 1$ and assume that the skill distribution at birth in the North first order stochastically dominates the distribution in the South, $F^N(\theta) \le F^S(\theta)\ \forall\theta$ with strict inequality for $\theta \in \left(\underline{\theta}^S, \bar{\theta}^N\right)$.

With these assumptions the model is similar to Dasgupta (2012) except that: (i) all workers are symmetric, meaning there is no incentive for high knowledge managers to match with high skill workers, and; (ii) the supply of training is endogenous to managers' time allocation. Using analogous reasoning to that employed in the proof of Proposition 2 it is straightforward to check that the greater supply of skills in the North implies the autarky no training wage is higher in the North than in the South. However, as in Dasgupta (2012) globalization leads to a factor price equalization equilibrium in which the no training wage and the training price are equal across countries. In the integrated equilibrium all managers are indifferent between locating in North and South.[25] Labor market clearing requires a net inflow of Northern managers to the South, but there is no sorting of managers with different knowledge levels across countries.

These comparisons highlight how cross-country efficiency differences generate the sorting which drives the key results in this paper.[26] Micro-level data on managers' characteristics and location decisions could be used to distinguish between this paper and the predictions of Monge-Naranjo (2011) and Dasgupta (2012) by testing for the existence of matching between high knowledge managers and high efficiency countries.[27]

### 3.2. Welfare

Returning to the baseline model with $Z^N > Z^S$ and assuming South is a small economy I will now consider how global integration affects the Southern welfare function, $V^S(\theta)$. Southern workers benefit from globalization. From Eq. (10) their expected lifetime utility is proportional to the no training wage, which increases following integration with the North. In addition, since $w^S$ is increasing in both $Z^S$ and $Z^N$, steady state workers' welfare is increasing in both Southern productivity $Z^S$ and Northern productivity $Z^N$. From Eq. (8) a manager's welfare is proportional to her profits. Using Eq. (18) a Southern manager who does not provide training makes profits:

$$\pi^S(\theta) = \theta Z^S \left(\theta_1^S\right)^{\alpha-1}.$$

Such managers lose from globalization because an increase in the no training wage raises $\theta_1^S$ and lowers profits. In the open economy equilibrium the welfare of managers who do not train is increasing in Southern productivity. However, it is decreasing in Northern productivity because $\theta_1^S = \theta_1^N$ which is increasing in $Z^N$ by Proposition 2. Since $\theta_1^S > \theta_1^{S,A}$ some agents switch from management to wage labor following globalization. The continuity of $V^S(\theta)$ implies that within this group of switchers relatively low skill agents' welfare increases, while relatively high skill agents experience a welfare decrease.

The effect of global integration on the welfare of Southern managers who do give training is ambiguous. On the one hand they face a higher efficiency wage which reduces the profitability of time allocated to production. On the other hand they can hire Northern workers which may increase the profitability of training. Due to the complementarity between managerial skill and host country efficiency the highest skill

[22] If $\frac{w^S}{Z^S} < \frac{w^N}{Z^N}$ then all managers that do not give training will hire workers in the South. Since a positive fraction of Northern managers have skill $\theta \in (\theta_1^N, \theta^*)$ and South is a small country it follows that labor demand will exceed labor supply in the South.

[23] If $\theta_2^{N,A} \ge \bar{\theta}$, then prior to globalization both the North and the South are in equilibria where there is no training and $\theta_1$ is independent of $Z$. In this case there is no incentive for managers to form global production teams and global integration has no effects.

[24] Since the relationship between $Z$ and $\theta_2$ in the closed economy is non-monotonic, the ordering of $\theta_2^{S,A}$ and $\theta^* = \theta_2^{N,A}$ is uncertain. However, provided $\frac{Z^S}{Z^N}$ is sufficiently small then $\theta_2^S > \theta^*$ and integration causes some high skill Southern managers who did not give training in the closed economy to start training.

[25] The only exception occurs if $\underline{\theta}^N$ is sufficiently large that all Northern agents select into management. In this case all managers must locate in the South.

[26] Remember that the equilibrium sorting characterized in Proposition 4 relies on the assumption that variation in $Z$ affects not only labor productivity, but also the productivity with which the managerial input and training are produced. See Section 4.2 for a discussion of the case where $Z$ only affects labor productivity.

[27] Although I am not aware of existing research that addresses this issue directly, Gibson and McKenzie (2011) report that emigration rates from developing countries are substantially higher for more educated individuals and that management is one of the top six occupations among tertiary educated developing country migrants to the US.

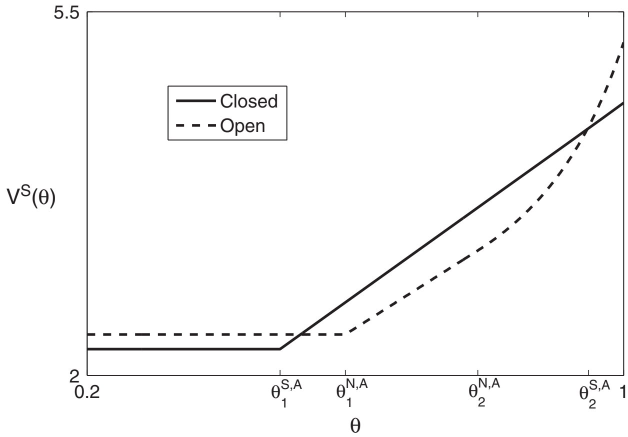


**Fig. 5.** Individual welfare effects of global integration on small South.

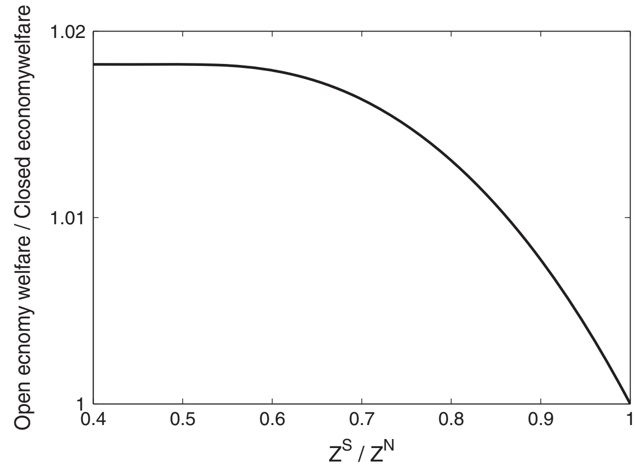


**Fig. 6.** Aggregate welfare effects of global integration on small South.

managers obtain the greatest relative benefit from globalization, but whether they gain in absolute terms is ambiguous. The lowest skill managers who give training must have lower welfare in the open economy because of the continuity of $V^S(\theta)$ at $\theta^*$. Fig. 5 plots $V^S(\theta)$ in both the closed and open economy.[28] Overall, globalization leads to a polarization in welfare outcomes. Low skill workers and possibly also the high skill managers who set up firms in the North benefit from globalization, but medium skill agents lose out.

Aggregate welfare, $W$, can be measured by the expected lifetime utility of an agent prior to learning her initial skill level. In steady state:

$$W^j = \frac{1}{p}\left[w^j F\left(\theta_1^j\right) + \int_{\theta_1^j}^{\bar{\theta}} \pi^j(\theta) f(\theta) d\theta\right] \quad j = S, N.$$

For a small economy where no training occurs, such as the South, this definition of welfare also equals the average expected utility from the current period onwards of all agents alive today, where aggregation gives equal weight to each agent. Fig. 6 plots aggregate Southern welfare in the open economy relative to autarky as a function of $\frac{Z^S}{Z^N}$. It shows that welfare in the South increases due to globalization and that the gains from integration are greatest when efficiency in the South is low meaning that globalization leads to a large increase in the Southern no training wage. As $Z^S \to Z^N$ differences between the two economies shrink, incentives to form international production teams disappear and the benefits of globalization vanish.

## 4. Efficiency, learning and sorting

This section considers the robustness of the results presented above to three extensions of the baseline model. First, I relax the assumption that each country has a single economy wide efficiency level by introducing sector specific efficiencies. Second, I allow for variation in the strength of the complementarity between country efficiency and managerial skill. Third, I modify the training technology to introduce heterogeneity in workers' ability to learn from managers. In each case I analyze whether changing the baseline model affects how managers sort across countries.

### 4.1. Sector specific efficiency

The baseline model assumes that country specific efficiency is the same in both the output sector and the training sector. Thus, in a closed economy both the output volume Eq. (2) and the training volume Eq. (3) are proportional to $Z$. In this section I analyze whether the complementarity between managerial skill and country efficiency which leads high skill managers to set up firms in the North is driven by output efficiency, training efficiency or some combination of the two. To answer this question I modify the open economy model in Section 3 to introduce sector specific efficiency levels $Z_O$ in output production and $Z_T$ in training production.[29]

When efficiency is sector specific profit maximization by a country $j$ manager who sets up a firm in country $k$ implies:

$$L^{jk*}(\theta) = \frac{1-\alpha}{\alpha}\left[1-h^{jk*}(\theta)\right]\theta \frac{Z_O^j}{Z_O^k}\left(\frac{w^k}{Z_O^k}\right)^{-\frac{1}{\alpha}}, \tag{13''}$$

$$h^{jk*}(\theta) = \begin{cases} 0 & \text{if } \frac{c^k(\theta)}{\theta} < \frac{1}{\tau}\frac{Z_O^j}{Z_T^j}\left(\frac{w^k}{Z_O^k}\right)^{\frac{\alpha-1}{\alpha}}, \\ \frac{1}{2}\left[1-\frac{\theta}{\tau c^k(\theta)}\frac{Z_O^j}{Z_T^j}\left(\frac{w^k}{Z_O^k}\right)^{\frac{\alpha-1}{\alpha}}\right] & \text{otherwise}, \end{cases} \tag{14''}$$

$$\pi^{jk}(\theta) = Z_O^j\theta\left[1-h^{jk*}(\theta)\right]\left(\frac{w^k}{Z_O^k}\right)^{\frac{\alpha-1}{\alpha}} + Z_T^j\tau h^{jk*}(\theta)\left[1-h^{jk*}(\theta)\right]c^k(\theta). \tag{20}$$

Comparing these expressions with Eqs. (13′), (14′) and (18) above two features stand out. First, the fraction of her time a manager allocates to training is no longer independent of her country meaning that her location and training choices will in general depend on which country she comes from. This implication of allowing for sector specific efficiency levels does not change any of the mechanisms that drive managerial sorting, but it does make characterizing the steady state considerably more complicated. Therefore, for the remainder of this section I will focus exclusively on the case where South is a small economy.

Second, conditional on the efficiency wage $\frac{w^k}{Z_O^k}$,[30] profits are supermodular in $\theta$ and either output efficiency $Z_O^j$ or training efficiency $Z_T^j$. Thus, managerial skill is complementary to both output efficiency and training efficiency. Since the training price inherits this complementarity from the profit function it follows that high skill managers will set up firms in the North provided $Z_O^N \geq Z_O^S$ and $Z_T^N \geq Z_T^S$ with strict inequality in at least one sector. In particular, when

[28] Fig. 5 is drawn with $Z^S = 0.57$. In both Figs. 5 and 6, $Z^N = 1.1$.

[29] With these technologies, $Z$ should be replaced by $Z_T$ in Assumptions (A1) and (A2).

[30] Note that when efficiency is sector specific the efficiency wage is defined in terms of output efficiency $Z_O$.

South is a small economy the following proposition holds. The proof is in Appendix A.

**Proposition 6.** *Suppose efficiency is sector specific and South is a small economy. Then no training occurs in the South in the open economy steady state whenever $Z_O^N \geq Z_O^S$ and $Z_T^N \geq Z_T^S$ with strict inequality in at least one case.*

Proposition 6 shows that in order for globalization to generate a brain drain from the South it is sufficient for the North to have higher efficiency in either output production or training production. The reason is that higher efficiency in either sector makes Northern managers more profitable than Southern managers and, since this North–South profit gap is increasing in managerial skill, the most skilled managers always obtain a higher training price in the North. Therefore, the baseline results do not depend on which sector drives higher profitability in the North.

### 4.2. Efficiency and managerial productivity

In the baseline model firm location is determined by a trade-off between the cost of labor and the price of training. Since the training price is convex in managerial skill, higher skill managers produce more training and their location decision is more sensitive to the relative price of training across countries. The relationship between a country's efficiency and its training price depends on two countervailing forces —— a wage effect and a profit effect. At low levels of managerial skill the former force prevails and training is more expensive in the South, but when $\theta$ is sufficiently high training is more expensive in the North. Therefore, the most skilled managers always set up firms in the more productive country. In this section I extend the model to analyze how the sorting of managers to countries depends on the strength of the profit effect, which in turn is driven by the elasticity of a manager's productivity with respect to her country's efficiency.

Suppose that we generalize the output and training technologies by assuming that a manager from country $j$ who allocates a fraction $1-h$ of her time to production produces $(1-h)\left(Z^j\right)^{\beta}\theta$ units of the managerial input and $h(1-h)\tau\left(Z^j\right)^{\beta}$ units of training, where $\beta \geq 0$.[31] $\beta$ is the elasticity of managerial productivity with respect to efficiency in the manager's country. With this technology it is straightforward to show that profit maximization implies $\pi^{jk}(\theta) = \theta\left(Z^j\right)^{\beta}\tilde{\pi}^k(\theta)$, where $\tilde{\pi}^k(\theta)$ is given by Eq. (18) and:

$$c^k(\theta) = \frac{1-p}{p}\left[\left(Z^k\right)^{\beta}\theta\tilde{\pi}(\theta) - w^k\right]. \tag{21}$$

As before time spent training is given by Eq. (14′) and both a manager's time allocation and her choice of location are independent of her home country.

Let us start by assuming that $\beta = 0$. In this case productivity depends only on host country efficiency and a manager's profits are independent of which country she comes from. This eliminates the profit effect that tends to make the price of training higher in the North. Therefore, the price of training is always higher in the South where the no training wage is lower. In equilibrium the efficiency wage is weakly lower in the North meaning that firms which do not give training always weakly prefer to locate in the North. In addition, whenever some Southern agents select into wage labor, either all training takes place in the South or some managers that train Southern workers are higher skilled than some managers that train Northern workers. This reverses the sorting of managers across countries obtained when $\beta = 1$. The intuition for this result is that when a firm's productivity is independent of its manager's home country, training is most valuable in the South where the no training wage is lowest and agents have the most to gain from learning new skills. Now let us consider managerial sorting for general $\beta > 0$. If we assume South is a small economy then we obtain the following result which is proved in Appendix A.

**Proposition 7.** *Suppose South is a small economy. Then no training occurs in the South in the open economy steady state whenever either: (i) the elasticity $\beta$ of managerial productivity with respect to country efficiency exceeds some threshold $\beta^* < 1$, or; (ii) the efficiency gap between South and North is sufficiently large.*

From Eq. (21) we see that the complementarity between managerial skill and country efficiency is stronger when $\beta$ is higher. Consequently, a high $\beta$ strengthens the profit effect on the price of training and when $\beta$ is sufficiently high, or the efficiency gap between countries is sufficiently large, the profit effect dominates the wage effect and all training takes place in the North. However, when $\beta < \beta^*$ the profit effect is weak and provided efficiency in the South exceeds some threshold $Z^*(\beta)$ then the wage effect can dominate. In this case labor market clearing in the South requires that the South has a strictly higher efficiency wage than the North and that training takes place in the South.

### 4.3. Learning capacity

The baseline model maintains the assumption that, within each economy, all workers are symmetric. Consequently, employment contracts are independent of workers' managerial knowledge. In this section I show that the paper's prediction that globalization precipitates a brain drain from South to North is robust to a modification of the training technology that introduces heterogeneity in workers' ability to acquire new skills. For simplicity I restrict attention to the case where the efficiency wage does not vary across countries. Suppose agents differ in their learning capacity $\phi$ and that a worker with learning capacity $\phi$ who receives $q$ units of training has probability $q\phi$ of acquiring her manager's skill level. Since the expected volume of training required to reach any given skill level is decreasing in $\phi$, it is reasonable to expect a positive correlation between $\phi$ and $\theta$. However, for present purposes greater generality can be maintained by leaving both the relationship between the two dimensions of agent heterogeneity and the distribution of $\phi$ unspecified.

Except for the addition of heterogeneity in learning capacity the model is unchanged from Section 3 above. We still have $\pi^{jk}(\theta) = Z^j\theta\tilde{\pi}^k(\theta)$ where $\tilde{\pi}^k(\theta)$ is given by Eq. (18). Thus, a manager's profit maximizing training and location choices are independent of her country of origin and depend only on the efficiency wages and training prices in different countries. However, the training price now depends on both the country where a firm locates and the learning capacity of the firm's trainees.

Let $c(\theta;\phi,Z)$ be the training price paid to a skill $\theta$ manager by a learning capacity $\phi$ worker in an efficiency $Z$ country. Using Eq. (7) the training price is:

$$c(\theta;\phi,Z) = \frac{1-p}{p}\phi Z\theta\tilde{\pi}(\theta) - (1-p)\phi V^W(\phi,Z) - \frac{1}{q}\left[pV^W(\phi,Z) - w(Z)\right],$$

[31] Set $\beta = 1$ to retrieve the technology used earlier in the paper. For the general technology Assumptions (A1) and (A2) become $\tau < \frac{4\zeta}{1-\zeta}\frac{1}{\left(Z^N\right)^{\beta}}$ and $\tau > \frac{p}{1-p}\frac{1}{\left(Z^S\right)^{\beta}}$, respectively.

where $V^W(\phi,Z)$ is the expected lifetime welfare of a learning capacity $\phi$ worker in an efficiency $Z$ country and the dependence of the no training wage on $Z$ is made explicit. When all workers in a country are symmetric $pV^W(Z)=w(Z)$ because managers appropriate the entire training surplus. However, when there is heterogeneity in learning capacity across workers, the training surplus may be shared between managers and workers and worker welfare may depend on $\phi$. There are two cases to consider, either: (i) $pV^W(\phi,Z)=w(Z)$, or; (ii) $pV^W(\phi,Z)>w(Z)$. In the later case the training price is increasing in the quantity of training $q$, implying that it is optimal for managers to provide each of their trainees with the maximum possible volume of training by choosing $q=\frac{1}{\phi}$. Consequently, all trainees successfully acquire their managers' skills. It follows that in both cases (i) and (ii) the training price can be rewritten as:

$$c(\theta;\phi,Z)=\frac{1-p}{p}\phi Z\theta\tilde{\pi}(\theta)-\phi\left[V^W(\phi,Z)-w(Z)\right]. \tag{22}$$

Observe that the training price is supermodular in $\theta$ and $\chi\equiv\phi Z$. Thus, there is complementarity between managerial skill and the product of a trainee's learning capacity with host country efficiency. The additional benefit of training a worker with higher $\chi$ is greater for more skilled managers. This complementarity is analogous to the training price complementarity between $\theta$ and $Z$ in the baseline model and it has two important implications for matching between managers and workers and the allocation of training within and across countries. First, within countries there is positive assortative matching between high $\phi$ workers and high $\theta$ managers. Only workers whose learning capacity exceeds some threshold value receive training[32] and trainees with a higher learning capacity are matched with better managers and learn more. Second, if the efficiency wage is constant across countries then in the open economy there is positive assortative matching between high $\chi$ workers and high $\theta$ managers. For a given efficiency wage, only the training price matters for profit maximization and the complementarity embedded in Eq. (22) ensures that more knowledgeable managers match with higher $\chi$ trainees. These predictions do not rely on the assumption that there are only two countries. Proposition 8 summarizes equilibrium matching with heterogeneity in learning capacity. The proof is in the Appendix A.

**Proposition 8.** *Suppose learning capacity $\phi$ differs across agents. In steady state: (i) there exists a threshold learning capacity in each country below which workers do not receive training and above which all workers become trainees, and: (ii) holding the efficiency wage constant, there is positive assortative matching between high skill managers and trainees with high $\phi*Z$.*

Proposition 8 shows that with heterogeneity in workers' learning capacity the training a worker receives is determined by the interaction of worker type and country efficiency. Assuming that the distribution of learning capacity does not vary between North and South, the most knowledgeable managers still choose to locate in the North precipitating a brain drain of the best Southern managers. However, provided there is sufficient variation in learning capacity,[33] then, unlike in the baseline model, there is an overlap between the skill distributions of managers who give training in the North and managers who locate in the South. High learning capacity Southern trainees are employed by more skilled managers than low learning capacity Northern trainees and, in contrast to Proposition 5, training occurs in the South even when it is a small economy. Variation in learning capacity creates an incentive for managers to set up firms in the South in order to match with high learning ability Southern workers, but it does not overturn the prediction that there exists a skill threshold above which all managers choose to locate in the high efficiency North. Consequently, Southern workers never learn the most advanced skills and in steady state there is a greater fraction of agents in the extreme right tail of the skill distribution in the North than in the South.[34]

Proposition 8 also has several interesting corollaries. First, holding learning capacity constant trainees in the North are trained by higher skill managers and attain a higher skill level than trainees in the South. Second, holding managerial skill constant, managers that locate in the South employ workers with greater learning ability than managers in the North. Third, in equilibrium firms may choose to employ two types of workers. Low learning capacity workers who do not receive training and high learning capacity workers who become trainees. This outcome occurs when the production labor supplied by trainees is insufficient to satisfy the firm's labor demand and leads to within firm heterogeneity in worker type, job type and wages.

## 5. Conclusion

This paper presents a theory in which by accident of birth Northern agents have higher productive efficiency —— be it caused by superior institutions, infrastructure or education. Country efficiency and individual knowledge are complements and if either the complementarity is sufficiently strong, or the efficiency gap between North and South is sufficiently large, the North has a comparative advantage in high knowledge production and there is a brain drain of high knowledge agents from South to North. However, if the complementarity is weak and the efficiency gap is small the pattern of comparative advantage is reversed and high knowledge agents are attracted to the South where workers with a low outside option will pay a higher price for learning.

Low levels of skill and technology transfer from developed to developing countries are often explained in terms of barriers to technology transfer or of developing countries' lack of absorptive capacity —— their missing ability to learn. This paper suggests an alternative perspective. Even if agents in all countries have the same capacity to learn and the learning technology does not discriminate between within-country and cross-country transfers, when training requires a rival input —— the time of high knowledge agents —— that is in limited supply, equilibrium knowledge transfer will depend on the income maximizing allocation of this factor.

The paper offers a benchmark model in which all markets are perfectly competitive. It would be instructive to extend the model to include labor market imperfections that require managers and workers to share the rents arising from knowledge transfer. Rent sharing would reduce the price of training from the manager's perspective leading managers to allocate less time to training and making their choice of location more sensitive to relative labor costs. Importantly, rent sharing would also mean that learning brings welfare gains in addition to higher knowledge. In this case, I hypothesize that a brain drain from South to North could result in a decrease in Southern welfare as Southern workers receive a lower share of the learning rents because of reduced training in the South. It would

[32] In general, the threshold may be sufficiently low that all workers in a country receive training or sufficiently high that no workers are trained.

[33] In particular, we require that the heterogeneity in learning capacity is sufficiently large relative to $\frac{Z^N}{Z^S}$ that the best Southern workers have a higher $\chi$ than the worst Northern trainees.

[34] To solve the full general equilibrium model with heterogeneity in learning capacity it is first necessary to specify the distribution of $\phi$ and the correlation between $\phi$ and $\theta$. However, the discussion above demonstrates that the complementarity between $\theta$ and $\chi$, which drives matching between managers and trainees, does not depend on these details.

also be interesting to consider how the model's welfare implications are affected by modifying the training technology to allow knowledge transfers to generate long run growth.

Finally, note that in applications where the amount of time a manager allocates to training is not of central interest a considerable simplification to the model can be obtained by assuming that managers spend all their time on production and skill transfer occurs through learning-by-doing. This assumption enhances the model's tractability because time allocation is no longer endogenous, while maintaining the cross-country trade-off between labor costs and the value of learning.

**Acknowledgments**

I am grateful to Pol Antràs, Gita Gopinath, Elhanan Helpman, Oleg Itskhoki, Germán Pupato and participants at LACEA-TIGN 2011 for their helpful comments and suggestions.

## Appendix A

**Proof of Lemma 1.** I will prove the result by contradiction. Let $\theta_a$ and $\theta_b$ be skill levels at which agents select into wage labor and suppose there exists an employment contract $\left(\theta'_a, \tilde{w}_a, q_a\right)$ that belongs to $\psi(\theta_a)$, but not to $\psi(\theta_b)$. Let $\left(\theta'_b, \tilde{w}_b, q_b\right)$ be an employment contract that belongs to $\psi(\theta_b)$. Since $\left(\theta'_i, \tilde{w}_i, q_i\right)\in\psi(\theta_i)$ it follows from Eq. (7) that:

$$V^W(\theta_i) = \frac{\tilde{w}_i + (1-p)q_i V\left(\theta'_i\right)}{1-(1-p)(1-q_i)}, \quad i = a, b.$$

Note that the right hand side of this expression is independent of $\theta_i$ implying that the expected lifetime utility of a worker employed on a given contract is independent of her initial managerial knowledge. In addition, since all workers supply the same quantity $Z$ of efficiency units of labor, the value to a firm of employing a worker under a given contract does not depend on the worker's managerial knowledge. Consequently, any employment contract triple $\left(\theta', \tilde{w}, q\right)$ offered by a firm, must be offered to workers of all skill levels. Thus, $\Psi(\theta_a) = \Psi(\theta_b)$.

There are now two possibilities to consider. First, $V^W(\theta_a) = V^W(\theta_b)$. In this case workers with skill $\theta_b$ are indifferent between the contract $\left(\theta'_a, \tilde{w}_a, q_a\right)$ and the contract $\left(\theta'_b, \tilde{w}_b, q_b\right)$, which contradicts $\left(\theta'_a, \tilde{w}_a, q_a\right)$ not belonging to $\psi(\theta_b)$. Second, $V^W(\theta_a) \neq V^W(\theta_b)$ and suppose without loss of generality that $V^W(\theta_a) < V^W(\theta_b)$. In this case workers with skill $\theta_a$ would obtain strictly higher expected lifetime utility under the contract $\left(\theta'_b, \tilde{w}_b, q_b\right)$ and this contradicts $\left(\theta'_a, \tilde{w}_a, q_a\right)\in\psi(\theta_a)$. It follows that $\psi(\theta_a) = \psi(\theta_b)$.

**Proof of Proposition 2.** To prove the result it is sufficient to show that the equilibrium value of $\theta_1$ is uniquely determined and is strictly decreasing in $\Delta \equiv \frac{p}{1-p}\frac{1}{\tau}\frac{1}{Z}$ if $\theta_2 < \bar{\theta}$. Noting that $\theta_2 = \frac{\theta_1}{1-\Delta}$ and substituting the steady state skill distribution Eq. (16) into Eq. (17) we can rewrite the labor market clearing condition with $\theta_1$ as the only endogenous variable:

$$F(\theta_1) - \frac{1-\zeta}{\zeta}\int_{\frac{\theta_1}{1-\Delta}}^{\bar{\theta}} \frac{\zeta H^*(\theta)}{\zeta-(1-\zeta)H^*(\theta)} f(\theta)d\theta$$
$$= \frac{1-\alpha}{\alpha}\frac{1}{\theta_1}\left[\int_{\theta_1}^{\frac{\theta_1}{1-\Delta}} \theta f(\theta)d\theta + \int_{\frac{\theta_1}{1-\Delta}}^{\bar{\theta}} \frac{\zeta[1-h^*(\theta)]}{\zeta-(1-\zeta)H^*(\theta)}\theta f(\theta)d\theta\right]. \quad (23)$$

Now, from Eq. (14), we have that $\forall\theta \geq \theta_2$:

$$h^*(\theta) = \frac{1}{2}(1-D),$$
$$H^*(\theta) = \frac{1}{4\Delta}\frac{p}{1-p}\left(1-D^2\right),$$
$$\frac{\zeta[1-h^*(\theta)]}{\zeta-(1-\zeta)H^*(\theta)} = \frac{2\Delta(1+D)}{4\Delta-\frac{p}{\delta\zeta}\left(1-D^2\right)} \equiv E, \quad (24)$$

where:

$$D \equiv \frac{\theta_1^{\alpha-1}}{\tau}\frac{\theta}{c(\theta)},$$
$$= \frac{\theta(4\Delta-1)}{\theta-2\theta_1+2\left[\theta_1^2-\theta\theta_1+\theta^2\Delta\right]^{\frac{1}{2}}}. \quad (25)$$

Noting that Assumption (A1) implies $4\Delta > 1$, differentiation of $D$ can be used to show that $\frac{\partial D}{\partial\theta_1} > 0$ and $\frac{\partial D}{\partial\Delta} > 0$ which imply $\frac{\partial H^*(\theta)}{\partial\theta_1} < 0$ and $\frac{\partial H^*(\theta)}{\partial\Delta} < 0 \ \forall\theta > \theta_2$.

In addition, by differentiating $E$ with respect to $D$ we obtain:

$$\frac{\partial E}{\partial D} = \frac{2\Delta}{\left[4\Delta-\frac{p}{\delta\zeta}\left(1-D^2\right)\right]^2}\left[4\Delta-\frac{p}{\delta\zeta}(1+D)^2\right].$$

Substituting for $D$ and using $\theta > \theta_2$ we have $(1+D)^2 > 4\Delta$ and since $\frac{p}{\delta\zeta} \geq 1$ it follows that $\frac{\partial E}{\partial D} < 0$. Combining this with $\frac{\partial D}{\partial\theta_1} > 0$ it immediately follows that $\frac{\partial E}{\partial\theta_1} < 0$. A similar argument shows that $\frac{\partial E}{\partial\Delta} < 0$.

With these results in hand we can proceed to prove the proposition. The left hand side of Eq. (23) gives labor supply as a function of $\theta_1$. It is continuous, non-positive when $\theta_1 = \underline{\theta}$, equal to 1 when $\theta_1 = \bar{\theta}$ and, since $\frac{\partial H^*(\theta)}{\partial\theta_1} < 0 \ \forall \ \theta > \theta_2$, it is strictly increasing in $\theta_1$. The right hand side of Eq. (23) gives labor demand. It is continuous, positive when $\theta_1 = \underline{\theta}$ and zero when $\theta_1 = \bar{\theta}$. Moreover, differentiating the right hand side of Eq. (23) with respect to $\theta_1$ and using $h^*(\theta_2) = H^*(\theta_2) = 0$ and $\frac{\partial E}{\partial\theta_1} < 0$ shows that it is strictly decreasing in $\theta_1$. It immediately follows that Eq. (23) defines a unique solution for $\theta_1$ on $\left(\underline{\theta}, \bar{\theta}\right)$.

To prove the second part of the proposition, first note that when $\theta_2 \geq \bar{\theta}$ Eq. (23) is independent of $\Delta$, meaning that $\theta_1$ does not depend on $Z$. When $\theta_2 < \bar{\theta}$ analogous arguments to those used above show that the left hand side of Eq. (23) is strictly increasing in $\Delta$, while the right hand side is strictly decreasing in $\Delta$. Consequently, we must have that $\theta_1$ is strictly decreasing in $\Delta$.

**Proof of Proposition 4.** Let us start by considering agents' selection into occupations. The skill threshold, $\theta_1^j$, at which country $j$ agents are indifferent between management and wage labor must satisfy $\pi^j\left(\theta_1^j\right) = w^j \Leftrightarrow \theta_1^j\,\tilde{\pi}\left(\theta_1^j\right) = \frac{w^j}{Z^j}$. Since $\frac{w^S}{Z^S} \leq \frac{w^N}{Z^N}$ and $\theta\tilde{\pi}(\theta)$ is strictly increasing in $\theta$ it immediately follows that $\theta_1^S \leq \theta_1^N$ with equality if and only if $\frac{w^S}{Z^S} = \frac{w^N}{Z^N}$. Moreover, using the expression for the price of training in Eq. (19) we have that $c^j(\theta_1^j) = 0$, $j = S, N$ and from the solution for the amount of time allocated to training in Eq. (14′) we know that $\frac{c^j\left(\theta_2^j\right)}{\theta_2^j} = \frac{1}{\tau}\left(\frac{w^k}{Z^k}\right)^{\frac{\alpha-1}{\alpha}} > 0$. Since higher skill training is more valuable $c^j(\theta)$ is strictly increasing in $\theta$ and, therefore, $\theta_1^j < \theta_2^j$, $j = S, N$.

Now consider a manager's choice of location when $\frac{w^S}{Z^S} < \frac{w^N}{Z^N}$. Since $\theta_1^j < \theta_2^j$, $j = S, N$ there is always a positive mass of managers who do not give training and, because the South has a lower efficiency wage, these managers will always locate in the South. Labor market clearing in the South then requires $\theta_1^S > \underline{\theta}$ and using the result proved above that $\theta_1^S \leq \theta_1^N$ we have that $\theta_1^N > \underline{\theta}$. Given that a positive mass of Northern agents choose to become workers labor market clearing in the North necessitates that a positive mass of managers choose to locate in the North.

Define $\theta^* \equiv inf\left\{\theta : \tilde{\pi}^N(\theta) > \tilde{\pi}^S(\theta)\right\}$. From the continuity of $\tilde{\pi}^N(\theta)$ and $\tilde{\pi}^S(\theta)$ it must be that $\tilde{\pi}^N(\theta^*) = \tilde{\pi}^S(\theta^*)$. Obviously, $\theta_2^N < \theta^* < \bar{\theta}$. By applying the envelope theorem to the definition of $\tilde{\pi}^k(\theta)$ in Eq. (18) we obtain:

$$\frac{d\tilde{\pi}^k(\theta)}{d\left(\frac{c^k(\theta)}{\theta}\right)} = \tau h^{k*}(\theta)\left[1-h^{k*}(\theta)\right].$$

Now note that: (i) $h < \frac{1}{2} \Rightarrow \frac{\partial}{\partial h}[h(1-h)] > 0$; (ii) $\tilde{\pi}^N(\theta) \geq \tilde{\pi}^S(\theta) \Rightarrow c^N(\theta) > c^S(\theta)$ since if the South has both a lower efficiency wage and a higher training price then firms must make strictly higher profits in the South, and; (iii) from Eq. (14′), $c^N(\theta) > c^S(\theta)$ and $\frac{w^S}{Z^S} < \frac{w^N}{Z^N}$ together imply $h^{N*}(\theta) > h^{S*}(\theta)$. Combining these three observations we have that $\frac{d\tilde{\pi}^N(\theta)}{d\left(\frac{c^N(\theta)}{\theta}\right)} > \frac{d\tilde{\pi}^S(\theta)}{d\left(\frac{c^S(\theta)}{\theta}\right)}$ whenever $\tilde{\pi}^N(\theta) \geq \tilde{\pi}^S(\theta)$.

Next, from Eq. (18) and Eq. (19) we have that if $\tilde{\pi}(\theta) = \tilde{\pi}^k(\theta)$ then:

$$\tilde{\pi}(\theta) = \frac{p\left[1-h^{k*}(\theta)\right]\left(\frac{w^k}{Z^k}\right)^{\frac{\alpha-1}{\alpha}} - (1-p)\tau h^{k*}(\theta)\left[1-h^{k*}(\theta)\right]\frac{w^k}{\theta}}{p-(1-p)\tau h^{k*}(\theta)\left[1-h^{k*}(\theta)\right]Z^k}.$$

It follows that $\tilde{\pi}(\theta)$ is non-decreasing in $\theta$ and using Eq. (19) this implies that both $\frac{c^S(\theta)}{\theta}$ and $\frac{c^N(\theta)}{\theta}$ are strictly increasing in $\theta$. Finally:

$$\frac{c^N(\theta)}{\theta}\frac{\theta}{c^S(\theta)} = \frac{Z^N\theta\tilde{\pi}(\theta)-w^N}{Z^S\theta\tilde{\pi}(\theta)-w^S},$$

which is strictly increasing in $\theta$ given $\frac{w^S}{Z^S} < \frac{w^N}{Z^N}$. Putting everything together we have that an increase in $\theta$ raises $\frac{c^N(\theta)}{\theta}$ by more than $\frac{c^S(\theta)}{\theta}$, which in turn increases $\tilde{\pi}^N(\theta)$ by more than $\tilde{\pi}^S(\theta)$ whenever $\tilde{\pi}^N(\theta) \geq \tilde{\pi}^S(\theta)$. Therefore, $\tilde{\pi}^N(\theta)$ and $\tilde{\pi}^S(\theta)$ satisfy a single crossing property and all managers with skill $\theta > \theta^*$ set up firms in the North.

If $\frac{w^S}{Z^S} = \frac{w^N}{Z^N}$ then $\theta_1^S = \theta_1^N$ and $\frac{c^N(\theta)}{c^S(\theta)} = \frac{Z^N}{Z^S}$. Managers who do not give training are indifferent between locating in the North and the South, while managers that do give training always prefer the North because it has a higher training price. In this case $\theta^* = \theta_2^N < \theta_2^S$.

**Proof of Proposition 6.** First, observe that since South is a small economy labor market clearing in the South implies that the efficiency wage is weakly lower in the North $\frac{w^N}{Z_O^N} \leq \frac{w^S}{Z_O^S}$. Therefore, a necessary condition for training to take place in the South is that there exists $\theta_a$ such that $c^S(\theta_a) \geq c^N(\theta_a)$. Let $k \equiv arg\ max_{k\in\{S,N\}}\pi^{Sk}(\theta_a)$ be the location choice of Southern managers with skill $\theta_a$. Using Eq. (11) and Eq. (20) we then have:

$$c^S(\theta_a) = \frac{1-p}{p}\left[Z_O^S\left(\theta_a\left[1-h^{Sk*}(\theta_a)\right]\left(\frac{w^k}{Z_O^k}\right)^{\frac{\alpha-1}{\alpha}} - \frac{w^S}{Z_O^S}\right) + Z_T^S\tau h^{Sk*}(\theta_a)\left[1-h^{Sk*}(\theta_a)\right]c^k(\theta_a)\right].$$

There are two cases to consider. First, suppose $\theta_a[1-h^{Sk*}(\theta_a)]\left(\frac{w^k}{Z_O^k}\right)^{\frac{\alpha-1}{\alpha}} > \frac{w^S}{Z_O^S}$. Then since $Z_O^N \geq Z_O^S$ and $Z_T^N \geq Z_T^S$ with strict inequality in at least one case and Northern managers can always choose to copy the location and training choices of Southern managers it immediately follows that $c^S(\theta_a) < c^N(\theta_a)$. Second, suppose $\theta_a\left[1-h^{Sk*}(\theta_a)\right]\left(\frac{w^k}{Z_O^k}\right)^{\frac{\alpha-1}{\alpha}} \leq \frac{w^S}{Z_O^S}$. Then:

$$\begin{aligned} c^S(\theta_a) &\leq \frac{1-p}{p}Z_T^S\tau h^{Sk*}(\theta_a)\left[1-h^{Sk*}(\theta_a)\right]c^k(\theta_a) \\ &\leq \frac{1-p}{4p}Z_T^S\tau c^k(\theta_a) \\ &< c^k(\theta_a), \end{aligned}$$

where the second line follows from noting that $h(1-h) \leq \frac{1}{4}$ and the final inequality comes from $Z_T^S \leq Z_T^N$ and the open economy version of Assumption (A1). Therefore, in both cases we must have $c^S(\theta_a) < c^N(\theta_a)$ implying that no training occurs in the South.

**Proof of Proposition 7.** First we will characterize the closed economy steady state for general $\beta$. Let $\Delta \equiv \frac{p}{1-p}\frac{1}{\tau}\frac{1}{Z^\beta}$. Then following the same steps used in Section 2 it is straightforward to show that $\theta_1 = \left(\frac{w}{Z}\right)^{\frac{1}{\alpha}}Z^{1-\beta}$ and $\theta_2 = \frac{\theta_1}{1-\Delta}$. With these expressions for $\theta_1$ and $\Delta$ the labor market clearing condition Eq. (23) is unaltered, $D$ is still given by Eq. (25) and $h^*(\theta)$, $H^*(\theta)$ and $E$ are the same functions of $D$ given in Eq. (24). Therefore, Proposition 2 implies that $\theta_1$ is uniquely determined and is strictly decreasing in $\Delta$. It follows that $\theta_1$ is strictly increasing in $Z$ whenever $\beta > 0$.

Now consider the open economy. Since South is a small country $w^N$, $\theta_1^N$ and $\theta_2^N$ take the same values as in autarky. All that remains is to solve for $w^S$. We cannot have $\frac{w^S}{Z^S} < \frac{w^N}{Z^N}$ because if the efficiency wage is lower in the South all managers who do not give training strictly prefer to locate in the South and this violates labor market clearing. Given $\frac{w^S}{Z^S} \geq \frac{w^N}{Z^N}$ Eq. (21) implies that if $\beta \geq 1$ then $c^N(\theta) > c^S(\theta)\,\forall\theta \geq \theta_1^N$ meaning that there will never be any training in the South since the North has both a higher training price and a weakly lower labor cost.

Suppose $0 < \beta < 1$. Using Eq. (21) we can show that if $\frac{w^S}{Z^S} = \frac{w^N}{Z^N}$ then $\frac{c^N(\theta)}{c^S(\theta)}$ is strictly increasing in $\theta$. Therefore, a sufficient condition for $\frac{w^S}{Z^S} = \frac{w^N}{Z^N}$ to hold in steady state is that $\theta_2^N \leq \theta_2^S$. This restriction ensures that all training takes place in the North allowing the Southern labor market to clear. Assuming $\theta_2^S < \theta_2^N$ we can solve for $\theta_2^S$ and compare the result to $\theta_2^N$ obtaining:

$$\begin{aligned} \frac{\theta_2^N}{\theta_2^S} &= \frac{Z^N}{Z^S}\frac{(1-p)\tau\left(Z^S\right)^\beta - p}{(1-p)\tau\left(Z^N\right)^\beta - p}, \\ &\equiv B. \end{aligned}$$

Observe: (i) as $(1-p)\tau\left(Z^S\right)^\beta - p \to 0, B \to 0$ and as $Z^S \to Z^N$, $B \to 1$; (ii) $B$ is strictly increasing in $Z^S$ if $Z^S < \left(\frac{p}{1-p}\frac{1}{\tau}\frac{1}{1-\beta}\right)^{\frac{1}{\beta}} \equiv A(\beta)$ and is strictly decreasing in $Z^S$ if $Z^S > A(\beta)$, and; (iii) $A(\beta)$ is unbounded above as $\beta \to 1$. Now define $\beta^*$ to be the smallest value of $\beta \in (0,1)$ such that $Z^N \leq A(\beta^*)\,\forall\beta' \geq \beta^*$ and let $Z^*(\beta)$ be the smallest positive value of $Z^S$ such that $B = 1$. The observations above ensure that $\beta^*$ exists and that $Z^*(\beta) < A(\beta)$ whenever $\beta < \beta^*$. If either $\beta \geq \beta^*$, or $\beta < \beta^*$ and $Z^S \leq Z^*(\beta)$, then $B \leq 1$ implying that $\theta_2^N \leq \theta_2^S$ and there is no training in the South in steady state.

If $\beta < \beta^*$ and $Z^S > Z^*(\beta)$ then $B > 1$ and $\theta_2^S < \theta_2^N$. Consequently, managers with skill $\theta \in (\theta_2^S, \theta_2^N)$ strictly prefer to locate in the South. Now if $\theta_2^S < \bar{\theta}$, which is guaranteed whenever $\theta_2^N = \theta_2^{N,A} < \bar{\theta}$ this contradicts labor market clearing in the South. Therefore, we must have $\frac{w^S}{Z^S} > \frac{w^N}{Z^N}$. It is possible that all Southern agents choose to become managers,[35] but if there is a non-zero labor supply in the South

[35] This possibility can be ruled out by setting $\underline{\theta} = 0$.

the only way to obtain labor market clearing in the South is if some managers provide training in the South.

**Proof of Proposition 8.** First, I will prove that in each country there exists a threshold learning capacity such that workers receive training if and only if their learning capacity exceeds the threshold. Suppose this is not the case. Then there exists a worker with learning capacity $\phi$ who does not receive training and a worker with learning capacity $\phi' < \phi$ who does receive training. Agents accept employment with training only if it is weakly preferred to employment without training. Consequently, training cannot reduce worker welfare implying that $V^W\left(\phi',Z\right) \geq \frac{w(Z)}{p} = V^W(\phi,Z)$. Then from Eq. (22) we have:

$$\begin{aligned} c(\theta;\phi,Z) - c\left(\theta;\phi',Z\right) &= \frac{1-p}{p}\left(\phi-\phi'\right)[Z\theta\tilde{\pi}(\theta)-w(Z)] \\ &\quad + \phi'\left[V^W\left(\phi',Z\right)-\frac{w(Z)}{p}\right], \\ &> 0. \end{aligned}$$

This implication is inconsistent with managerial profit maximization. Hence, it cannot be the case that the $\phi'$ worker receives training, but the worker with learning capacity $\phi$ does not. The result follows.

To prove the second part of the proposition, I will again use a proof by contradiction. Suppose positive assortative matching does not occur. Then there exists a pair of trainees indexed by $i = a,b$, where trainee $i$ comes from an efficiency $Z_i$ country, has learning capacity $\phi_i$ and is employed by a skill $\theta_i$ manager, such that $\theta_b > \theta_a$, but $\phi_a Z_a \equiv \chi_a > \chi_b \equiv \phi_b Z_b$. Since the efficiency wage is constant across countries by assumption, for managerial profit maximization to hold we must have $c(\theta_a;\phi_a,Z_a) \geq c(\theta_a;\phi_b,Z_b)$. Using Eq. (22) this implies:

$$\begin{aligned} &\frac{1-p}{p}\theta_a\tilde{\pi}(\theta_a)(\chi_a-\chi_b)-\phi_a\left[V^W(\phi_a,Z_a)-w(Z_a)\right]+\phi_b\left[V^W(\phi_b,Z_b)-w(Z_b)\right]\geq 0, \\ &\Rightarrow\frac{1-p}{p}\theta_b\tilde{\pi}(\theta_b)(\chi_a-\chi_b)-\phi_a\left[V^W(\phi_a,Z_a)-w(Z_a)\right]+\phi_b\left[V^W(\phi_b,Z_b)-w(Z_b)\right]>0, \\ &\Rightarrow c(\theta_b;\phi_a,Z_a)>c(\theta_b;\phi_b,Z_b). \end{aligned}$$

This contradicts profit maximization by managers with skill $\theta_b$. Thus, there must be positive assortative matching.

## References


Acemoglu, Daron, 1997. Training and innovation in an imperfect labour market. Review of Economic Studies 64 (3), 445–464 (July).

Acemoglu, Daron, Pischke, Jorn-Steffen, 1998. Why do firms train? Theory and evidence. Quarterly Journal of Economics 113 (1), 79–119 (February).

Aghion, Philippe, Howitt, Peter, Mayer-Foulkes, David, 2005. The effect of financial development on convergence: theory and evidence. Quarterly Journal of Economics 120 (1), 173–222 (February).

Almeida, Rita K., Aterido, Reyes, 2010. Investment in Job Training: Why are SMEs Lagging so Much Behind? The World Bank, Policy Research Working Paper Series.

Antràs, Pol, Garicano, Luis, Rossi-Hansberg, Esteban, 2006. Offshoring in a knowledge economy. Quarterly Journal of Economics 121 (1), 31–77 (February).

Arrow, Kenneth J., 1969. Classificatory notes on the production and transmission of technological knowledge. American Economic Review 59 (2), 29–35 (May).

Balsvik, Ragnhild, 2011. Is labor mobility a channel for spillovers from multinationals? Evidence from Norwegian manufacturing. The Review of Economics and Statistics 93 (1), 285–297 (February).

Beaudry, Paul, Francois, Patrick, 2010. Managerial skills acquisition and the theory of economic development. Review of Economic Studies 77 (1), 90–126 (January).

Becker, Gary S., 1964. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. National Bureau of Economic Research; Columbia University Press, New York.

Bloom, Nicholas, Van Reenen, John, 2010. Why do management practices differ across firms and countries? Journal of Economic Perspectives 24 (1), 203–224 (Winter).

Bloom, Nicholas, Eifert, Benn, Mahajan, Aprajit, McKenzie, David, Roberts, John, 2011. Does Management Matter? Evidence from India. NBER Working Paper.

Burstein, Ariel T., Monge-Naranjo, Alexander, 2009. Foreign know-how, firm control, and the income of developing countries. Quarterly Journal of Economics 124 (1), 149–195 (February).

Cabral, Luis, Wang, Zhu, 2008. Spin-offs: Theory and Evidence. Federal Reserve Bank of Kansas City, Research Working Paper.

Caselli, Francesco, 2005. Accounting for Cross-country Income Differences. In: Aghion, Phillipe, Durlauf, Steven (Eds.), The Handbook of Economic Growth, vol. 1. Elsevier, Amsterdam.

Chatterjee, Satyajit, Rossi-Hansberg, Esteban, 2012. Spinoffs and the market for ideas. International Economic Review 53 (1), 53–93 (February).

Comin, Diego, Hobijn, Bart, 2009. Lobbies and technology diffusion. The Review of Economics and Statistics 91 (2), 229–244 (May).

Dasgupta, Kunal, 2012. Learning and knowledge diffusion in a global economy. Journal of International Economics 87 (2), 323–336 (July).

Fosfuri, Andrea, Motta, Massimo, Ronde, Thomas, 2001. Foreign direct investment and spillovers through workers' mobility. Journal of International Economics 53 (1), 205–222 (February).

Franco, April Mitchell, Filson, Darren, 2006. Spin-outs: knowledge diffusion through employee mobility. The RAND Journal of Economics 37 (4), 841–860 (Winter).

Gibson, John, McKenzie, David, 2011. Eight Questions About Brain Drain. The World Bank, Policy Research Working Paper Series.

Glass, Amy Jocelyn, Saggi, Kamal, 2002. Multinational firms and technology transfer. The Scandinavian Journal of Economics 104 (4), 495–513 (December).

Görg, Holger, Strobl, Eric, 2005. Spillovers from foreign firms through worker mobility: an empirical investigation. The Scandinavian Journal of Economics 107 (4), 693–709 (December).

Hausmann, Ricardo, Rodrik, Dani, 2003. Economic development as self-discovery. Journal of Development Economics 72 (2), 603–633 (December).

Helpman, Elhanan, 1984. A simple theory of international trade with multinational corporations. Journal of Political Economy 92 (3), 451–471 (June).

Helpman, Elhanan, Melitz, Marc J., Yeaple, Stephen R., 2004. Export versus FDI with heterogeneous firms. American Economic Review 94 (1), 300–316 (March).

Howitt, Peter, Mayer-Foulkes, David, 2005. R&D, implementation, and stagnation: a Schumpeterian theory of convergence clubs. Journal of Money, Credit, and Banking 37 (1), 147–177 (February).

Lucas Jr., Robert E., 1978. On the size distribution of business firms. Bell Journal of Economics 9 (2), 508–523 (Autumn).

Malchow-Møller, Nikolaj, Markusen, James R., Schjerning, Bertel, forthcoming. Foreign firms, domestic wages. Scandinavian Journal of Economics.

Markusen, James R., Trofimenko, Natalia, 2009. Teaching locals new tricks: foreign experts as a channel of knowledge transfers. Journal of Development Economics 88 (1), 120–131 (January).

Monge-Naranjo, Alexander, 2011. Foreign firms and the diffusion of knowledge. Penn State University.

Parente, Stephen L., Prescott, Edward C., 1999. Monopoly rights: a barrier to riches. American Economic Review 89 (5), 1216–1233 (December).

Poole, Jennifer, forthcoming. Knowledge transfers from multinational to domestic firms: evidence from worker mobility. The Review of Economics and Statistics.

Rosen, Sherwin, 1982. Authority, control, and the distribution of earnings. Bell Journal of Economics 13 (2), 311–323 (Autumn).