University of Toronto
Department of Economics

Working Paper 364

Learning, Knowledge Diffusion and the Gains from Globalization

By Kunal Dasgupta

July 30, 2009

Learning, Knowledge Diffusion and the Gains from Globalization

Kunal Dasgupta*
Princeton University
December, 2008
(Job Market Paper)

Abstract

We develop a dynamic, general equilibrium model to understand how multinationals affect host countries through knowledge diffusion. Workers learn from their managers and knowledge diffusion takes place through worker mobility. We identify two forces that determine wages: the labour demand effect and the learning effect. The former tends to raise wages while the latter tends to reduce it. We show that in a model without learning, an integrated steady-state equilibrium, in which incumbent host country managers operate alongside multinationals, can never be a Pareto improvement for the host country. In contrast, we present a novel mechanism through which a Pareto improvement occurs in the presence of learning dynamics. We study how integration affects the life time earnings of agents and the degree of inequality in the host country, as well as, analyze the pattern of multinational activity. In the quantitative section of the paper, we calibrate our model to fit key moments from the U.S. wage distribution and quantify gains from integration. Our estimates suggest that learning produces welfare gains that range from 2% for the middle-income countries to 43% for the low-income countries.

KEYWORDS : Multinationals, knowledge diffusion, learning, welfare gains, worker mobility.

1 Introduction

One of the most important asset that a Multinational Enterprise (MNE) brings to a foreign market is its possession of superior knowledge.[1] Naturally, economists are concerned about whether and how this knowledge diffuses to domestically owned firms and the consequence of this diffusion for

*E-mail : kdasgupt@princeton.edu. I am indebted to Esteban Rossi-Hansberg, without whose guidance and support this paper would never have been possible. I am also grateful to Gene Grossman and Marc Melitz for their valuable advice and encouragement. I would also like to thank Alicia Adsera, Pol Antràs, Costas Arkolakis, David Atkin, Thomas Chaney, Cecilia Fieler, Luis Garicano, Nobuhiro Kiyotaki, Marc Muendler, Miklos Koren, Per Krusell, John McLaren, Alexander Monge-Naranjo, Stephen Redding, Daniel Trefler, Vinayak Tripathi, Jon Vogel and workshop participants in Princeton University and the Federal Reserve Bank of Philadelphia for their comments and suggestions. Financial support from the International Economics Section, Princeton University is greatly appreciated.

[1] This is a widely held belief in the business literature. See Kogut and Zander (1993).

the firms, the workers and the economy as a whole (Lipsey and Sjöholm (2005), Blomström and Kokko (1998)).

Existing studies of MNEs have focused on different channels for knowledge diffusion.[2] One such channel, that has received relatively less attention, is diffusion through worker mobility. There is evidence that MNEs provide more and better training to its workers compared to domestic firms (Gershenberg (1987)).[3] MNEs provide domestic workers with access to a vast pool of advanced knowledge. By learning from MNEs, these workers increase their productive capabilities. Some of them even go on to start their own firms. Giarratana *et al* (2003) talk about the spin-offs from MNEs that were created in India after the country liberalized in 1991. Based on interviews conducted with the founders of some of these spin-offs, Giarratana *et al* (2003) conclude that the founders bring a high-level of technological expertise from the MNEs to the new firms.

Moreover, in many developing countries, the educational system is not geared towards meeting the needs of the industry. In these countries, foreign MNEs provide the domestic workers with the opportunity to acquire marketable skills. For example, in China, potential managers think of the MNEs as schools where they can train themselves, and then, once they have the required expertise to start their own business, leave (The Economist, 2005). This could partly explain the high turnover rates that the county has witnessed in recent years.[4]

Knowledge acquisition by the workers has an impact both on their life time earnings, as well as, the productivity and profitability of firms which hire them. At the same time, as workers learn, the entire knowledge distribution of the host-country changes, which in turn has aggregate implications. In this paper, we build a model which not only allows us to study the impact of MNE entry on welfare and earning dynamics at the individual level, but also the impact on inequality and the pattern of multinational activity at the aggregate level.

We develop a dynamic, general equilibrium model, in which agents have different levels of knowledge. Our model adds learning to a framework that is similar to Antràs, Garicano and Rossi-Hansberg (2006). In our model, production is carried out in firms by workers who are supervised by a manager. Agents make two decisions: (1) whether to be a worker or a manager and (2) who to match with, i.e. a manager chooses his workers while a worker chooses for whom to work. The production technology exhibits complementarity.[5] Agents are endowed with knowledge at birth, but can acquire more by working in firms. We assume that workers learn on the job. Learning is stochastic and depends both on the worker's knowledge and the manager's knowledge. In particular,

[2] Rodríguez-Clare (1996) focuses on the impact of multinationals on developing countries through the generation of backward and forward linkages. In Markusen and Venables (1999), the effect on host country firms depends on the relative strengths of competition and linkage effects.

[3] Görg et al (2007) find that workers who are trained in subsidiaries of multinational firms have a steeper wage gradient compared to workers who receive training in local firms. They take this as evidence that foreign subsidiaries provide more effective training to workers.

[4] The Economist reports that employee turnover rates have gone up from 8.3% in 2001 to 11.3% in 2004. See "China's people problem", *The Economist*, 14th April, 2005.

[5] This means that the marginal productivity of an agent is enhanced if he matches with more knowledgeable agents.

expected learning is an increasing function of both the worker's and the manager's knowledge. Thus, there is also 'complementarity' in learning.

In our model, when an agent is born, he draws his knowledge from an exogenously given distribution. Agents also face a constant probability of death every period. Apart from the exogenous entry and exit of agents, there is endogenous movement of agents within the knowledge distribution due to learning. We show that the complementarity of the production and learning technologies leads to positive assortative matching (PAM), whereby more knowledgeable workers team with more knowledgeable managers to produce and learn. The equilibrium is characterized by a threshold such that every agent below the threshold is a worker and those above are managers. However, the combination of PAM and learning implies that every agent who starts his life as a worker, works for better and better managers till he himself becomes a manager, provided that he survives long enough. Thus, agents move up the knowledge ladder and their earnings increase over their lifetime. Birth, learning and death determine the evolution of the knowledge distribution.

Our initial analysis focuses on a country in autarky. Next, we allow the Home country (or the host country) to integrate with a Foreign country (or the source country) that has a different knowledge distribution.[6] In particular, the Foreign newborn distribution dominates the Home newborn distribution in the sense of first-order stochastic dominance. This is a way to capture the assumption that the Foreign country has relatively more knowledgeable agents compared to the Home country. Integration, in our model, means that managers are able to form international teams.[7] At the same time, managers are rival factors.[8] To understand the impact of integration on the welfare of host-country agents, we first consider a version of the model where agents do not learn. Integration leads to a re-adjustment among host country managers. The least knowledgeable among them exit, whereas among those who persist as managers, some are now matched with less knowledgeable workers. Under this situation, we show that some agents are necessarily worse-off compared to autarky. That is, we show that in the absence of learning, integration can never be a Pareto improvement in an equilibrium where some incumbent host country managers operate.

Introducing learning complicates the analysis because the knowledge distribution becomes endogenous. We identify two effects that determine wages. Upon integration, the wage schedule is pushed up by the competition for workers coming from foreign MNEs. We call this the labour demand effect. But now there is a second effect on wages. The entry of MNEs creates the possibility for the workers to be matched with more knowledgeable managers. By working for the MNEs, workers can learn more and earn more than under autarky. The result that MNEs hire the more knowledgeable workers however implies that the less knowledgeable workers can expect to work for

[6] Throughout the paper, we use the terms "globalization", "integration" and "opening up" interchangeably.

[7] In this paper, MNEs are synonymous with international production teams. We abstract from the issue related to the boundaries of international firms. For some recent papers which deal with this issue, see Antrás (2003), Antrás and Helpman (2004), and Grossman and Helpman (2003).

[8] Whether managers travel from the source-country to the host-country or not is irrelevant.

the MNEs in the future only if they learn from their current managers. Thus learning creates a rent. The managers extract a part of this rent by paying lower wages and thereby internalize the knowledge "spillover". We call this the learning effect. We show that if learning is fast enough, this effect dominates and the wage schedule shifts down enough so as to make the incumbent managers better off. The workers are better off too, because the increase in their continuation value outweighs the reduction in current wage. We believe that this is a novel mechanism through which integration, in the presence of learning, can lead to Pareto improvement.[9]

We analyze the model in more detail in the numerical section. Our model allows us to study the evolution of individual earnings over the lifetime. By improving the matches, integration increases the amount of knowledge that agents can acquire in each period. This increases the gradient of the lifetime earnings function. For slow learning, integration also reduces consumption inequality. Inequality, however rises if agents are learning faster. In this case, integration amplifies the initial inequality in the Home country. Our model also sheds light on the pattern of multinational activity. First, domestic firms and MNEs coexist. Furthermore, depending on the relative endowments of knowledge in the two countries, we may observe multinational activity in both directions.

In the quantitative section, we calibrate the closed-economy model to match key moments of the U.S. wage distribution. Then we ask the following question : How much do the developing countries gain by moving from autarky to frictionless integration with the average developed country? We focus on bilateral integration rather than multilateral integration since we want to show how the host country benefits from integrating with a country which, in some sense, has a relatively greater endowment of knowledgeable agents. Using parameter values obtained from the calibration, we find welfare gains that range from two percent for the richest middle-income countries to almost forty-three percent for the poorest countries. Most of these gains can be attributed to learning, rather than efficient allocation of managerial talent across countries.

Our paper is related to Monge-Naranjo (2007) who also studies the impact of MNEs on the domestic accumulation of skills. He develops a two period overlapping generations model where young agents can either be workers or potential managers. The latter can acquire knowledge and become managers when they are old. When an economy opens up, foreign entrepreneurs, who have higher knowledge than their domestic counterparts, relocate and carry out production with local workers. In his model, potential managers and entrepreneurs are homogenous, and consequently, every potential manager learns the same. In contrast, workers and managers in our model are heterogenous and learning is specific to the worker-manager pair. This assumption, by generating a non-trivial knowledge distribution, allows us to study the impact of integration on the distribution of consumption, the dynamics of individual earnings and the pattern of multinational activity.[10]

[9] See Fosfuri *et al* (2001) and Glass and Saggi (2002) for game-theoritic treatments of knowledge diffusion through worker mobility.

[10] For quantitative models that compute static welfare gains associated with multinational production see Ramondo (2008), Garetto (2008) and Burstein and Monge-Naranjo (2008). Rodríguez-Clare (2007) develops a model of trade

On the empirical side, Poole (2006) uses matched employer-employee data from Brazil to investigate whether knowledge spillovers occur through worker mobility. She finds that (1) higher-skilled former MNE workers are better able to convey knowledge while higher-skilled incumbent domestic workers are better able to absorb knowledge and (2) incumbent production workers learn more from former MNE workers or managers. Her findings form the basis of our assumption about how workers learn within firms.

The remainder of the paper proceeds as follows. In Section 2, we describe the model while in Section 3, we study the properties of a stationary equilibrium. In Section 4, we analyze how integration affects welfare in the host country. We study a numerical example in Section 5 and use it to further characterize the equilibrium. Section 6 discusses the calibration and the quantitative results. Section 7 concludes. All the proofs are in the Appendix.

2 The Model

Our model introduces learning and dynamics to a framework that is similar to Antràs, Garicano and Rossi-Hansberg (2006), henceforth AGR. There is a continuum of heterogeneous agents who vary in terms of their knowledge. Knowledge is embodied in an agent, but can be acquired through interactions, i.e. an agent can learn from others. Knowledge includes any quality that enhances productivity, for example, management skills and experience. One can think of knowledge as some composite of different attributes that affects an agent's productive capability.[11]

Time is discrete and the time horizon is infinite. Every period, agents are born. A newborn agent draws his knowledge from an exogenously given distribution $\Phi(k)$ with support $[\underline{k}, \overline{k}]$ ($\phi(k)$ being the corresponding density). At the same time, each agent dies in any period with a constant probability δ. The actual knowledge distribution at time t, $\Psi_t(k)$ ($\psi_t(k)$ being the corresponding density) as well as the invariant distribution, will, of course, depend on $\Phi(k)$.[12] We assume that knowledge is perfectly observable and thereby abstract from asymmetric informational issues. Agents are risk-neutral.

Production : Production of a single, non-storable good is carried out in firms. We call this good GDP. A firm comprises of a manager and production workers. Production workers do routine jobs and each production worker combines with the manager to produce $f(y)$ units of output, where y is the knowledge of the manager. Thus, *"f(y) captures the indivisibility of management-type decisions and implies a scale economy because it improves productivity of all the workers in the firm, irrespective of their numbers"* as in Rosen (1982, p. 314). Notice that the productivity of workers in a firm run by a manager with knowledge y is simply $f(y)$. The manager pays wages to

and diffusion where growth is caused by technological progress. Unlike our model, however, diffusion of ideas is an exogenous process.

[11] For a trade model where agents have two attributes, see Ohnsorge and Trefler (2007).

[12] In the absence of learning, any initial distribution will ultimately converge to $\Phi(k)$.

the workers and is the residual claimant of the output.[13]

The span of control of a manager depends only on the knowledge of the workers he hires. The span of control is given by $n(x;\beta)$, where x is the knowledge of the worker.[14][15] For a given x, β measures the span of control with $\frac{\partial n}{\partial \beta} > 0$. Henceforth, we suppress the dependence of n on β and introduce it only when necessary. Apart from the knowledge of the worker and the manager, output also depends on local conditions like government policies, infrastructure, political stability, etc.[16] We denote local conditions by σ. A firm faces the σ of the country in which it is producing. Total output of a firm is then given by

$$q = \sigma f(y) n(x) \tag{1}$$

We make the following assumptions regarding technology -

ASSUMPTION 1a : $f'(y) > 0, f''(y) \geq 0, n'(x) > 0, n''(x) < 0$.

ASSUMPTION 1b : $\frac{\partial}{\partial y}[\frac{f'(y)}{f(y)}] \leq 0$.

ASSUMPTION 1c : $\frac{f'(\underline{k})}{f(\overline{k})} > \frac{n'(\underline{k})}{n(\underline{k})}$.

Assumptions 1a, 1b and 1c together imply that output is more sensitive to the knowledge of the manager relative to that of the worker. Assumption 1c also says that for a given knowledge distribution, there should be sufficient asymmetry between the manager and the worker's contribution to output.[17] This is a sufficient condition for the existence and uniqueness of equilibrium.[18]

Learning : Agents also learn in firms.[19] Since the seminal work of Gary Becker (Becker (1962)), economists have been studying on-the-job training. In this paper, we abstract from formal training provided by firms and instead focus on the knowledge that workers acquire simply by associating

[13] Here, as in Monge-Naranjo (2007), we assume that there is no difference between the managers and entrepreneurs. For a model which make this distinction, see Holmes and Schmitz (1990).

[14] Managers could potentially choose different types of workers. However measure consistency implies that a manager can never be matched with an interval of workers. Given the technology, it can also be shown that in equilibrium, the manager would not want to hire more than one type of worker. See AGR for the proofs.

[15] For a micro-foundation of such a technology, see Garicano (2000).

[16] Our assumption that labour is the only factor of production is without loss of generality. We can always introduce capital. The cost of capital usually has three components - sunk cost, fixed cost and variable cost. In the absence of uncertainty in production and credit market imperfections, the first two do not really have any effect. So we just normalize those to zero. As for variable capital requirement, we think of it as being subsumed in $f(y)$.

[17] To see this, note that the output elasticity of the manager's knowledge is $\frac{\partial q}{\partial y}\frac{y}{q} = \frac{f'(y)y}{f(y)}$, while that of the worker is $\frac{n'(x)x}{n(x)}$. Assumption 1b says that $\frac{f'(y)}{f(y)}$ is non-increasing in y while Assumption 1a implies that the same is true for $\frac{n'(x)}{n(x)}$. Moreover, $\frac{f'(\overline{k})}{f(\overline{k})} > \frac{f'(\overline{k})}{f(\overline{k})}\frac{f'(\underline{k})}{f'(\overline{k})} > \frac{n'(\underline{k})}{n(\underline{k})}$. This inequality, combined with the previous observation, implies that $\frac{f'(k)k}{f(k)} > \frac{n'(k)k}{n(k)}$ $\forall k$.

[18] If we relax this assumption, proving uniqueness becomes very difficult.

[19] As pointed out by Rosen (1972, p. 326), "education is not produced only in schools and learning does not cease after graduation.......Rather learning and work are complementary..........In fact, learning in the workplace is extremely widespread and characterizes almost all labour market activities."

with the manager. We follow Jovanovic and Rob (1989) in defining our learning technology. Within each firm, the worker learns from the manager.[20] Learning is stochastic and depends both on the knowledge of the manager and the worker. The randomness in learning does not necessarily reflect any randomness in the knowledge transfer process but rather, is a simplistic way of modelling the heterogeneity in the capacity to absorb knowledge. A worker with knowledge x at time t has knowledge x' at time $t+1$. The learning distribution is given by $\mathcal{L}(x'|x,y)$, $x' \in [x,y]$. We make the following assumptions about the learning technology:

ASSUMPTION 2a : $\frac{\partial}{\partial x}[\int h(x')d\ \mathcal{L}(x'|x,y)] > 0 \quad \forall\ h(x')$ increasing in x'.

ASSUMPTION 2b : $\frac{\partial}{\partial y}[\int h(x')d\ \mathcal{L}(x'|x,y)] > 0 \ \ \forall\ h(x')$ increasing in x'.

The above conditions are the familiar ones for first-order stochastic dominance.[21] These conditions imply that expected learning is increasing in the knowledge of both the workers and the managers.

Agent's problem : Every agent is a price-taker. There are two prices in the economy. First, the managers hire workers and pay a price for their marginal product. Second, the workers learn from the managers and pay a price for the acquired knowledge. It is inconsequential whether there are two transactions within the firm or whether the managers simply pay the wage net of the rent (on this point, see Rosen(1972)). What matters is the net payment to workers, and hence we focus on *net wages* w_t.

Given a sequence of wage functions $\{w_t\}_{t=0}^{\infty}$, the manager's problem is defined recursively as

$$V_M(y,w_t) = \sup_x\{\sigma f(y)n(x) - w_t(x)n(x) + (1-\delta)\max[V_W(y,w_{t+1}), V_M(y,w_{t+1})]\}. \qquad (2)$$

where $V_W(y,w_t)$ is the value function of an agent with knowledge y, if he chooses to be a worker while $V_M(y,w_t)$ is the value function if, instead, he chooses to be a manager.[22] The value of a manager depends on the current distribution $\Psi_t(k)$ through the net wage schedule w_t. That is why w_t is treated as a state variable. The second term on the right captures the fact that an agent, who is a manager at time t, might choose to be a worker at time $t+1$ if the wage schedule changes.

$V_W(y,w_t)$ is given by

$$V_W(x,w_t) = w_t(x) + (1-\delta)\int_x^{m_t(x)} \max[V_W(x',w_{t+1}), V_M(x',w_{t+1})]d\mathcal{L}(x'|x,m_t(x)). \qquad (3)$$

[20] Unlike Jovanovic and Rob (1989), learning is one-sided. Assuming that managers also learn from workers could be an interesting extension and would be one channel through which growth can be introduced in this model.

[21] For a distribution to first-order stochastic dominate another distribution, their supports have to be the same. In this case, $\Phi(x'|x_1,y)$ and $\Phi(x'|x_2,y)$, $x_1 > x_2$, have different support. But we can always think of $\Phi(x'|x_1,y)$ as having support $[x_2,y]$ with zero mass in $[x_2,x_1]$. The same logic applies to $\Phi(x'|x,y_1)$ and $\Phi(x'|x,y_2)$.

[22] Notice the absence of time discounting in the above formulation. This is due to the fact that a positive probability of death acts as a discount factor.

where $m_t(x)$ is the knowledge of the manager who hires a worker with knowledge x. The term within the integral denotes the expected value of the worker if he works for a manager with knowledge $m_t(x)$. Depending on how much he learns, the worker might become a manager or continue as a worker at $t+1$. As before, this decision will depend not only on the worker's own knowledge but also on the market wage schedule. The following two lemmas establish existence and some properties of the value functions.

Lemma 1 $V_M(y, w)$ exists, and is continuous and strictly increasing in y.

Lemma 2 $V_W(x, w)$ exists, and is continuous and strictly increasing in x.

Matching : Notice that the problem of the manager is essentially static since he does not learn. Therefore the manager chooses his workers in order to maximize his current profits

$$\pi_t(y) = \sigma f(y)n(x) - w_t(x)n(x).$$

The first-order condition (FOC) for the manager's problem is

$$(\sigma f(y) - w_t(x))n'(x) - w_t'(x)n(x) = 0.$$

Re-arranging, we have $w_t'(x) = \frac{(\sigma f(y) - w_t(x))n'(x)}{n(x)}$. Profit-maximization implies that the numerator is positive. Thus, $w_t'(x) > 0$. It can easily be shown that $w_t(x)$ is continuous.

Totally differentiating the FOC,

$$-[(\sigma f(y) - w_t(x))n''(x) - 2w_t'(x)n'(x) - w_t''(x)n(x)] = \sigma f'(y)n'(x)\frac{dy}{dx}.$$

Profit-maximization implies that the LHS is positive.[23] So is $f'(y)n'(x)$. Therefore $\frac{dy}{dx} > 0$, i.e. more knowledgeable workers will work for more knowledgeable managers. Hence we have Positive Assortative Matching (PAM).[24] PAM implies that $m_t(x)$ exists and $m_t'(x) > 0$.[25]

In equilibrium, there exists a threshold k_t^* such that all agents with knowledge less than k_t^* are workers, while those with knowledge above k_t^* are managers (We formally define equilibrium in the

[23] Profit-maximization implies that the second-order condition is satisfied : $(\sigma f(y) - w_t(x))n''(x) - 2w_t'(x)n'(x) - w_t''(x)n(x) < 0$.

[24] Here it is worth mentioning why we need the technological restriction linking the span of control to the knowledge of the workers. Consider a production function of the form $q = f(y)g(nx)$. The manager chooses both n and x to maximize $\pi = f(y)g(nx) - nw(x)$. The FOC with respect to n gives $f(y)g'(nx)x = w(x)$ while that with respect to x gives $f(y)g'(nx)n = nw'(x)$. Combining we get $w'(x) = \frac{w(x)}{n}$. Thus the wage schedule is linear which means that the workers are perfect substitutes. Hence, as in Lucas (1978), the manager will be indifferent between hiring workers with different levels of knowledge.

[25] The fact that complementarity in production leads to PAM is a standard result in the matching literature.

next section). Under threshold matching, the labour-market clearing condition is given by

$$\int_{\underline{k}}^{k} d\Psi_t(s) = \int_{m_t(\underline{k})}^{m_t(k)} n(m_t^{-1}(s))d\Psi_t(s) \qquad \forall k \leq k_t^* \tag{4}$$

Note that the labour market condition is not standard. The LHS denotes the supply of workers in the interval $[\underline{k}, k]$, while the RHS denotes the demand for workers coming from managers in the interval $[m_t(\underline{k}), m_t(k)]$. Measure consistency requires that these two values must be equal for every k. That is because, the workers hired by managers with knowledge in $[m_t(\underline{k}), m_t(k)]$ must have knowledge in $[\underline{k}, k]$. An implication of PAM is that $m_t(k)$ is monotone increasing. This allows us to invert the matching function. Differentiating eqn (4) with respect to k, we have

$$m_t'(k) = \frac{\psi_t(k)}{n(k)\psi_t(m_t(k))} \tag{5}$$

The above differential equation, along with the boundary conditions $m_t(\underline{k}) = k_t^*$ and $m_t(k_t^*) = \overline{k}$, allows us to solve for the matching function. As the following lemma shows, given a $\Psi_t(k)$, the threshold k_t^* and consequently the matching function are uniquely determined.

Lemma 3 Given a $\Psi_t(k)$, k_t^* exists and is unique.

The above discussion suggests that the matches can be determined completely once we know the knowledge distribution; we do not need to know the wage schedule in order to determine the matches. Rather, once the matches are determined, wages adjust so as to support the matches that emerge. Of course, this does not mean that how agents match does not depend on wages. In this economy, wages (and profits) not only determine the remuneration of the agents but they also play an allocative role (Sattinger(1993)). But for the purpose of solving the model, we can derive the matching function without any information on the wage function.

Dynamics : When workers work for managers, they learn. Accordingly, their knowledge increases over time. But agents also die every period with probability δ and are replaced by newborns who draw knowledge from the exogenous distribution $\Phi(k)$. Birth, learning and death implies a rule for the evolution of the knowledge distribution $\Psi_t(k)$:

$$\Psi_{t+1}(k) = \delta\Phi(k) + (1-\delta)\int_{\underline{k}}^{k}\int_{s}^{k} d\mathcal{L}(s'|s, m_t(s))d\Psi_t(s) \text{ for all } k \in [\underline{k}, \overline{k}] \tag{6}$$

The first term on the RHS denotes the fraction of agents who are born in period $t+1$ with knowledge less than k. The second term denotes the agents from period t who remain below k despite learning from their managers. There is another way of looking at the evolution. Let A be any Borel set of $[\underline{k}, \overline{k}]$. Then the transition function for the knowledge distribution satisfies, for every $k \in [\underline{k}, \overline{k}]$

$$P_t(k, A) = \begin{cases} (1-\delta)\int\limits_A d\mathcal{L}(s|k, m_t(k)) + \delta \int\limits_A d\Phi(s) & \text{if } k \leq k_t^* \\ \delta \int\limits_A d\Phi(s) & \text{otherwise} \end{cases}$$

Equation (6) implies that Ψ_{t+1} is determined by how individuals acquire knowledge in period t, and the acquisition of knowledge by individuals is determined only by who they match with at time t, which in turn depends only on Ψ_t. Therefore, Ψ_{t+1} is a function of Ψ_t. We seek a fixed point of Ψ_t, i.e. an invariant knowledge distribution Ψ^*. As the following proposition shows, such a fixed point exists and is unique.

Proposition 1 A unique, invariant knowledge distribution Ψ^* exists and any initial distribution Ψ_0 weakly converges to it.

Therefore, in the long run, the knowledge distribution converges to Ψ^*, with threshold k^*. Agents, who are born with knowledge above k^*, become managers instantaneously. Since managers do not learn, these agents are stuck with the level of knowledge they are born with. On the other hand, agents, who are born with knowledge below k^*, start their lives as workers. These agents learn in every period and move up, till they eventually cross the threshold and become managers themselves. For these agents, the lifetime earnings profiles are positively sloped.

3 Equilibrium

3.1 Autarky

Now, we formally define an equilibrium of this model. A competitive equilibrium of this economy consists of :

1. value functions $V_W(x, w_t) : [\underline{k}, k_t^*] \to \mathbb{R}$ and $V_M(y, w_t) : [k_t^*, \overline{k}] \to \mathbb{R}$;
2. a matching function $m_t(x) : [\underline{k}, k_t^*] \to [k_t^*, \overline{k}]$;
3. prices $w_t(x) : [\underline{k}, k_t^*] \to \mathbb{R}$ and $\pi_t(y) : [k_t^*, \overline{k}] \to \mathbb{R}$; and
4. an occupational choice structure : workers $\in [\underline{k}, k_t^*]$ and managers $\in [k_t^*, \overline{k}]$

such that

(a) $V_W(x, w_t)$ and $V_M(y, w_t)$ satisfy the worker's and manager's problems respectively ($V_W(k_t^*, w_t) = V_M(k_t^*, w_t)$ and $V'_W(k_t^*, w_t) \leq V'_M(k_t^*, w_t)$);[26]

[26] If this condition is not satisfied, then $\overline{k}$ can hire $k_t^* + \epsilon$ and both of them can be made strictly better-off. But then k_t^* can not be an equilibrium. The formal proof is in the Appendix.

(b) $m_t(x)$ and $m_t^{-1}(y)$ are the corresponding policy functions;

(c) markets for worker clears, that is equation (4) is satisfied; and

(d) the distribution evolves according to equation (6).

The following proposition provides for the existence and uniqueness of the equilibrium.

Proposition 2 There exists a δ^*, such that $\forall\delta > \delta^*$, an equilibrium exists and it is unique. Moreover, this equilibrium is efficient.

A stationary equilibrium of this model consists of value functions, a matching function, prices and an occupational choice (all independent of time), such that workers and managers are maximizing welfare, labour market clears and $\Psi_{t+1} = \Psi_t = \Psi^*$. Henceforth, we focus on the stationary distribution.

It might seem natural to write the wage as $w(x, y)$ given that (a) the same manager can produce different levels of output by hiring different types of workers and (b) the same worker can acquire different levels of knowledge by working for different managers. (a) suggests that the price for labour should be specific to a worker-manager pair while (b) suggests that the same should be true for the price of knowledge. In order to understand why $w(x)$ only has the knowledge of the worker as its argument, we look at the underlying mechanism by which these objects are determined.

First, we consider an economy without learning. In this economy, every agent with knowledge y, in the role of a manager, offers a wage schedule $\widetilde{w}_{NL}(x, y)$. This wage schedule is such that y *is indifferent between hiring any* x. Denote the corresponding profit of y as $\widetilde{\pi}(y)$. The following lemma establishes some properties of $\widetilde{w}_{NL}(x, y)$.

Lemma 4 $\widetilde{w}_{NL}(x, y)$ is continuous and increasing in x for all y.

At the same time, each agent, in the role of a worker, faces a series of wages offered by the other agents. If $\widetilde{\pi}(x) > \max_y \widetilde{w}_{NL}(x, y)$, agent x becomes a manager. Otherwise, he becomes a worker and works for $y(x)$ where $y(x) = \arg\max \widetilde{w}_{NL}(x, y)$. Hence, the occupation of agents is endogenously determined. Equilibrium is obtained when every agent is employed and is maximizing utility. The equilibrium wage schedule is given by $\widetilde{w}_{NL}(x, y(x)) \equiv w_{NL}(x)$, i.e. $w_{NL}(x)$ is the upper envelope of the individual wage schedules.[27]

Consider what happens when we introduce learning. Now, the worker acquires knowledge from the manager, which raises the former's continuation value. Thus, learning creates a rent and

[27] To understand why this is an equilibrium, consider an arbitrary worker with knowledge x^*. By the definition of $w_{NL}(x)$, he maximizes his utility by getting $w_{NL}(x^*)$. Let the manager who offers this wage be denoted by y^*. To

the manager tries to extract a part of this rent by paying a wage that is lower than what he would have paid in the absence of learning. If $\widetilde{w}(x,y)$ is the wage under learning then $\widetilde{w}(x,y) = \widetilde{w}_{NL}(x,y) - C(x,y)$.[28] Each manager now offers a pair $\{\widetilde{w}(x,y), k(x,y)\}$ where $k(x,y)$ is the expected knowledge that a worker x can acquire by working for manager y.[29] That is, each manager effectively offers a $\widetilde{V}_W(x,y)$ schedule since the current wage and expected future knowledge level determines the present value of the workers. The equilibrium $V_W(x, w_t)$ schedule is the upper envelope of the individual $\widetilde{V}_W(.)s$. The $V_W(x, w_t)$ and $w_t(x)$ schedules are related such that the manager who maximizes the present value of earnings of x, also offers the highest wage to x. This is stated formally in the next lemma.

Lemma 5 If y^* is the solution to $\arg\max_y \widetilde{w}(x,y)$, then it also solves $\arg\max_y \widetilde{V}_W(x,y)$.

Lemma 5 implies that if worker x maximizes the profit of manager y, then it must be the case that manager y also maximizes the welfare of worker x. Before proceeding any further, we make the following assumption about the learning technology :

ASSUMPTION 2c : If, in period t, a worker with knowledge x works for a manager with knowledge y, then in period $t+1$, the worker has knowledge x with probability θ and knowledge y with probability $1-\theta$.

Thus, learning is an all-or-nothing proposition for the worker. This assumption gives us analytical tractability. In the next section, we relax this assumption and work with a more general learning technology. Notice that, in spite of the learning distribution having just two points, it still satisfies assumptions 2a and 2b.[30] Recall that the density function for the newborn distribution is given by $\phi(k)$. The learning technology, along with the newborn distribution, implicitly defines the invariant distribution and allows us to solve for the threshold k^*.

---

see why y^* is also maximizing his utility, note that

$$\begin{aligned} \pi(y^*) &= f(y^*)n(x^*) - w_{NL}(x^*)n(x^*) \\ &= f(y^*)n(x^*) - \widetilde{w}_{NL}(x^*, y^*)n(x^*) \\ &= f(y^*)n(x) - \widetilde{w}_{NL}(x, y^*)n(x) \quad \forall x \\ &\geq f(y^*)n(x) - w_{NL}(x)n(x) \end{aligned}$$

where the third line follows from the definition of $\widetilde{w}_{NL}(.)$ and the last line follows from the fact that $w_{NL}(.)$ is the upper envelope of the individual $\widetilde{w}_{NL}(.)$ schedules. Therefore, y^* maximizes his utility by hiring x^*. Since, x^* was chosen arbitrarily, the result follows. Note that $w_{NL}(x)$ increasing in x and continuous (follows from the properties of $\widetilde{w}_{NL}(x,y)$).

[28] Note that the wage schedule offered by the manager is such that the manager is indifferent between hiring any worker. Consequently, the wage schedule must shift down.

[29] This is similar to Boyd and Prescott (1987), where the old workers in a firm offer the young a package of current consumption and future expertise.

[30] To see this, note that for any $h(.), h' > 0$, the expected value of $h(x)$ is $\theta h(x) + (1-\theta)h(y)$. This expression is increasing in both x and y.

Proposition 3 k^* is defined implicitly by the following equation

$$\int_{\underline{k}}^{k^*} \frac{\phi(k)}{n(k;\beta)} dk = \int_{k^*}^{\overline{k}} \phi(k)dk + (1-\theta)(\frac{1-\delta}{\delta}).$$

Moreover, k^* has the following properties :

$$\frac{\partial k^*}{\partial \theta} < 0 \ ; \ \frac{\partial k^*}{\partial \delta} < 0 \ ; \ \frac{\partial k^*}{\partial \beta} > 0.$$

The above expression sheds light on how the distribution changes as the rate of learning increases. δ, being the probability of death in a period, proxies for the length of a time period. A lower δ, holding θ unchanged, implies that agents are acquiring the same expected knowledge over a smaller interval of time. On the other hand, a lower θ, holding δ unchanged, implies that agents are acquiring more expected knowledge over the same interval of time. Both these cases translate into faster learning for the agents. As the rate of learning increases, the knowledge distribution becomes negatively-skewed as more and more mass is shifted to the upper tail. Consequently, labour market clearing requires that the threshold shift to the right. The threshold also rises with an increase in the span of control (higher β). Intuitively, a greater span of control implies that fewer managers are required to employ the workers.

Recall that a worker with knowledge k produces $f(m(k))$ units of output. Hence, total output produced in this economy is given by

$$Y = \int_{\underline{k}}^{k^*} f(m(k))d\Psi(k) \tag{7}$$

Total welfare is given by

$$W = \int_{\underline{k}}^{k^*} V_W(k)d\Psi(k) + \int_{k^*}^{\overline{k}} V_M(k)d\Psi(k) \tag{8}$$

In this model, individual welfare equals the present value of consumption (or income, since the good is non-storable) because agents are risk-neutral.

3.2 International Integration

Next, we allow the host country to integrate with another country. Globalization or international integration, in the context of our model, means that managers from one country can hire workers in another country, i.e. integration leads to the creation of MNEs. However, the managerial input is rival and as a result, managers can not operate plants in both countries. The motive behind the formation of MNEs is exploiting differences in factor prices.[31] In this paper, we focus on full integration, i.e. we assume that MNEs are formed costlessly. In particular, we assume away any cost that might be associated with opening a plant in another country. We do acknowledge that these costs are important, but the introduction of such costs will increase the complexity of the model (see AGR on this issue).

Let us introduce some notation. Define the subscripts $i = \{A, I\}$, $j = \{H, F\}$, where A and I stand for autarky and integration respectively, while H and F stand for Home and Foreign respectively. We label the host country Home. The Home newborn distribution is denoted by $\Phi_H(k)$ with support $[\underline{k}, \overline{k}_H]$, while the Foreign newborn distribution is $\Phi_F(k)$, with $[\underline{k}, \overline{k}_F]$ being the corresponding support. We assume that $\overline{k}_F > \overline{k}_H$ and that $\Phi_F(k)$ first-order stochastic dominates $\Phi_H(k)$. This assumption captures the relative abundance of more knowledgeable agents in the Foreign country. We also assume that $\sigma_F = \sigma_H$, where σ_H and σ_F denote the Home and Foreign country-specific productivity respectively. The steady-state knowledge distributions are indexed by i and j. So, for example, $\Psi_{AH}(k)$ is the Home steady-state knowledge distribution under autarky, $\Psi_I(k)$ is the integrated distribution, and so on. P_H and P_F are the population in the two countries. With integration, the fundamental change is in the distribution of newborns, which is given by

$$\Phi_I(k) = \begin{cases} \frac{P_H}{P_H+P_F}\Phi_H(k) + \frac{P_F}{P_H+P_F}\Phi_F(k) & \text{for } k \in [\underline{k}, \overline{k}_H] \\ \frac{P_H}{P_H+P_F} + \frac{P_F}{P_H+P_F}\Phi_F(k) & \text{for } k \in [\overline{k}_H, \overline{k}_F] \end{cases} \tag{9}$$

$\Phi_I(k)$, combined with the learning technology, determines the integrated knowledge distribution $\Psi_I(k)$. The new threshold, k_I^*, would typically be different from k_A^*, the autarky threshold. In order to derive the relation between the thresholds under autarky and integration, we need the following lemma.

Lemma 6 If a knowledge distribution G first-order stochastic dominates another distribution H, then $k_G^* > k_H^*$, where k_G^* and k_H^* are the thresholds under G and H respectively.

Now, $\Phi_I(k)$ first-order stochastic dominates $\Phi_H(k)$.[32] In the benchmark case of no-learning, the knowledge distributions in both the countries coincide with the newborn distributions. Consequently, under no-learning, $k_I^* > k_A^*$ (this follows directly from Lemma 6). With learning, however, the knowledge distributions are no longer exogenous and this complicates the analysis. Still, we can derive a relation between k_A^* and k_I^*, as shown by the following proposition.

[31] This motive for establishing subsidiaries in other countries is the same as in Helpman (1984).
[32] This follows from eqn (9) and the fact that $\Phi_F(k)$ first-order stochastic dominates $\Phi_H(k)$.

Proposition 4 $k_I^* > k_A^*$ where k_I^* is defined implicitly by the following equation

$$\int_{\underline{k}}^{k_I^*} \frac{p_H \phi_H(k) + p_F \phi_F(k)}{n(k;\beta)} dk = \int_{k_I^*}^{\overline{k}_F} (p_H \phi_H(k) + p_F \phi_F(k)) dk + (1-\theta)(\frac{1-\delta}{\delta})$$

where $p_H = \frac{P_H}{P_H + P_F}$ and $p_F = \frac{P_F}{P_H + P_F}$

Proposition 4 implies that the knowledge range for workers at Home expands under integration. The agents belonging to $[k_A^*, k_I^*]$, who were managers under autarky, decide to become workers under integration. The entry of highly knowledgeable managers from the Foreign country raises the opportunity cost of being a manager for a domestic agent. An incumbent Home manager weighs the cost of becoming a worker for a MNE (forgone current profits) against the benefit (higher expected profits in the future). For the managers in $[k_A^*, k_I^*]$, benefits outweigh costs and consequently they switch.

Although Proposition 4 indicates the direction of change for the threshold, it says nothing about its magnitude. In particular, we could have the following two scenarios :

(I) $k_I^* > \overline{k}_H$: In this case, every agent born in the Home country starts his life as a worker. The support of $\Psi_{IH}(k)$ is $[\underline{k}, \overline{k}_F]$, despite the fact that, the Home newborn distribution $\Phi_H(k)$ still has the smaller support.[33] Though theoretically an interesting case, this situation is quite extreme because it implies that integration results in the destruction of all incumbent firms (managers), who are replaced by a new class of bigger and more productive firms.

(II) $k_I^* < \overline{k}_H$: This situation is characterized by the birth of a new class of Home firms (with knowledge in $[\overline{k}_H, \overline{k}_F]$), who are on par with the Foreign MNEs in terms of size and productivity. But unlike Case I, a set of incumbent Home managers with knowledge in $[k_I^*, \overline{k}_H]$ continues to operate in the integrated economy.

Whether we are in Case I or Case II depends on the parameters of the model. For a given $\overline{k}_H$, there exists a k' such that $\overline{k}_F < k'$ implies that $k_I^* < \overline{k}_H$.[34] Hence, as long as $\overline{k}_F$ is not too different from $\overline{k}_H$, there will be some incumbent managers in the Home country. Intuitively, $\overline{k}_F$ being much greater than $\overline{k}_H$ implies that following integration, the Home agents have an opportunity to work for very knowledgeable managers. This is also true for the incumbent Home managers, who would rather work in Foreign MNEs, learn and become much better managers in the future than remain managers with low levels of knowledge.

[33] It is not the case that every Home agent is a worker. There are Home managers in $[k_I^*, m(\overline{k}_H)]$. This, however, means that the Home managers in the integrated economy have knowledge greater than $\overline{k}_H$.

[34] This follows from the result that k_I^* is monotone increasing in $\overline{k}_F$, and $k_I^* < \overline{k}_H$ when $\overline{k}_F = \overline{k}_H$.

Integration also affects matching. An immediate implication of Proposition 4 is that $m_I(\underline{k}) > m_A(\underline{k})$, where $m_A(.)$ and $m_I(.)$ are the matching functions under autarky and integration respectively.[35] Therefore, the least knowledgeable worker in the Home country, and by continuity, a set of less knowledgeable workers, is matched with better managers. On the other hand, some of the Home managers are now matched with less able workers. Since the output of firms depends positively on workers' knowledge, the output of some of the Home firms under integration are necessarily lower than in autarky.

Corollary 1 Under integration, the output of a positive measure of Home firms goes down.

Note that since the output produced by a worker depends only on the knowledge of the manager he is matched with, the productivity of a firm, as measured by the value-added per worker, does not change.

Total Home output (GDP) produced in the integrated equilibrium is given by

$$Y = \int_{\underline{k}}^{k_I^*} f(m_I(k))d\Psi_{IH}(k)P_H \tag{10}$$

This would be different from Gross National Income (GNI), which is given by

$$GNI = \int_{\underline{k}}^{k_I^*} w_I(k)d\Psi_{IH}(k)P_H + \int_{k_I^*}^{\overline{k}_F} \pi_I(k)d\Psi_{IH}(k)P_H \tag{11}$$

The difference between the two arises from the fact that, in an integrated equilibrium, a part of the Home output goes to the Foreign country as profits of Foreign MNEs, while some of the Home firms may become multinationals and earn profits from their operations in the Foreign country. Aggregate welfare of Home agents is given by

$$W = \int_{\underline{k}}^{k^*} V_W(k)d\Psi_{IH}(k)P_H + \int_{k^*}^{\overline{k}_F} V_M(k)d\Psi_{IH}(k)P_H \tag{12}$$

To sum up, with integration, the threshold of the knowledge distribution shifts to the right. This necessarily means that some of the Home workers work for more knowledgeable managers. These workers also learn more compared to autarky. At the same time, some of the incumbent firms suffer a decline in output. In the next section, we study the impact of integration on the welfare of the Home workers and managers.

[35] To see this, note that $m_A(\underline{k}) = k_A^*$ and $m_I(\underline{k}) = k_I^*$. Proposition 4 then gives the result.

4 Welfare

In this section, we analyze how learning alters the nature of welfare gains for the workers and managers in the Home country. We focus our attention on the case where there are surviving Home managers, i.e. Case II.[36] First we look at the benchmark case of no learning. This is, essentially, the static framework presented in AGR. We ask the interested reader to look at their paper for a detailed and insightful analysis of this case. Shutting down the learning channel simplifies the analysis because the knowledge distributions coincide with the newborn distributions, which are exogenously given. Absence of learning also means that agents can be labelled workers or managers depending on their knowledge at birth.

A result that emerges from AGR is that integration raises aggregate consumption, and with risk-neutral agents, the aggregate welfare of the Home economy. What about individual welfare? In the previous section, we showed that the output produced by the less knowledgeable Home managers goes down under integration.[37] The actual change in profits and welfare, however, depends on the wages they pay. These wages would be different from those under autarky. Of course, as wages change, the welfare of the workers change too. The next proposition shows us how the welfare of Home agents changes following integration.

Proposition 5 In a no-learning world, an integrated steady-state equilibrium with incumbent Home firms can never be a Pareto improvement relative to the autarky steady-state equilibrium in the Home country.

The above proposition tells us that in the absence of learning, integration creates winners and losers. The identity of the winners and losers, though, will depend on the specific parameter values. If we think of workers and managers as two separate factors of production, Proposition 5 essentially gives us a Heckscher-Ohlin like result.[38]

Does Proposition 5 continue to hold when we introduce learning? In order to prove otherwise, we have to show that every agent in the Home economy is strictly better off under integration. Corollary 1 implies that some of the Home managers earn lower revenue compared to autarky.[39] Hence, for them to be better-off under integration, the wage bill has to go down more than revenue.

In this model, there are two forces that determine wages. First, there is a labour demand effect. The entry of Foreign MNEs increases the demand for Home workers. Integration also increases competition faced by the Home workers from their Foreign counterparts. As shown by AGR, (1)

[36] The reason for this is the following : If $\overline{k}_F$ is very different from $\overline{k}_H$, then irrespective of whether agents learn, every Home agent is better off working for the more knowledgable Foreign managers. Thus we get Pareto improvement but Home firms disappear completely.

[37] This is true for both the learning and no-learning case.

[38] To be technically correct, we have infinitely many factors.

[39] Recall that there is only one good. Hence output equals revenue.

if the two countries are not too similar and (2) if the span of control is not too small, the labour demand effect raises the wages of all Home workers.[40]

But there is now a learning effect. In our model, a worker can be hired by any manager with a positive probability. Working for a better manager means higher expected learning and consequently, higher revenue. Hence, the entry of highly knowledgeable Foreign managers raises the continuation value of the Home workers. PAM implies that the most knowledgeable managers hire only the most knowledgeable workers. Therefore, the less knowledgeable workers can hope to work for the MNEs if they learn and acquire enough knowledge. Thus learning creates a rent. Moreover, these workers can learn only from their current managers, some of whom are the incumbent Home managers. This allows the managers to compress the wage. The workers accept this wage reduction because they expect to be compensated in the future (when they become managers). So the learning effect tends to lower the wage schedule. The strength of this effect depends on how fast agents are learning.

The total impact on wages depends on the relative strengths of the two effects. The following Proposition shows us the condition under which the Home economy realizes Pareto gains.

Proposition 6 Under Assumption 2c, the integrated steady-state equilibrium is a Pareto improvement for the Home country if

$$(1-\theta)(\frac{1-\delta}{\delta}) > \frac{2f(\overline{k}_H)[n(\overline{k}_H)-n(\underline{k})]+n(\underline{k})[\mu(\overline{k}_H)f(\overline{k}_H)-\mu(\underline{k})f(\underline{k})]}{\mu(\underline{k})[f(\overline{k}_F)+f(\underline{k})]n(\underline{k})-\mu(\overline{k}_H)[n(\underline{k})+n(\overline{k}_H)]f(\overline{k}_H)}$$

where $\mu(\underline{k}) = \frac{n(\underline{k})}{1+n(\underline{k})}$ and $\mu(\overline{k}_H) = \frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$.

Let us denote by Ω, the set of pairs of θ and δ that satisfy the above condition. The LHS of the above expression is positive by definition. The numerator of the fraction on the RHS is positive too.[41] For Ω to be non-empty, the denominator has to be positive.

We can show that, the denominator is an increasing function of $\frac{f'(k)}{f(k)} - \frac{n'(k)}{n(k)}$, i.e. the degree of asymmetry between the manager's and the worker's contribution to output. The greater is this asymmetry, the greater is the increase in the earnings of the worker when he becomes a manager and the greater is the wage cut that the worker is willing to accept in order to learn.

Proposition 6 also sheds light on how welfare changes as we change the rate of learning. Assuming that the RHS of the expression in Proposition 6 is positive, the inequality will not be satisfied if learning is very slow. In the limiting case of no-learning, $\delta = 1$ (or $\theta = 1$), the LHS is equal to zero. As we start reducing δ (or θ) i.e. raise the rate of learning, the LHS starts to increase and at some point, it is greater than the RHS. Recall from Proposition 3 that k_I^* is decreasing in

[40] For a similar effect in a trade model, see Melitz (2003).

[41] Since $\frac{n(k)}{1+n(k)}$ is increasing in k.

both δ and θ. Therefore, if the rate of learning continues to increase, eventually k_I^* exceeds $\overline{k}_H$. In this case, integration still generates Pareto gains for the Home country, but at the cost of all the incumbent Home firms (see previous section).

Corollary 1 stated that some of the incumbent firms produce less under integration relative to autarky. For these firms to be better off, they have to pay lower wages to the workers. Corollary 2 follows.

Corollary 2 If all Home agents gain from integration, Home workers must be earning a lower wage compared to autarky.

In our model, if a MNE has more knowledge than an incumbent Home firm, PAM implies that, the workers in the MNE are more knowledgeable than the ones in the Home firm. PAM also implies that after working for the MNE, a worker never works for the Home firm. Therefore, there is no flow of knowledge from the MNE to the Home firm. Despite this, the incumbent firm can be made better off if learning is fast enough.[42] Of course, some of the former multinational workers set up their own firms and these managers directly benefit from the superior knowledge of MNEs.[43]

5 Numerical analysis

To further our understanding of the model, we resort to numerical analysis. We study how the equilibrium changes with the rate of learning, and whether the welfare results from the previous section go through for a more general learning technology. We go on to analyze the evolution of individual earnings over the lifetime, the change in inequality due to integration and the pattern of multinational activity. Throughout, our focus is on the Home economy.

The only change from the last section is in the learning technology. We assume that a worker with knowledge x, and working for a manager with knowledge y, draws his knowledge in the next period from a distribution which is uniform on $[x, y]$. The production function is given by $f(y) = y^{\alpha}$, $n(x; \beta) = x^{\beta}$. Finally, we assume that the distribution of newborns is a truncated exponential in $[1, \overline{k}]$ with parameter λ. By setting $\underline{k} = 1$, we are implicitly setting the size of the smallest firm to two (one manager and one worker). The following figures are drawn for $\alpha = 1$, $\beta = 0.5$, $\lambda = 1$, $\sigma_H = \sigma_F = 1$, $\overline{k}_H = 1.5$, $\overline{k}_F = 1.75$.

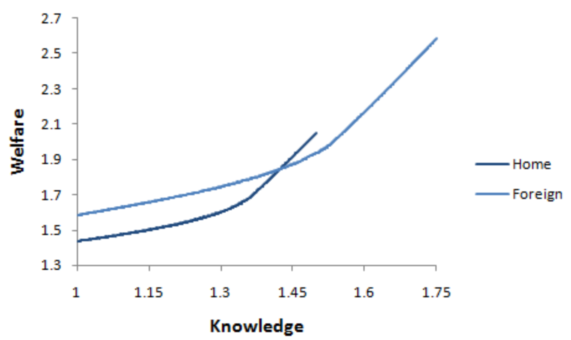
5.1 Earnings and Welfare

In Figure 1, we compare the Home and Foreign economies under autarky. Figure 1a shows the welfare of an agent *as a function of his knowledge at birth.* The only difference between the

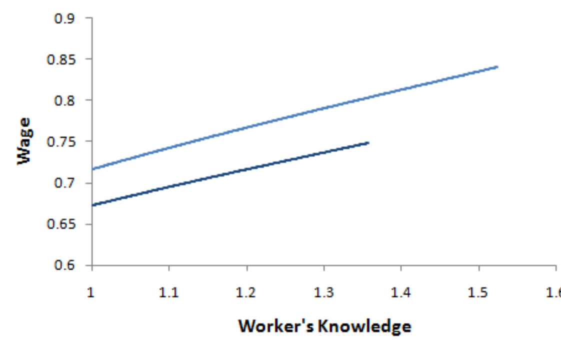
[42] The traditional view regarding knowledge spillover is that workers with experience in MNEs are hired by domestic managers. These workers bring with them knowledge regarding better technology and management practises and the domestic firms benefit from this. This superior knowledge leads to an increase in firm productivity, which translates into higher earnings for all. See, for example, Barba Navaretti and Venables (2004).

[43] This effect is similar to Monge-Naranjo (2007) where the transfer of skills fom MNEs materialize in a new sector of firms, not in the pre-existing sector of firms.

two countries is in the distribution of newborns. In particular, the Foreign country has a larger knowledge support. This translates into relatively greater endowment of more knowledgeable agents in the Foreign country.[44]



1a: Welfare under autarky



1b: Wages under autarky

Figure 1: Comparison of the Home and Foreign economy under autarky

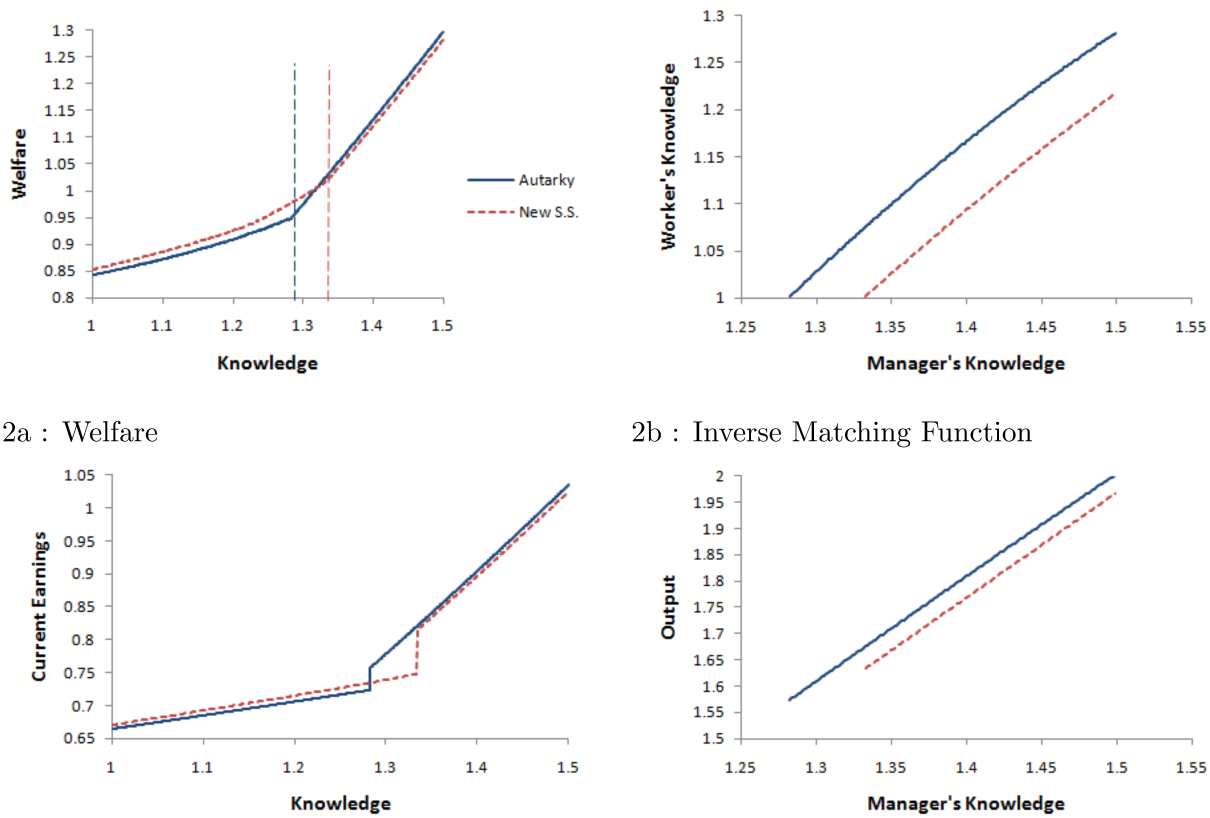
The less knowledgeable agents are relatively scarce in the Foreign country and hence, are better off compared to their Home counterparts. But the most knowledgeable agents at Home are better off compared to their foreign counterparts. The latter are low-level managers in the Foreign country, who have to compete against managers with superior knowledge. The relative abundance of less knowledgeable agents at Home translates into a Home wage schedule that lies below the Foreign wage schedule. Thus, labour is cheap at Home and this motivates the formation of MNEs.

Next, we allow the countries to integrate. First, we consider a case where learning is slow. Keeping the learning distribution unchanged, this means a higher δ. We choose $\delta = 0.8$. In Figure 2, we compare the two steady-states: autarky and integration. In Figure 2a, we see that in the steady-state under integration (New S.S.), the welfare of individuals who are born with less knowledge is higher while the welfare of those with high levels knowledge is lower. Figure 2b indicates that the incumbent managers have a worse match; thus every incumbent manager is now producing less. This is confirmed in Figure 2d. This implies lower revenue. But the effect on profits, which determines the managers' welfare, also depends on the wage bill.

From the discussion in the previous section, we know that, the effect of integration on wages depends on the relative strength of the labour demand effect and the learning effect. For a slow learning rate, the former effect dominates and the wage schedule shifts up, thereby lowering the profits of incumbent managers. This is shown in Figure 2c. This explains the reduction in welfare of the more knowledgeable agents. Notice that we restrict our attention to the agents who are born in

[44]If both Φ_H and Φ_F are truncated exponentials with the same parameter and $\overline{k}_H < \overline{k}_F$, then Φ_F first-order stochastic dominates Φ_H.

$[\underline{k}, \overline{k}_H]$. Although, in the new steady state, there are Home agents with knowledge in $[\overline{k}_H, \overline{k}_F]$, but at the time of birth, these agents still draw their knowledge from the Home newborn distribution, which does not change with integration. Home agents attain knowledge in $[\overline{k}_H, \overline{k}_F]$ *through learning, not through birth.*



2a : Welfare

2b : Inverse Matching Function

2c : Current Earnings

2d : Output

Figure 2 : Change in the values of different variables due to integration ($\delta = 0.8$)

In the above scenario, we compared the steady-state under autarky and integration. To compute the welfare gains from the integration policy, however, one should compare the autarky steady-state equilibrium with the integrated equilibrium in the period just after the policy is put in place. With the policy, there is no immediate change in the knowledge distribution. Consequently, the matches do not change. However, as soon as the policy is implemented, the agents' expectations about the future knowledge distribution change. The agents know that when the Home country opens up, the matches will be determined by the integrated knowledge distribution. Therefore, the matches will change, and by changing the way agents learn, will affect the distribution. The knowledge distribution will evolve slowly, until it reaches the new steady-state.

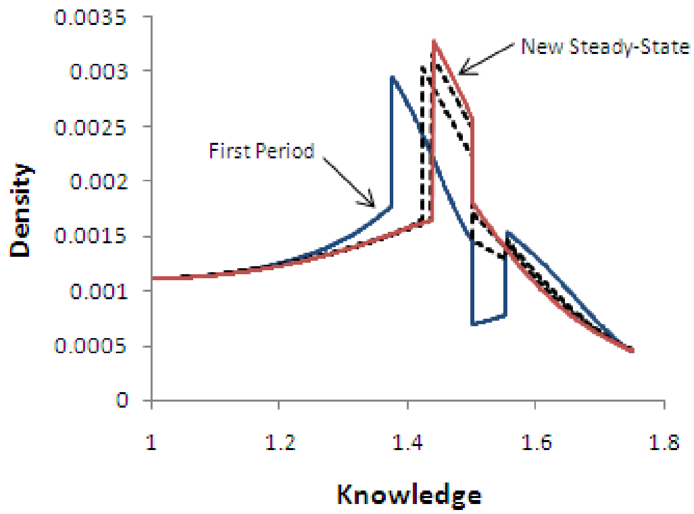


Figure 3: Evolution of the Knowledge Distribution

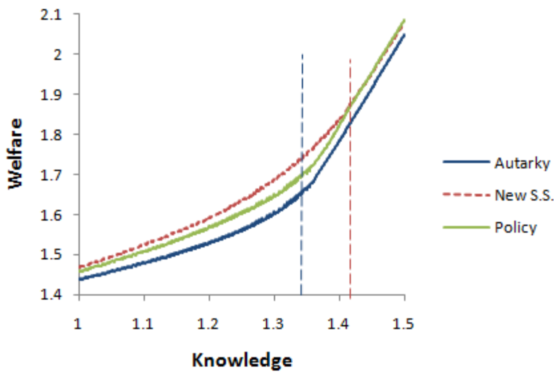
The above figure shows the transition of the integrated knowledge distribution from the time the Home country integrates until it reaches the new steay-state. The intial distribution is simply the sum of the Home and Foreign steady-state distributions. There are three discontinuities in the initial distribution. The first one occurs at k_A^*.[45] The second one occurs at $\overline{k}_H$. This is due to the fact that no Home agents are born to the right of $\overline{k}_H$. The third discontinuity is at the Foreign threshold. In the new steady-state, the discontinuities occur at k_I^* and $\overline{k}_H$.

The agents know exactly how the distribution is going to evolve and hence, they know the wages and profits at each period during the transition. This, in turn, allows them to compute their welfare in every period. It can be shown that for $\delta = 0.8$, not only are there losers in the new steady-state, but the policy itself creates losers.

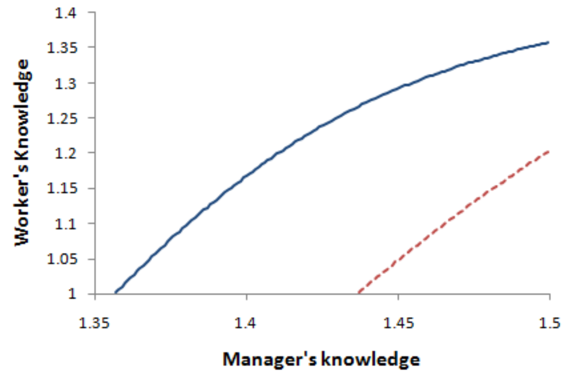
The outcomes change if we increase the learning rate. The results are displayed in Figure 4 for $\delta = 0.5$. Now, the learning effect dominates the labour demand effect, thereby lowering the wage schedule. Although the output of the incumbent Home managers fall (Figure 4d) due to a worsening of their matches (Figure 4b), the wage bill decreases by so much, that it outweighs the fall in revenue, resulting in higher profits. This makes the home managers better-off. The less knowledgeable agents are better-off too, as the increase in their continuation value outweighs the decline in wages. This is true both in the new steady-state, as well as, in the period following the

[45] There is always a discontinuity at the threshold. In a small interval to the left of the threshold (where all agents are workers) there is both an inflow of workers and an outflow of workers. But in a small interval just to the right of the threshold, there is only inflow and no outflow (since managers do not learn).

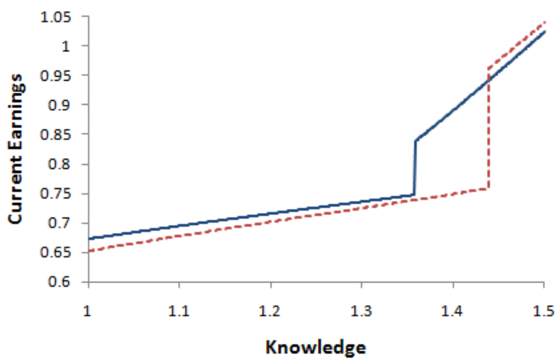
policy implementation.[46]



4a: Welfare



4b: Inverse Matching Function



4c: Current Earnings

4d: Output

Figure 4: Change in the values of different variables due to integration ($\delta = 0.5$)

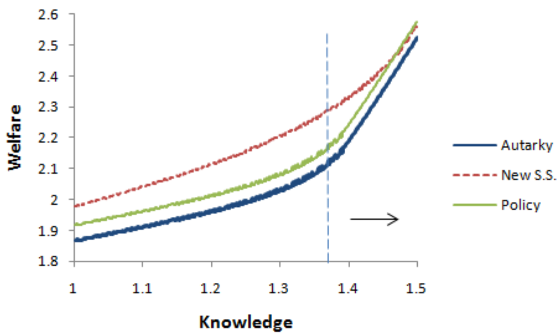
Finally, Figure 5 shows the results for $\delta = 0.4$. For this rate of learning, k_I^* exceeds $\overline{k}_H$. This situation corresponds to Case I from the last section. Integration makes every agent better-off. However, every newborn agent is a worker in the integrated economy. Under this situation, none of the incumbent Home managers operate.

The above plots suggest that the incumbent firms experience a decline in output, irrespective of whether agents are learning at a slow or a fast rate. According to Aitken and Harrison (1999), foreign direct investment was accompanied by a decline in the productivity of domestically owned firms in Venezuela. This decline, the authors report, is due to a contraction in output of domestic firms due to the "market stealing effects" of foreign firms.[47] In our model, the output reduction

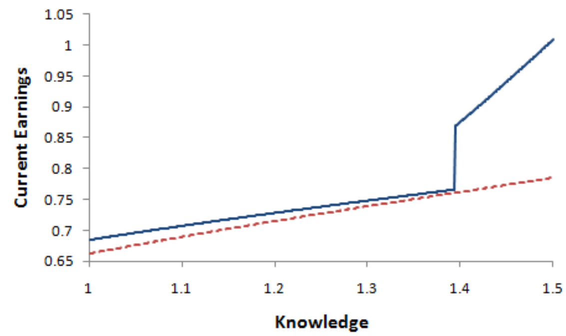
[46] It can be shown that the evolution of the value function during the transition is monotonic. This implies that for $\delta = 0.5$, agents are better-off compared to autarky at every period during the transition.

[47] With fixed costs of production, foreign firms with lower marginal costs can expand their output at the expense of domestic firms.

is a natural consequence of complementarity in production and learning. Under integration, the most knowledgeable workers are hired by the MNEs leaving less knowledgeable workers to work for incumbent domestic firms. We call this the "worker stealing effect".[48] In spite of this, the Home managers are actually better off if learning is fast enough.



5a: Welfare



5b: Current Earnings

Figure 5: Change in the values of different variables due to integration ($\delta = 0.4$)

Evidence regarding the impact of multinational production on wages has been mixed. Aitken *et al* (1996) report that in Mexico and Venezuela, the wage spillover to domestic firms is negative and significant. On the other hand, Lipsey and Sjöholm (2004) find significant positive wage spillovers to domestic firms in Indonesia. In the previous section, we saw that if learning is fast enough, the wage schedule shifts down. This result, however, is not inconsistent with the finding of positive wage spillovers. In the above mentioned studies, the wage is computed as the average of wages paid by all domestic firms. The average wage depends not only on the level of the wage schedule but also on the distribution of agents. With integration, as workers get matched with better managers and learn more, the mass shifts to the right. Therefore, a lowering of the wage schedule and a higher average wage can go hand in hand.

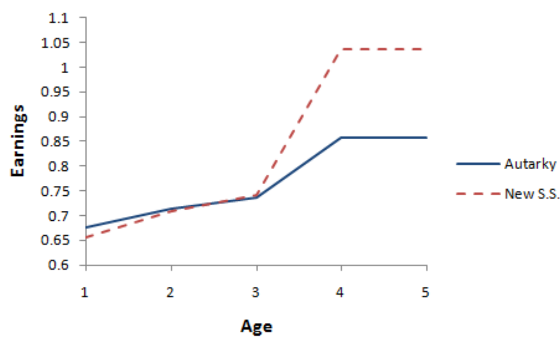
The numerical results of this section confirm the analytical results obtained in the last section. When learning is slow, integration creates winners and losers. In the above example, the more knowledgeable agents in the host country lose. But if agents learn fast enough, integration can make every agent better-off. In this case, there is a decline in the output of the incumbent firms, as well as, the wages of the workers. Therefore, current wages or productivity could be misleading when it comes to assessing welfare gains from integration.

[48] I am grateful to Thomas Chaney for suggesting this term.

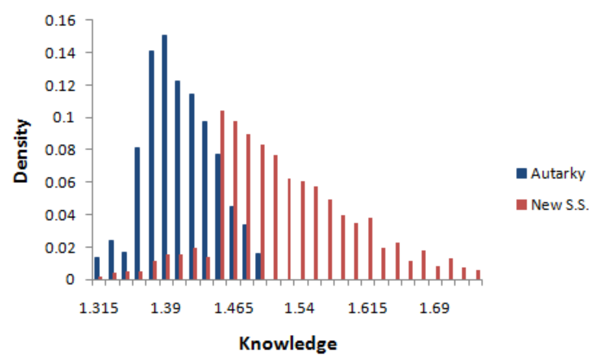
5.2 Earnings Dynamics

Although the economy is not growing in the steady-state, individual earnings grow over the lifetime. Figure 6 displays some of the dynamic aspects of the model for $\delta = 0.5$.

Figure 6a plots the earnings path of the worker with median knowledge. We assume that the knowledge that he acquires every period is the expected knowledge that an agent with his level of knowledge would acquire. In the figure, the agent works for the first three periods and manages from the fourth period onwards.[49] Under integration, a lower wage in the first two periods is more than compensated by the increase in future profits. The lifetime earnings schedule under integration is also steeper than that under autarky. As long as the agent is a worker, under integration, he gets matched with more knowledgeable managers. This implies that the relative jump in wages every period is greater under integration.



6a: Evolution of earnings over the lifetime



6b: Distribution of knowledge at death

Figure 6: Dynamics of knowledge and earnings

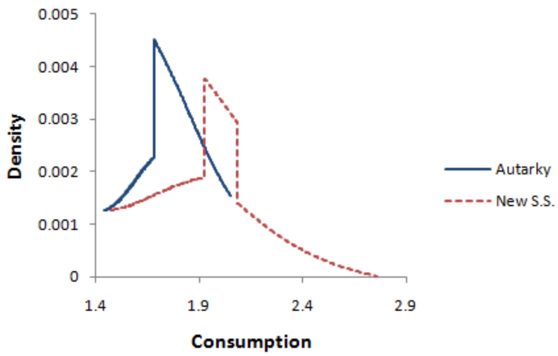
Figure 6b provides an explanation for the greater jump in future profits. It shows the distribution of knowledge of the median agent at the time of death.[50] With integration, the distribution shifts to the right. On average, the agent becomes a more knowledgeable manager compared to autarky and hence, his expected profits are higher.

5.3 Inequality

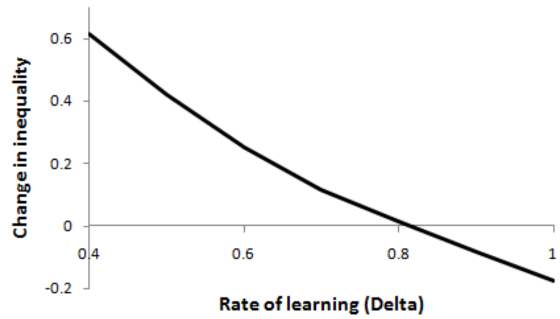
By generating a non-degenerate consumption distribution, the model also allows us to talk about inequality. Figure 7 examines the effect of integration for inequality. δ is set to 0.5.

[49] Note that once the agent becomes a manager, his earnings do not change because he stops learning.

[50] Given the parameter values, the probability of the agent living for more than 5 periods is extremely small. Here we are assuming that the agent dies after 5 periods.



7a: Consumption distributions



7b: Change in Gini coefficients

Figure 7: Income inequality

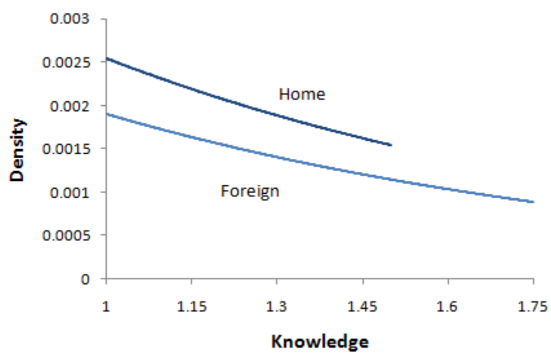
Figure 7a plots the Home consumption distributions under autarky and integration. With integration, the distribution stretches out and more mass is shifted to the upper tail (The maximum consumption under autarky is 2.05, while that under integration is 2.77). From the figure, it is not evident whether inequality goes up or down. Figure 7b plots the percentage change in the gini coefficient as the Home country integrates, for different values of δ. From the figure we see that for $\delta = 0.5$, inequality rises by about 40%. But this rise in inequality is not a general phenomenon. For higher values of δ, integration actually leads to a reduction in inequality. For lower values of δ, however, integration increases inequality. Moreover, there is a monotonic relation between inequality and the the rate of learning.

Integration, when agents learn fast enough, gives an advantage to those who are born with a lot of knowledge, but not enough to become managers. These are the agents who work for the most knowledgeable Foreign managers, learn and in turn, become knowledgeable managers in the future. Agents who are born with very little knowledge are matched to the less knowledgeable incumbent Home managers and accordingly, learn less. Learning amplifies the initial inequality in the economy.

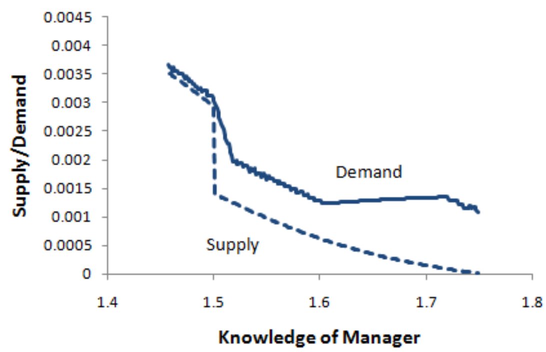
5.4 Pattern of MNE activity

Our model also allows us to analyze the pattern of MNE activity. In the previous section, we had mentioned that integration leads to the creation of a new class of Home managers, who are as productive as their counterparts in the Foreign country. This can be seen in Figure 8. Figure 8a shows the newborn distribution of the two countries. Under autarky, the support of the Home knowledge distribution coincides with that of the Home newborn distribution and hence, the knowledge of the best manager is bounded above by $\overline{k}_H$. Figure 8b plots the supply and demand for managers at Home in the integrated steady-state equilibrium. The supply of managers is simply the part of the knowledge distribution that lies above the threshold. Recall that the upper bound of the Home newborn distribution is 1.5. Therefore, there is a discrete drop in the density of newborn agents to the right of 1.5 and this explains the discontinuity at 1.5. The demand for managers is obtained by looking at the number of workers of each type and the demand for manager per worker.[51]

Figure 8b suggests that, in the integrated steady-state equilibrium, there are Home managers who are as knowledgeable as their Foreign counterparts. At the same time, the supply of Home managers is not sufficient to meet the demand. In the new equilibrium, some of the Foreign managers hire Home workers and hence, Home firms and Foreign MNEs operate together.[52] Figure 8b also throws light on the pattern of multinational activity. The supply of Home managers falls short of demand, and there are no Home MNEs in this equilibrium.[53] Moreover, most of the MNEs operating at Home are the best Foreign firms. Thus, the MNEs, on average, are bigger and more productive than the Home firms. PAM implies that a worker in a MNE, on average, is more knowledgeable than a worker in a Home firm. Therefore, the former employees of MNEs are also more productive managers.



8a: Newborn distributions



8b: Supply and demand for managers

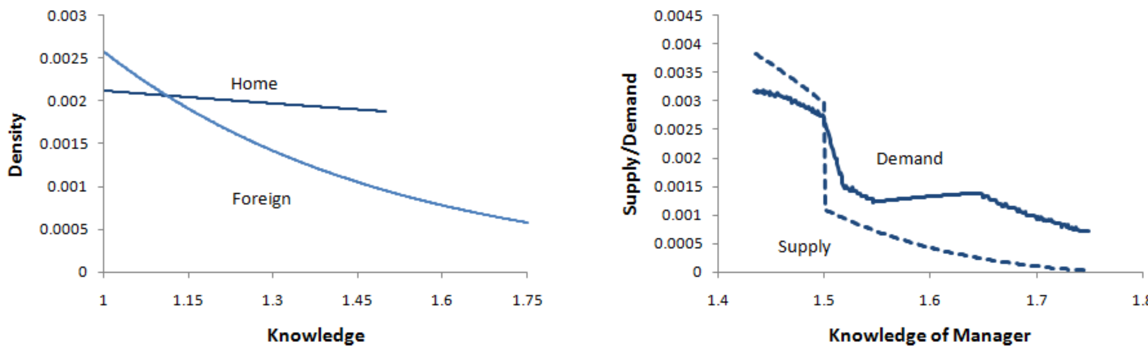
Figure 8: MNE activity at Home

Alternate newborn distributions : The pattern of MNE activity described above depends on the newborn distributions in the two countries. In the above example, we had assumed that the Foreign newborn distribution dominates the Home newborn distribution in the sense of first-order stochastic dominance. The pattern changes if we relax this assumption. Figure 9 considers a case where the Foreign newborn distribution has relatively more less knowledgeable people relative to the Home newborn distribution.

[51] The demand for manager per worker is simply the reciprocal of the span of control. This is a feature of the span of control being only a knowledge of the worker.

[52] See Markusen and Venables (1999) for the case where FDI leads to the development of local industry but is driven out as the industry develops enough.

[53] Here we follow AGR in assuming that, a manager will hire workers in the other country only if he strictly prefers doing so.



9a: Newborn distributions

9b: Supply and demand for managers

Figure 9: MNE activity at Home (different degree of inequality)

Figure 9a shows the two newborn distributions, where the $\overline{k}_F$ is still greater than $\overline{k}_H$. The pattern of MNE activity that results from such distributions is shown in Figure 9b. The supply of less knowledgeable Home managers exceeds demand, while the converse is true for the more knowledgeable Home managers. In the integrated equilibrium, we witness multinational activity in both directions, with the smaller MNEs originating in the Home country and the larger MNEs originating in the Foreign country.[54]

In a survey of firms in Ghana, Görg and Strobl (2005) investigate whether knowledge spillovers occur through worker mobility. They combine information on whether or not the owner of a domestic firm had previous experience in a multinational with information on firm-level productivity. They show that firms which are run by owners who worked for foreign multinationals in the same industry immediately prior to opening their firm are more productive than other domestic firms. Using data on Danish firms, Malchow-Møller *et al* (2006) show that previous experience in foreign-owned firms increases a worker's current wage. Both pieces of evidence are consistent with our model.

Therefore, in our model, domestic firms and MNEs coexist. On average, the foreign MNEs tend to be bigger and more productive. Our model also generates multinational activities by firms in both countries.

6 Gains from Globalization

In this section, we measure the gains from bilateral integration. We perform the following counterfactual exercise. First, we compute the GDP and associated welfare of the host country under autarky. Then we allow it to integrate with a foreign country which has relatively more knowledge-

[54] For a model with two-way FDI flows, see Nocke and Yeaple (2008). In their model, these FDI flows take the form of cross-border mergers and acquisitions.

able agents. This leads to the formation of international teams. By looking at the resulting GDP and welfare in the host country and comparing it with the autarky levels, we get an estimate of the gains from integration. In order to carry out this exercise, we need the values for the technology parameters α, β, the parameter of the exponential distribution λ, the probability of death δ, and the upper bound $\overline{k}$.

The most difficult parameter to obtain is δ. To our knowledge, there is no paper which tries to understand how workers learn within firms. In order to come up with an estimate for δ, we proceed as follows. In an influential paper, Topel (1991) showed that ten years of current job seniority raises the wage of the typical male worker in the United States by over 25 percent. This is an estimate of what the typical worker would lose if his job were to end exogenously. One can consider this as evidence that workers accumulate firm-specific knowledge and it gives an indication of the speed with which workers learn. We show in the Appendix that this translates into the unit of time being 17 years.

We also choose the exit rate for firms. The probability of exit for firms, however, varies widely in the U.S. economy. Larger and older firms have a smaller probability of exit compared to smaller and new firms (Dunne *et al* (1989)). In this paper, we assume that all firms face the same probability of exit. Dunne *et al* (1988) report an average exit rate that varies between 0.3 and 0.39 between each pair of census years over the period 1963-82,[55] where the gap between consecutive census years is 5 years. We take the average exit rate in 5 years to be 0.34. δ, which is the probability of exit of a firm over 17 years, turns out to be approximately 0.7.

We choose the three parameters α, β, λ to match three key moments of the U.S. wage distribution. A measure of inequality that is quite popular in the literature, is the ratio of the 90th percentile to the 10th percentile (Juhn *et al*). We ignore the observations at the tails and instead choose a different measure : the ratio of the 75th percentile to the 25th percentile. This ratio was around 2.5 in 2007 (Source : U.S. Bureau of Labor Statistics). This ratio is a measure of dispersion of the wage distribution. We also choose the ratio of the mean to the median, which can be interpreted as a measure of skewness. In 2007, this value was around 1.25. The third moment that we match is the ratio of the median wage of managers to the overall median wage. The corresponding value in 2007 was about 2.6. The implied values of the parameters are shown in the following table.

Parameter	δ	α	β	λ
Value	0.7	1.9	0.55	1.4

Table 1: The values of the parameters

We assume that the parameter values in Table 1 are common to all countries. This means that there are other country-specific factors that generate different levels of output and factor prices.

[55] This excludes the smallest firms.

The country-specific parameters in this model are $\overline{k}$ and σ. We interpret $\overline{k}$ (and hence the support of the knowledge distribution) as being determined by the education policies of the government and the general research environment while σ is a measure of business risk. These two, combined with technology, result in final output. In order to calibrate these two parameters for the different countries, we proceed as follows.

We divide the countries into groups based on their autarky per capita GDP relative to the U.S. Since those figures are not available, we choose the relative per capita GDP of these countries in 1970.[56] The figures for per capita GDP have been obtained from the Penn World Tables.

We assume that every country within an income group faces the same σ. We also assume that the richest country in an income group has the same $\overline{k}$ (and hence the same distribution) as the average country belonging to the income group just above it. The difference in income between these two countries arises due to a difference in σ. Therefore, once we know the average $\overline{k}$ of an income group, we can figure out the σ for the next income group. Within an income group, the difference in income arises only due to a difference in $\overline{k}$. Once we know the $\overline{k}$ of the richest country in a group, we can compute the average $\overline{k}$ for that group. We repeat this procedure for every group. Our classification scheme, and σ and $\overline{k}$ of the average country in each income group, is shown in Table 2.

Groups	Basis of classification[57]	Average p.c. GDP	Average $\overline{k}$	σ
High-income	Above 60%	1 (normalized)	2	0.7
Middle-income (high)	40%-60%	0.64	1.85	0.5
Middle-income (low)	20%-40%	0.35	1.5	0.38
Low-income (high)	10%-20%	0.19	1.21	0.29
Low-income (low)	Below 10%	0.07	1.05	0.18

Table 2: Country specific parameters

Assumption 1c imposes some restriction on the support of the knowledge distribution. For the High-income countries, we fix $\overline{k}$ at 2. We also normalize the average per capita GDP of the High-income countries to 1. This gives a value of σ for this group equal to 0.7. Following the procedure outlined above, we back out the $\overline{k}$ and σ for the other income groups.

Once we have all the parameters, we perform the following exercise. We let the average country

[56] Our choice of 1970 as the base year is due to the fact that the Bretton Woods collapsed in 1971. During the entire post-war period, until the fall of the Bretton Woods, most of the countries had a system of fixed exchange rates. Consequently, there was a lot of capital control in place which limited cross border investment, both in financial assets and FDI (Irwin(2002)). This can be judged from the fact that FDI inflows in 1970-73 was around 1.5% of what it was in 1998-2001 (UNCTAD). Of course, world GDP has also been growing during this time. But the growth in real FDI inflows at 17.7% per year far outstripped the corresponding growth in real GDP of 2.5% per year during the period 1985-99.

[57] per capita GDP as percentage of U.S. per capita GDP.

from each of the income groups integrate with the average High-income country. The gains from integration are displayed in Table 3.

	Percentage change in			Percentage change
	GNI	GDP	Welfare	due to learning
Middle-income (high)	1.6%	10.9%	2%	98%
Middle-income (low)	5.7%	57.1%	13.2%	88%
Low-income (high)	10.2%	157.9%	32.1%	83%
Low-income (low)	22.2%	300%	42.9%	74%

Table 3: Percentage change in variables due to integration

Under autarky, GDP and Gross National Income (GNI) are the same. That is not true anymore under integration. Table 3 suggests that the gain in per capita GDP is much larger than the gain in per capita GNI. Under integration, the workers in the host country are matched with better managers. Since output per worker depends on the knowledge of the manager he is matched with, integration has a significant impact on output. The effect on GNI is much more muted. As the analysis in the last section suggests, higher wages are associated with lower profits and vice versa. The positive effect on welfare, however, is much more pronounced. Under integration, the Home agents learn from more knowledgeable managers and this raises the present value of their income, even when current wages go down. These welfare gains range from a low 2% for the relatively rich Middle-income countries to almost 43% for the poorest countries of the world.

The gains from integration can be decomposed into two parts. There are gains from more efficient allocation of managers to workers. These gains are purely static in nature. On top of this, there are dynamic gains because learning changes the knowledge distribution. The last column in Table 3 shows the percentage of gains that can be attributed to learning. This percentage is higher for the richer developing countries. The Middle-income countries are more similar to the High-income countries in terms of their knowledge distribution. As a result, the gains from re-allocation are quite small.

7 Conclusion

In this paper, we develop a dynamic, general equilibrium model to study the effects of knowledge diffusion from MNEs on host countries. In our model, agents are heterogenous in terms of the knowledge they possess. Every period, an agent chooses his occupation (worker or manager) and with whom he wants to work. We assume that workers learn on the job. Both the production and the learning technology exhibit complementarity. We show that the complementarity of the production and learning technologies results in positive assortative matching (PAM). In the stationary equilibrium, every agent born above a knowledge threshold is a manager. Agents born with knowledge less than the threshold, work for more knowledgeable managers and learn. Some of them eventually become managers.

We allow the Home country to integrate with a Foreign country, where the Foreign country has relatively more knowledgeable agents. By integration, we mean that managers are able to form international teams. First, we consider a version of the model where agents do not learn. Integration leads to a re-adjustment among host country managers whereby the least knowledgeable among them exit, whereas among those who persist as managers, some are now matched with less knowledgeable workers. Under this situation, we show that integration can never generate Pareto gains for the Home country.

In the presence of learning, there are two effects that determine wages: the labour demand effect and the learning effect. The former tends to raise the wage schedule while the latter tends to lower it. We show that if learning is fast enough, the learning effect dominates and the wage schedule shifts down enough so as to make the incumbent managers better-off. At the same time, the continuation value of the workers outweighs the decline in their current wages, thereby making them better-off. We believe that this is a novel mechanism through which integration, in the presence of learning, can lead to Pareto improvement for the Home country.

In the quantitative section, we calibrate the closed-economy model to match key moments of the U.S. wage distribution. Then we perform the following counterfactual: How much do the developing countries gain by moving from autarky to frictionless integration with the average developed country? Using parameter values obtained from the calibration, we find welfare gains that range from two percent for the richest middle-income countries to almost forty-three percent for the poorest countries.

To conclude, we do not have a good understanding of, for example, how knowledge evolves and the distribution of skills change as a country gradually integrates with the rest of the world. We believe that our paper is a step in that direction. Our quantitative analysis, although performed under some strong assumptions about the knowledge distribution of newborns, suggests that the dynamic gains from integration are non-negligible, especially for the poorest countries of the world. We believe that introducing growth in our model will generate new insights about the relation between growth and openness and we leave that for future work.

Reference

1. **Aitken, B. and A.E. Harrison** (1999), "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela", *The American Economic Review*, 89(3), 605-618.

2. **Aitken, B., A.E. Harrison, R. Lipsey** (1996), "Wages and foreign ownership : A comparative study of Mexico, Venezuela, and the United States", *Journal of International Economics*, 40(3-4), 345-371.

3. **Antràs, P.** (2003), "Firms, Contracts and Trade Structure", *The Quarterly Journal of Economics*, 118(4), 1375-1418.

4. **Antràs, P. , L. Garicano and E. Rossi-Hansberg** (2006), "Offshoring in a Knowledge Economy", *The Quarterly Journal of Economics*, 121(1), 31-77.

5. **Antràs, P. and E. Helpman** (2004), "Global Sourcing", *The Journal of Political Economy*, 112(3), 552-580.

6. **Barba Navaretti, G. and A. Venables** (2004), "Multinational Firms in the World Economy", Princeton University Press.

7. **Becker, G.** (1962), "Investment in Human Capital: A Theoretical Analysis", *The Journal of Political Economy,* 70(S5) , 9-49.

8. **Blomström, M. and A. Kokko** (1998), "Multinational Corporations and Spillovers", *Journal of Economic Surveys*, 12(3), 247-277.

9. **Boyd, J. H. and E. C. Prescott** (1987), "Dynamic Coalitions: Engines of Growth", *The American Economic Review*, 77(2), 63-67.

10. **Burstein, A. and A. Monge-Naranjo** (2008), "Foreign Know-How, Firm Control and the Income of Developing Countries", forthcoming *The Quarterly Journal of Economics.*

11. **Dunne T., M. J. Roberts and L. Samuelson** (1988), "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries", *The RAND Journal of Economics*, 19(4), 495-515.

12. **Dunne T., M. J. Roberts and L. Samuelson** (1989), "The Growth and Failure of U. S. Manufacturing Plants", *The Quarterly Journal of Economics*, 104(4), 671-698.

13. **Fosfuri A., M. Motta and T. Rønde** (2001), "Foreign direct investment and spillovers through workers' mobility", *Journal of International Economics*, 53(1), 205-222.

14. **Garetto, S.** (2008), "Input Sourcing and Multinational Production", mimeo University of Chicago.

15. **Garicano, L.** (2000), "Hierarchies and the Organization of Knowledge in Production", *The Journal of Political Economy*, 108(5), 874-904.

16. **Gershenberg, I.** (1987), "The training and spread of managerial know-how, a comparative analysis of multinational and other firms in Kenya", *World Development*, 15(7), 931-939.

17. **Giarratana, M., A. Pagano and S. Torrisi** (2004), "The Role of Multinational Firms in the Evolution of the Software Industry in India, Ireland and Israel", in A. Arora and A. Gambardella (eds), From Underdogs to Tigers: Bridging the Gap, Oxford University Press, New York.

18. **Görg, H. and E. Strobl** (2005), "Spillovers from Foreign Firms through Worker Mobility: An Empirical Investigation", *Scandinavian Journal of Economics*, 107(4), 693-709.

19. **Görg, H., E. Strobl and F. Walsh** (2007), "Why Do Foreign-Owned Firms Pay More? The Role of On-the-Job Training", *Review of World Economics*, 143(3), 464-482.

20. **Grossman, G. and E. Helpman** (2003), "Outsourcing versus FDI in Industry Equilibrium", *Journal of the European Economic Association*, 1(2-3), 317-327.

21. **Helpman, E.** (1984), "A Simple Theory of International Trade with Multinational Corporations", *The Journal of Political Economy*, 92(3), 451-471.

22. **Holmes, T. and J. Schmitz** (1990), " A Theory of Entrepreneurship and Its Application to the Study of Business Transfers", *The Journal of Political Economy*, 98(2), 265-294.

23. **Irwin, D.** (2002), "Free Trade under Fire", Princeton University Press.

24. **Jovanovic, B. and R. Rob** (1989), "The Growth and Diffusion of Knowledge", *The Review of Economic Studies*, 56(4), 569-582.

25. **Kogut, B. and U. Zander** (2003), "Knowledge of the Firm and the Evolutionary Theory of the Multinational Corporation", *Journal of International Business Studies,* 34(6), 516-529.

26. **Kremer, M. and E. Maskin** (1996), "Wage Inequality and Segregation by Skill", NBER Working Paper W5718.

27. **Legros, P. and A. Newman** (2002), "Monotone Matching in Perfect and Imperfect Worlds", *The Review of Economic Studies*, 69(4), 925-942.

28. **Lipsey, R. and F. Sjohölm** (2004), "FDI and Wage Spillovers in Indonesian Manufacturing", *Review of World Economics*, 140(2), 287-310.

29. **Lipsey, R. and F. Sjohölm** (2005), "The Impact of Inward FDI on Host Countries: Why Such Different Answers?" in "Does Foreign Direct Investment Promote Development? : New Methods, Outcomes and Policy Approaches" by T. H. Moran, E. M. Graham, M. Blomström, Peterson Institute, 23-44.

30. **Lucas, R. E.** (1978), "On the Size Distribution of Business Firms", *The Bell Journal of Economics*, 9(2), 508-523.

31. **Lucas, R. E. and N. Stokey with E. C. Prescott** (1989), "Recursive Methods in Economic Dynamics", Harvard University Press.

32. **Malchow-Møller, N., J. R. Markusen and B. Schjerning** (2006), "Foreign Firms, Domestic Wages", CEBR Copenhagen working paper.

33. **Markusen, J. and A. Venables** (2007), "Foreign direct investment as a catalyst for industrial development", *European Economic Review*, 43, 335-356.

34. **Melitz, M. J.** (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica*, 71(6), 1695-1725.

35. **Monge-Naranjo, A.** (2007), "Foreign Firms, Domestic Entrepreneurial Skills and Development", mimeo Northwestern University.

36. **Nocke, V. and S. Yeaple** (2008), "An Assignment Theory of Foreign Direct Investment", *The Review of Economic Studies*, 75(2), 529-557.

37. **Ohnsorge, F. and D. Trefler** (2007) "Sorting It Out: International Trade with Heterogeneous Workers", *The Journal of Political Economy*, 115(5), 868-893.

38. **Poole, J. P.** (2006), "Multinational Spillovers through Worker Turnover", mimeo University of California, San Diego.

39. **Ramondo, N.** (2008), "Size, Geography and Multinational Production", mimeo University of Texas at Austin.

40. **Rodríguez-Clare, A.** (1996), "Multinationals, Linkages, and Economic Development", *The American Economic Review*, 86(4), 852-873.

41. **Rodríguez-Clare, A.** (2007), "Trade, Diffusion and the Gains from Openness", NBER Working Paper 13662.

42. **Rosen, S.** (1972), "Learning and Experience in the Labor Market", *The Journal of Human Resources*, 7(3), 326-342.

43. **Rosen, S.** (1982), "Authority, Control and the Distribution of Earnings", *The Bell Journal of Economics*, 13(2), 311-323.

44. **Sattinger, M.** (1993), "Assignment Models of the Distribution of Earnings", *Journal of Economic Literature*, 31(2), 831-880.

45. **Topel, R.** (1991), "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority", *The Journal of Political Economy*, 99(1), 145-176.

Appendix

Proof of Lemma 1. Let $\mathbb{C}$ be the space of bounded, continuous functions defined on $\mathbb{R}$. Define the operator $T(V_M(.,.))$ as

$$T(V_M(y, w_{t+1})) = \sup_{x \in [\underline{k}, y]} \{\sigma f(y)n(x) - w_t(x)n(x)\} + (1-\delta)\max[V_W(y, w_{t+1}), V_M(y, w_{t+1})]$$

We need to show that T preserves continuity and boundedness. Continuity follows from the fact that $f(.), n(.)$ and $w_t(.)$ are continuous and so is the *max* operator. To prove boundedness of T, we note that $f(y)$ and $n(x)$ are continuous functions defined on the compact set $[\underline{k}, y]$ and hence must be bounded. Furthermore, $w_t(x) \leq \sigma f(y)$ (otherwise profits will be negative) and hence $w_t(x)$ must be bounded too. Hence T preserves continuity and boundedness, i.e. $T(V_M(.)) : C \to C$.

It's easy to see that if $V_M^1(y,w) \geq V_M^2(y,w)$ for all x, then $T(V_M^1(y,w)) \geq T(V_M^2(y,w))$. Furthermore, $T(V_M(y,w)+a) \leq T(V_M(y,w)) + (1-\delta)a$. Since $(1-\delta) < 1$, T satisfies Blackwell's Sufficiency Conditions for a contraction and hence, by the Contraction Mapping Theorem, there exists a unique, fixed point of $V_M(y,w)$.

Let $V_W(y,x')$ and $V_M(y,x')$ be increasing in y. Then $\max[V_W(y,x'), V_M(y,x')]$ is also increasing in y. Furthermore, $f'(y) > 0$. Combining, we get

$$T(V_W(y_1)) > T(V_W(y_2))$$

if $y_1 > y_2$. That is, the contraction is increasing in y. Hence it must be the case that the fixed point is increasing in y too, i.e.$V_W(y,x)$ is increasing in y. ■

Proof of Lemma 2. Let $\mathbb{C}$ be the space of bounded, continuous functions defined on $\mathbb{R}$. Define the operator $T(V_W(.))$ as

$$T(V_W(x, w_{t+1})) = w_t(x) + (1-\delta)(\int_x^{m_t(x)} \max[V_W(x', w_{t+1}), V_M(x', w_{t+1})]d\mathcal{L}(x'|x, m_t(x)))\}$$

It can be shown that the contraction defined above preserves boundedness and continuity and therefore $T(V_W(.)) : C \to C$. , It is relatively simple to check that Blackwell's Sufficiency Conditions are satisfied too. Hence, $V_W(x,w)$ exists.

Let $V_W(x,w')$ and $V_M(x,w')$ be increasing in x. Define

$$h(x) = \max[V_W(x,w), V_M(x,w)]$$

Then $h(x)$ is an increasing function of x. Pick x_1 and x_2 such that $x_1 > x_2$. As we prove later, this implies that $m(x_1) > m(x_2)$. From the assumption about the learning technology, this implies,

$$\int_{x_1}^{m(x_1)} h(x')d\mathcal{L}(x'|x_1, m(x_1)) > \int_{x_1}^{m(x_2)} h(x')d\mathcal{L}(x'|x_1, m(x_2)) > \int_{x_2}^{m(x_2)} h(x')d\mathcal{L}(x'|x_2, m(x_2))$$

The above inequality, combined with $w_t'(x) > 0$, implies that

$$T(V_W(x_1, w)) > T(V_W(x_2, w))$$

i.e. the contraction is increasing in x. Hence it must be the case that the fixed point is increasing in x too, i.e. $V_W(x)$ is increasing in x. ■

Proof of Lemma 3. Equilibrium in the labour market implies that

$$\int_{\underline{k}}^{k^*} \psi(s)ds = \int_{k^*}^{\overline{k}} n(m^{-1}(s))\psi(s)ds$$

where the LHS is the supply of workers while the RHS is the demand for workers. Define

$$L(k^*) = \int_{\underline{k}}^{k^*} \psi(s)ds - \int_{k^*}^{\overline{k}} n(m^{-1}(s))\psi(s)ds$$

Now $L(\underline{k}) = -\int_{\underline{k}}^{\overline{k}} n(m^{-1}(s))\psi(s)ds < 0$ while $L(\overline{k}) = \int_{\underline{k}}^{\overline{k}} \psi(s)ds > 0$. Moreover, $\frac{\partial L(k^*)}{\partial k^*} = [1 + n(m^{-1}(k^*)]\psi(k^*) > 0$. Hence, by the Intermediate Value Theorem, $\exists$ a unique k^* such that $L(k^*) = 0$. ■

Proof of Proposition 1. Suppose P is monotone, has the Feller property and satisfies a mixing condition. Then P has a unique, invariant probability measure Ψ^*.(Stokey, Lucas with Prescott, 1989). Define the operator T as

$$(Tf)(k) = \int f(k')P(k, dk'), \qquad \text{all } k \in [\underline{k}, \overline{k}]$$

where $f : [\underline{k}, \overline{k}] \to \mathbb{R}$ is a bounded function. If f is non-decreasing, then the first-order stochastic dominance property of the learning distribution implies that Tf is also non-decreasing. (Monotone Property) It is straight-forward to verify that if f is bounded and continuous, then the same holds for Tf, i.e. $T : C(k) \to C(k)$ (Feller Property). The mixing condition requires that $\exists c \in [\underline{k}, \overline{k}]$, $\epsilon > 0$ and $N \geq 1$ such that $P^N(\underline{k}, [c, \overline{k}]) \geq \epsilon$ and $P^N([\underline{k}, c], \overline{k}]) \geq \epsilon$. Choose $k' \in [\underline{k}, \overline{k}]$. Define $\epsilon_1 = \int_{[k', \overline{k}]} d\Psi_N(s)$ and $\epsilon_2 = \int_{[\underline{k}, k']} d\Psi_N(s)$. By the assumption on $\Psi_N(.)$, we know that both these objects are greater than 0. Choose $\epsilon = \delta \min\{\epsilon_1, \epsilon_2\}$ and $N = 1$. Then $P(\underline{k}, [k', \overline{k}]) \geq \epsilon$ and $P([\underline{k}, k'], \overline{k}]) \geq \epsilon$. Therefore all the conditions for the existence and uniqueness of the invariant distribution are satisfied. ■

Deriving the equilibrium conditions. Since k^* must be indifferent between being a worker and a manager, we must have $V_W(k^*, w) = V_M(k^*, w)$. Furthermore, for k^* to be the threshold, it must be the case that $\overline{k}$ can not hire $k^* + \epsilon$ and be strictly better-off. If $k^* + \epsilon$ is a manager, he earns $V_M(k^* + \epsilon)$. In order to hire $k^* + \epsilon$, the manager has to pay him a wage such that he is just indifferent between being a manager and a worker. Let this wage be ω. ω should satisfy

$$\omega + (1 - \delta) \int V_M(k) d\mathcal{L}(k|, k^* + \epsilon, \overline{k}) = V_M(k^* + \epsilon)$$

Therefore, period profit of $\overline{k}$ if he hires $k^* + \epsilon$ is given by

$$\begin{aligned}\pi_{k^*+\epsilon}(\overline{k}) &= (\sigma f(\overline{k}) - \omega) n(k^* + \epsilon) \\ &= \sigma f(\overline{k}) n(k^* + \epsilon) - n(k^* + \epsilon)(V_M(k^* + \epsilon) - (1 - \delta) \int V_M(k) d\mathcal{L}(k|, k^* + \epsilon, \overline{k}))\end{aligned}$$

For k^* to be a threshold equilibrium, it must be the case

$$\underset{\epsilon \to 0}{lim} \frac{\partial \pi_{k^*+\epsilon}(\overline{k})}{\partial \epsilon} \leq 0$$

Now,

$$\begin{aligned}\lim_{\epsilon \to 0} \frac{\partial \pi_{k^*+\epsilon}(\overline{k})}{\partial \epsilon} = &\sigma f(\overline{k}) n'(k^*) - n(k^*)(V_M'(k^*) - (1 - \delta) \frac{\partial}{\partial k^*}(\int V_M(k) d\mathcal{L}(k|, k^*, \overline{k}))) \\ &- n'(k^*)(V_M(k^*) - (1 - \delta)(\int V_M(k) d\mathcal{L}(k|, k^*, \overline{k})))\end{aligned}$$

From the manager's profit-maximizing problem, we have

$$\sigma f(\overline{k}) n'(k^*) = w'(k^*) n(k^*) + w(k^*) n'(k^*)$$

Also, for a worker with knowledge k^*,

$$V_W(k^*) = w(k^*) + (1 - \delta)(\int V_M(k) d\mathcal{L}(k|k^*, \overline{k}))$$

Since k^* is the threshold, $\max[V_W(k), V_M(k)] = V_M(k)$ $\forall k \geq k^*$. Differentiating w.r.t. k^*

$$V_W'(k^*) = w'(k^*) + (1 - \delta) \frac{\partial}{\partial k^*}(\int V_M(k) d\mathcal{L}(k|k^*, \overline{k}))$$

Replacing in (A) and usign the fact that $V_W(k^*) = V_M(k^*)$, we have

$$\lim_{\epsilon \to 0} \frac{\partial \pi_{k^*+\epsilon}(\overline{k})}{\partial \epsilon} = w'(k^*) - (V_M'(k^*) - V_W'(k^*) + w'(k^*))$$
$$= V_W'(k^*) - V_M'(k^*)$$

where we use the fact that $V_W(k^*, w) = V_M(k^*, w)$. Hence $\underset{\epsilon \to 0}{lim} \frac{\partial \pi_{k^*+\epsilon}(\overline{k})}{\partial \epsilon} < 0$ implies that

$$V_W'(k^*) < V_M'(k^*)$$

■

Proof of Proposition 2. We shall prove this proposition in a slightly different way. First we prove the existence of the threshold equilibrium, assuming that the equilibium is unique. Then we show that the suficient condition for existence is also sufficient for uniqueness.

By assuming uniqueness, we are basically assuming that the set of workers and managers has to be connected in equilibrium (See AGR). Given that there exists a unique market-clearing threshold k^* (Lemma 5), we check whether it satisfies the equilibrium condition. Here we have

$$V_M(k^*) = \frac{1}{\delta}(\sigma f(k^*) - w(\underline{k}))n(\underline{k})$$

Using the Envelope Theorem,

$$V_M'(k^*) = \frac{1}{\delta}\sigma f'(k^*)n(\underline{k})$$

Also,

$$V_W'(k^*) = w'(k^*) + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)d\mathcal{L}(k|k^*, \overline{k}))$$
$$= \frac{(\sigma f(\overline{k}) - w(k^*))n'(k^*)}{n(k^*)} + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)d\mathcal{L}(k|k^*, \overline{k}))$$

where the second line follows from the manager's profit-maximization condition. Therefore, for k^* to be an equilibrium, it must be the case that

$$\frac{(\sigma f(\overline{k}) - w(k^*))n'(k^*)}{n(k^*)} + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)d\mathcal{L}(k|k^*, \overline{k})) \leq \frac{1}{\delta}\sigma f'(k^*)n(\underline{k})$$

If $\delta = 1$, this condition reduces to

$$\frac{(\sigma f(\overline{k}) - w(k^*))n'(k^*)}{n(k^*)} \leq \sigma f'(k^*)n(\underline{k})$$

Since $w(k^*) > 0$, for the above inequality to hold, we need to find the conditions under which $\frac{f(\overline{k})n'(k^*)}{n(k^*)} \leq f'(k^*)n(\underline{k})$, or $f(\overline{k})n'(k^*) \leq f'(k^*)n(\underline{k})$, since $n(k^*) \geq 1$.

But $f(\overline{k})n'(k^*) \leq f(\overline{k})n'(\underline{k})$ ($\because n''(.) \leq 0$) and $f'(k^*)n(\underline{k}) \geq f'(\underline{k})n(\underline{k})$ ($\because n''(.) \geq 0$). Hence it follows that

$$f(\overline{k})n'(k^*) \leq f(\overline{k})n'(\underline{k}) \leq f'(\underline{k})n(\underline{k}) \leq f'(k^*)n(\underline{k})$$

where the inequality in the middle follows from Assumption 3. Thus for $\delta = 1$, the conditon on technology is sufficient for an equilibrium. But when $\delta \neq 1$, we need to determine the magnitude of $\frac{\partial}{\partial k^*}(\int V_M(k)d\mathcal{L}(k|k^*, \overline{k}))$, since this term is positive by assumption on the learning technology. Now this term is endogeneous and it depends on the invariant distribution, which in turn is determined by the learning distribution. This term is bounded above, since the domain is compact. Hence by the Least Upper Bound Property, the supremum exists. Let

$$\zeta = \sup\{\frac{\partial}{\partial k^*}(\int V_M(k)d\mathcal{L}(k|k^*, \overline{k}))\}$$

Define δ^* as the value of δ that satisfies

$$f(\overline{k})n'(\underline{k}) + (1-\delta^*)\zeta = f'(\underline{k})n(\underline{k})$$

This can be re-written as

$$\frac{n'(\underline{k})}{n(\underline{k})} + (1-\delta^*)\frac{\zeta}{f(\overline{k})n(\underline{k})} = \frac{f'(\underline{k})}{f(\overline{k})}$$

The fact that $\frac{n'(\underline{k})}{n(\underline{k})} < \frac{f'(\underline{k})}{f(\overline{k})}$ implies that $\delta^* < 1$. Hence $\forall \delta \in [\delta^*, 1]$, we have

$$f(\overline{k})n'(\underline{k}) + (1-\delta)\zeta \leq f'(\underline{k})n(\underline{k})$$

Thus

$$\begin{aligned} f(\overline{k})n'(k^*) + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)d\mathcal{L}(k|k^*, \overline{k})) &\leq f(\overline{k})n'(\underline{k}) + (1-\delta)\zeta \\ &\leq f'(\underline{k})n(\underline{k}) \\ &\leq f'(k^*)n(\underline{k}) \end{aligned}$$

$\delta \leq 1$ implies that $f'(k^*)n(\underline{k}) \leq \frac{1}{\delta}f'(k^*)n(\underline{k})$. Therefore

$$f(\overline{k})n'(k^*) + (1-\delta)\frac{\partial}{\partial k^*}(\int V_M(k)d\mathcal{L}(k|k^*, \overline{k})) \leq \frac{1}{\delta}f'(k^*)n(\underline{k})$$

This completes our proof about the existence of equilibrium. As mentioned before, showing uniqueness entails showing that the set of workers and managers is connected. Suppose not. WLOG let's assume that the knowledge distribution has the following partition - $([1, k_1], [k_1, k_2], [k_2, k_3], [k_3, k_4])$. Workers in $[\underline{k}, k_1]$ work for managers in $[k_1, k_2]$ while workers in $[k_2, k_3]$ work for managers in $[k_3, k_4]$. For this to be an equilibrium, it must be the case that k_2 must be indifferent between being a worker and a manager. In other words, a deviation involving k_3 hiring $k_2 - \epsilon$ should not make both k_3 and $k_2 - \epsilon$ better off. Using a similar logic as developed above, one can show that that the condition for equilibrium is $V'_W(k_2) > V'_M(k_2)$. One can then show that if $\frac{n'(\underline{k})}{n(\underline{k})} < \frac{f'(\underline{k})}{f(\bar{k})}$, then for δ high enough, this conditon will always be violated. Therefore, an allocation with disconnected sets of workers and managers can never be sustained as an equilibrium implying that the only equilibrium is the threshold equilibrium.

To prove efficiency, we follow Legros and Newman (2002). Let us define $\lambda(k)$ as the value that the social planner attaches to knowledge k. In this set-up, when a manager with knowledge y and workers with knowledge x get together, they produce $f(y)n(x) + n(x)[\lambda(E(x'|x, y) - \lambda(x)]$, where $E(x'|x, y)$ is the expected knowledge that workers acquire. Let us consider the no-learning case first. In the absence of learning, the total value produced is simply $f(y)n(x)$. It is simple to show that complementarity implies that the planner uses Positive Assortative Matching. We need to show that the planner maximizes value from threshold matching. We prove this by contradiction. Supppose there's segregation, i.e. the set of workers and managers is disconnected. In this case, one can always find $k_1 < k_2 < k_2 + \epsilon < k_3$ such that k_1 works for k_2 and $k_2 + \epsilon$ works for $k_3(\epsilon)$, i.e. k_2 is a threshold. Since the planner is maximizing value, this implies that the planner can not do better my switching the team members, i.e. it must not be the case that

$$f(k_2 + \epsilon)n(k_1) + f(k_3(\epsilon))n(k_2) > f(k_2)n(k_1) + f(k_3(\epsilon))n(k_2 + \epsilon)$$

Re-arranging, we have

$$n(k_1)(f(k_2 + \epsilon) - f(k_2)) > f(k_3(\epsilon))(n(k_2 + \epsilon) - n(k_2))$$

Dividing by ϵ and taking limit as $\epsilon \to 0$, we get

$$n(k_1)f'(k_2) > f(k_3(0))n'(k_2)$$

In the above inequality, the $LHS = n(k_1)f'(k_2) > n(\underline{k})f'(k_2) \geq n(\underline{k})f'(\underline{k})$ and the $RHS = f(k_3(0))n'(k_2) < f(\bar{k})n'(k_2) \leq f(\bar{k})n'(k_1)$. Since $n(\underline{k})f'(\underline{k}) > f(\bar{k})n'(\underline{k})$, the above inequality holds. But then we get a contradiction. Hence the social planner maximizes value by threshold matching. Now we introduce learning. Larger is δ, smaller is the value of learning. Since the above inequality holds for $\delta = 1$, by continuity, it will also hold for δ large enough. Consequently, for δ large enough, threshold matching is also efficient. ■

Proof of Lemma 4. Since y is indifferent along $\widetilde{w}_{NL}(x,y)$, for any x_1 and x_2 we must have $\sigma f(y)n(x_1) - \widetilde{w}_{NL}(x_1)n(x_1) = \sigma f(y)n(x_2) - \widetilde{w}_{NL}(x_2)n(x_2)$. Letting $x_2 = x_1 + h$ and re-arranging, we have $\widetilde{w}_{NL}(x_1+h)n(x_1+h) - \widetilde{w}_{NL}(x_1)n(x_1) = \sigma f(y)n(x_1+h) - \sigma f(y)n(x_1)$. Dividing by h and taking the limit as $h \to 0$, we get $\widetilde{w}'_{NL}(x_1) = \frac{n'(x_1)(\sigma f(y) - w(x_1))}{n(x_1)}$. Since $\sigma f(y) - w(x_1) \geq 0$ and $n'(x_1) > 0$, the RHS≥ 0. Moreover, since the derivative is well-defined, it follows that $\widetilde{w}_{NL}(x_1)$ is also continuous at x_1. Since x_1 was chosen randomly, the result follows. ■

Proof of Lemma 5. Let us denote the continuation value of x, when he works for y, by $C(x,y)$. Conituity of the value functions implies that $C(x,y)$ is continuous too. We shall prove this lemma by contradiction. Suppose that $\exists\, x'$ and y' such that $V(x') = \widetilde{V}_W(x',y')$ but $\widetilde{w}(x',y') < \max\limits_{y} \widetilde{w}(x',y) = \widetilde{w}(x',\overline{y})$ (say). Then, given the properties of $\widetilde{w}(x,y)$, we can have either of the following situations - (a) Suppose $y' < \overline{y}$. At $x = x'$, $\widetilde{w}(x',y') < \widetilde{w}(x',\overline{y})$. Moreover, from the learning technology we know that the continuation value of $x\prime$ is higher if he works for $\overline{y}$ rather than y'. But then $\widetilde{V}_W(x',y') < \widetilde{V}_W(x',\overline{y})$. Hence $V(x') \neq \widetilde{V}_W(x',y')$ and we get a contradiction. (b) Suppose $y' > \overline{y}$. Then either $\exists\, x'' < x'$ and $y'' < \overline{y}$ such that $w(x'') = \widetilde{w}(x'',y'') < \widetilde{w}(x'',\overline{y})$ in which case we are back to case (a). Or, $\exists\, x'' < x'$ such that there's a discontinuity in $w(x)$ at $x = x''$. In particular, $\lim\limits_{\epsilon\to 0} w(x''-\epsilon) = \lim\limits_{\epsilon\to 0} \widetilde{w}(x''-\epsilon,\overline{y}) > \widetilde{w}(x'',y') = w(x'')$. Since $V(x)$ must be continuous at x'', $\lim\limits_{\epsilon\to 0} C(x''-\epsilon,\overline{y}) < C(x'',y')$. Now consider the sequence $x \to x''$, $(x < x'')$. Continuity of $\widetilde{w}(.)$ implies that $\lim\limits_{\epsilon\to 0}\widetilde{w}(x''-\epsilon,y') = \widetilde{w}(x'',y')$ and $\lim\limits_{\epsilon\to 0}\widetilde{w}(x''-\epsilon,\overline{y}) = \widetilde{w}(x'',\overline{y})$. Similarly, continuity of $C(.)$ implies that $\lim\limits_{\epsilon\to 0} C(x''-\epsilon,y') = C(x'',y')$ and $\lim\limits_{\epsilon\to 0} C(x''-\epsilon,\overline{y}) = C(x'',\overline{y})$. Now $\lim\limits_{\epsilon\to 0}\widetilde{V}_W(x''-\epsilon,y') = \lim\limits_{\epsilon\to 0}\widetilde{w}(x''-\epsilon,y') + \lim\limits_{\epsilon\to 0} C(x''-\epsilon,y') = \widetilde{w}(x'',y') + C(x'',y') > \widetilde{w}(x'',\overline{y}) + C(x'',\overline{y}) = \lim\limits_{\epsilon\to 0}\widetilde{w}(x''-\epsilon,\overline{y}) + \lim\limits_{\epsilon\to 0} C(x''-\epsilon,\overline{y}) = \lim\limits_{\epsilon\to 0}\widetilde{V}_W(x''-\epsilon,\overline{y})$. But then, as $\epsilon \to 0$, $argmax \widetilde{V}_W(x''-\epsilon,y) \neq \overline{y}$ and we get a contradiction. The reverse implication can be proved in a similar fashion. ■

Proof of Proposition 3. Let the number of people being born every period be normalized to 1. Cohort t at time t are all newborns. All agents in $[\underline{k}, k^*_A]$ are workers. The measure of these agents is $\int_{\underline{k}}^{k^*_A} \phi_H(k)dk$. A worker with knowledge k demands $\frac{1}{n(k)}$ managers. Therefore, the total demand for managers by cohort t workers is $\int_{\underline{k}}^{k^*_A} \frac{\phi_H(k)}{n(k;\beta)}dk$.The supply of cohort t managers is simply the measure of agents in $[k^*_A, \overline{k}_H]$. This is given by $\int_{k^*_A}^{\overline{k}_H} \phi_H(k)dk$. Let us consider the distribution of cohort $t-1$ agents at time t. A fraction $1-\delta$ of every type of agent in $[\underline{k}, k^*_A]$ survive in period t. Out of the ones that survive, a fraction π of every type of agent do not learn and remain where they are. Hence, the total demand for managers by cohort $t-1$ workers is $\int_{\underline{k}}^{k^*_A} \frac{\pi(1-\delta)\phi_H(k)}{n(k;\beta)}dk$. Similarly, a fraction $1-\delta$ of the cohort $t-1$ managers in $[k^*_A, \overline{k}]$ survive in period t. These agents do not learn. Moreover, $(1-\theta)(1-\delta)\int_{\underline{k}}^{k^*_A} \phi_H(k)dk$ agents move into this interval from $[\underline{k}, k^*_A]$. They are the cohort $t-1$ agents who were workers in period $t-1$ but become managers in period t. Therefore, the supply of cohort $t-1$ managers is $(1-\delta)[\int_{k^*_A}^{\overline{k}_H} \phi_H(k)dk + (1-\pi)\int_{\underline{k}}^{k^*_A} \phi_H(k)dk]$. The supply and demand for managers in other cohorts can be obtained in a similar fashion. Adding up the demand for managers and the supply of managers in each cohort, we get

$$\text{Demand for managers } = \frac{1}{1-\theta(1-\delta)} \int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)} dk$$

$$\text{Supply of managers} = \frac{1}{\delta(1-\theta(1-\delta))}[[1-\theta(1-\delta)] \int_{k_A^*}^{\overline{k}_H} \phi_H(k)dk + (1-\theta)(1-\delta) \int_{\underline{k}}^{k_A^*} \phi_H(k)dk]$$

In equilibrium, supply must equal demand. Equating the above two expressions and after a bit of algebra, we obtain the following :

$$\int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)} dk = \int_{k_A^*}^{\overline{k}_H} \phi_H(k)dk + (1-\theta)(\frac{1-\delta}{\delta})$$

In order to derive the properties of k_A^*, we use the Implicit Function Theorem. Differentiating the above equation w.r.t. π,

$$\frac{\partial k_A^*}{\partial \pi} \frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} = -\frac{\partial k_A^*}{\partial \pi} \phi_H(k_A^*) - (\frac{1-\delta}{\delta})$$

Therefore,

$$\frac{\partial k_A^*}{\partial \pi} [\frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} + \phi_H(k_A^*)] = -(\frac{1-\delta}{\delta})$$

Since the LHS is positive while the RHS is negative, $\frac{\partial k_A^*}{\partial \pi} < 0$. In a similar fashion it can be shown that $\frac{\partial k_A^*}{\partial \delta} < 0$.

Differentiating the labour market clearing condition w.r.t. β,

$$\frac{\partial k_A^*}{\partial \beta} \frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} + \int_{\underline{k}}^{k_A^*} \frac{-\phi_H(k)}{n(k;\beta)^2} \frac{\partial n(k;\beta)}{\partial \beta} dk = -\frac{\partial k_A^*}{\partial \beta} \phi_H(k_A^*)$$

Re-arranging terms, we have

$$\frac{\partial k_A^*}{\partial \beta} [\frac{\phi_H(k_A^*)}{n(k_A^*;\beta)} + \phi_H(k_A^*)] = \int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)^2} \frac{\partial n(k;\beta)}{\partial \beta} dk$$

Since $\frac{\partial n(k;\beta)}{\partial \beta} > 0$, both the LHS and the RHS are positive. Therefore, $\frac{\partial k_A^*}{\partial \beta} > 0$ ■

Proof of Lemma 6. Let G f.o.s.d. H. Let g and h be the corresponding densities. Also, let $\xi(k)$ be the demand for manager per worker, where the worker has knowledge k. Since the span of control is only a function of the worker's knowledge, a worker with knowledge k works in a firm of size $n(k)$. Hence $\xi(k)$ is simply the reciprocal of $n(k)$. Therefore $\xi'(k) < 0$ (this follows from $n'(k) > 0$). Also, let k^* be the threshold under H.

We shall prove the lemma by contradiction. Let k^* also be the threshold for G. We can have two possibilities - (i) $g(k) < h(k)$ for all $k < k^*$. In this case, the demand for managers under $G = \int_{\underline{k}}^{k^*} \xi(k)g(k)dk < \int_{\underline{k}}^{k^*} \xi(k)h(k)dk =$ demand for managers under H. But the supply of managers under $G = 1 - G(k^*) > 1 - H(k^*) =$supply of managers under H. Hence at k^*, there's an excess supply of managers under G. This means that the threshold for G must be greater than k^*. (ii) There are n intervals $A_i \subset [\underline{k}, k^*]$, $i = 1, ...n$ such that

$$g(k) > h(k) \forall k \in A_i, \forall i$$

$$g(k) < h(k) \text{ otherwise}$$

. Rank the $A_i s$ such that $A_i < A_j \Rightarrow maxA_i < minA_j$. We proceed as follows - We know that $\underline{k} < minA_1 = a_1(say)$ (otherwise H would f.o.s.d. G). Let $B = [\underline{k}, a_1]$. Then it must be the case that $g(k) < h(k)$ for all $k \in B$. G f.o.s.d. H implies that

$$\int_B g(k)dk + \int_{A_1} g(k)dk < \int_{A_1} h(k)dk + \int_B h(k)dk$$

Re-arranging the above equation,

$$\int_{A_1} [g(k) - h(k)]dk < \int_B [h(k) - g(k)]dk$$

Multiplying both sides by $\xi(a_1)$,

$$\int_{A_1} \xi(a_1)[g(k) - h(k)]dk < \int_B \xi(a_1)[h(k) - g(k)]dk$$

Now, $\xi'(k) < 0$ implies that $\xi(a_1) < \xi(k) \forall k \in B$ and $\xi(a_1) > \xi(k) \forall k \in A_1$. Replacing $\xi(a_1)$ in the above equation,

$$\int_{A_1} \xi(k)[g(k) - h(k)]dk < \int_B \xi(k)[h(k) - g(k)]dk$$

Here we are using the fact that $h(k) - g(k) > 0 \forall k \in B$ and $g(k) - h(k) > 0 \forall k \in A_1$. We re-arrange again to obtain

$$\int_B \xi(k)g(k)dk + \int_{A_1} \xi(k)g(k)dk < \int_{A_1} \xi(k)h(k)dk + \int_B \xi(k)h(k)dk$$

The LHS and the RHS are the demand for managers by workers in $B \cup A_1$ under G_1 and G_2 respectively. Define $maxA_1 = a_1'$ and $minA_2 = a_2$. Let $C = [a_1', a_2]$. G f.o.s.d. H implies that

$$\int_B g(k)dk+\int_{A_1} g(k)dk+\int_C g(k)dk+\int_{A_2} g(k)dk < \int_B h(k)dk+\int_{A_1} h(k)dk+\int_C h(k)dk+\int_{A_2} h(k)dk$$

Re-arranging, we have

$$\int_{A_2}[g(k)-h(k)]dk < (\int_B[h(k)-g(k)]dk - \int_{A_1}[g(k)-h(k)]dk) + \int_C[h(k)-g(k)]dk$$

Multiplying both sides by $\xi(a_2)$,

$$\int_{A_2}\xi(a_2)[g(k)-h(k)]dk < \xi(a_2)(\int_B[h(k)-g(k)]dk - \int_{A_1}[g(k)-h(k)]dk) + \int_C \xi(a_2)[h(k)-g(k)]dk$$

Since $\xi'(k) < 0$, we have

$$\begin{aligned}\int_{A_2}\xi(k)[g(k)-h(k)]dk < \xi(a_2)(\int_B[h(k)-g(k)]dk - \int_{A_1}[g(k)-h(k)]dk) \\ + \int_C \xi(k)[h(k)-g(k)]dk\end{aligned}$$

Again, using $\xi'(k) < 0$ in the above inequality

$$\begin{aligned}LHS < \xi(a_1)(\int_B[h(k)-g(k)]dk - \int_{A_1}[g(k)-h(k)]dk) \\ + \int_C \xi(k)[h(k)-g(k)]dk \\ \int_B \xi(k)[h(k)-g(k)]dk - \int_{A_1}\xi(k)[g(k)-h(k)]dk \\ + \int_C \xi(k)[h(k)-g(k)]dk\end{aligned}$$

Re-arranging gives us that the demand for managers by workers in $B \cup A_1 \cup C \cup A_2$ under G is less than that under H. We can repeat this argument by expanding the set till we reach k^*. But then we have shown that the demand for managers under G is less than that under H. However the supply of managers under G is greater than that under H. Therefore, at k^*, there is an excess supply of managers under G. Hence the threshold under G has to be greater than k^*. ■

Proof of Proposition 4. The derivation of the threshold is the same as in the proof of Proposition

3. Equating the supply of managers and the demand for managers, we have

$$\int_{\underline{k}}^{k_I^*} \frac{\phi_H(k) + \phi_F(k)}{2n(k;\beta)} dk = \int_{k_I^*}^{\overline{k}_F} \frac{\phi_H(k) + \phi_F(k)}{2} dk + (1-\theta)(\frac{1-\delta}{\delta})$$

Re-arranging, we have

$$\int_{\underline{k}}^{k_I^*} \frac{\phi_H(k)}{n(k;\beta)} dk - \int_{k_I^*}^{\overline{k}_H} \phi_H(k) dk - (1-\theta)(\frac{1-\delta}{\delta}) = \int_{k_I^*}^{\overline{k}_H} \phi_F(k) dk + (1-\theta)(\frac{1-\delta}{\delta}) - \int_{\underline{k}}^{k_I^*} \frac{\phi_F(k)}{n(k;\beta)} dk$$

Not that the LHS is the excess demand for managers in the Home country if the threshold is k_I^*, while the RHS is the excess supply of managers in the Foreign country if the threshold is k_I^*. If $k_I^* = k_A^*$, the LHS is equal to 0, i.e.

$$\int_{\underline{k}}^{k_A^*} \frac{\phi_H(k)}{n(k;\beta)} dk - \int_{k_A^*}^{\overline{k}_H} \phi_H(k) dk = (1-\theta)(\frac{1-\delta}{\delta})$$

Since $\phi_F(k)$ f.o.s.d. $\phi_H(k)$, from Lemma 6, we know that

$$\int_{\underline{k}}^{k_A^*} \frac{\phi_F(k)}{n(k;\beta)} dk - \int_{k_A^*}^{\overline{k}_H} \phi_F(k) dk < (1-\theta)(\frac{1-\delta}{\delta})$$

Therefore for $k_I^* = k_A^*$, the RHS is positive. But this means that $k_I^* \neq k_A^*$. In particular, since the LHS is increasing in k_I^* and the RHS is decreasing, it must be the case that $k_I^* > k_A^*$ ■

Proof of Proposition 5. We know that an allocation A is a Pareto improvement over allocation B if $u(x_i^A) \geq u(x_i^B)$ for all i, and $u(x_j^A) > u(x_j^B)$ for some j. This suggests that in order to show that A *is not* a Pareto improvement over B, it is sufficient to show that ∃ individuals 1 and 2 s.t. $u(x_1^A) \geq u(x_1^B) \Rightarrow u(x_2^A) < u(x_2^B)$ and vice versa. From Lemma 6, we have

$$k_{A,NL}^* < k_{I,NL}^*$$

where NL refers to no-learning. If there are incumbent firms in the Home economy, this also means that

$$k_{I,NL}^* < \overline{k}_H$$

The above inequality suggests that under Integration, there are incumbent Home managers who continue to operate ($k \in [k_{I,NL}^*, \overline{k}_H]$). At the same time, under Autarky, $m_{A,NL}^{-1}(k_{A,NL}^*) = \underline{k} \Rightarrow$

$m^{-1}_{A,NL}(k^*_{I,NL}) > \underline{k}$ (follows from PAM). While under Integration, $m^{-1}_{A,NL}(k^*_{I,NL}) = \underline{k}$, i.e. under Integration, the manager with knowledge $k^*_{I,NL}$ has a worse match. The present value of $k^*_{I,NL}$ is just the period profits $\pi_{I,NL}(k^*_{I,NL})$ divided by δ.

$$\begin{aligned}\pi_{A,NL}(k^*_{I,NL}) &= (\sigma f(k^*_{I,NL}) - w_{A,NL}(m^{-1}_{A,NL}(k^*_{I,NL})))n(m^{-1}_{A,NL}(k^*_{I,NL}))\\ &\geq (\sigma f(k^*_{I,NL}) - w_{A,NL}(k))n(k) \qquad \forall k\\ &\geq (\sigma f(k^*_{I,NL}) - w_{A,NL}(\underline{k}))n(\underline{k})\end{aligned}$$

Note that $\pi_{I,NL}(k^*_{I,NL}) = (\sigma f(k^*_{I,NL}) - w_{I,NL}(\underline{k}))n(\underline{k})$. Therefore the relation between $\pi_{A,NL}(k^*_{I,NL})$ and $\pi_{I,NL}(k^*_{I,NL})$ depends on the relation between $w_{A,NL}(\underline{k})$ and $w_{I,NL}(\underline{k})$. Let us consider the following cases -

(a) $w_{A,NL}(\underline{k}) < w_{I,NL}(\underline{k})$: In this case, $\pi_{A,NL}(k^*_{I,NL}) > \pi_{I,NL}(k^*_{I,NL}) \Rightarrow \underline{k}$ is strictly better-off under Integration but $k^*_{I,NL}$ is strictly worse-off

(b) $w_{A,NL}(\underline{k}) \gg w_{I,NL}(\underline{k})$: Then it is possible to have, $\pi_{A,NL}(k^*_{I,NL}) < \pi_{I,NL}(k^*_{I,NL}) \Rightarrow k^*_{I,NL}$ is strictly better-off under Integration but $\underline{k}$ is strictly worse-off

(c) $w_{A,NL}(\underline{k}) = w_{I,NL}(\underline{k})$: In this case, $\pi_{A,NL}(k^*_{I,NL}) \geq \pi_{I,NL}(k^*_{I,NL})$. This is not a negation of Pareto improvement. However let us choose the agent with knowledge $k^*_{I,NL} + \epsilon$ such that $m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon) < m^{-1}_{A,NL}(k^*_{I,NL} + \epsilon)$. Since $m(.)$ is continuous, we can always find such an ϵ. Moreover, since $m(.)$ is a function, its inverse must be strictly monotonic. Hence $m^{-1}_{I,NL}(k^*_{I,NL}+\epsilon) > m^{-1}_{I,NL}(k^*_{I,NL}) = \underline{k}$. Now

$$w'_{A,NL}(\underline{k}) = \frac{(\sigma f(k^*_{A,NL}) - w_{A,NL}(\underline{k}))n'(\underline{k})}{n(\underline{k})} < \frac{(\sigma f(k^*_{I,NL}) - w_{I,NL}(\underline{k}))n'(\underline{k})}{n(\underline{k})} = w'_{I,NL}(\underline{k})$$

Combined with $w_{A,NL}(\underline{k}) = w_{I,NL}(\underline{k})$, this means that in the neighbourhood of $k = \underline{k}$, $w_{A,NL}(k) < w_{I,NL}(k)$. Hence

$$\begin{aligned}\pi_{A,NL}(k^*_{I,NL} + \epsilon) &= (\sigma f(k^*_{I,NL} + \epsilon) - w_{A,NL}(m^{-1}_{A,NL}(k^*_{I,NL} + \epsilon)))n(m^{-1}_{A,NL}(k^*_{I,NL} + \epsilon))\\ &\geq (\sigma f(k^*_{I,NL} + \epsilon) - w_{A,NL}(m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon)))n(m^{-1}_{I,NL}(k^*_{I,NL} + \epsilon))\end{aligned}$$

Using the fact that $w_{A,NL}(k^*_{I,NL} + \epsilon) < w_{I,NL}(k^*_{I,NL} + \epsilon)$, we have

$$\pi_{A,NL}(k^*_{I,NL}+\epsilon) > (\sigma f(k^*_{I,NL}+\epsilon) - w_{I,NL}(m^{-1}_{I,NL}(k^*_{I,NL}+\epsilon)))n(m^{-1}_{I,NL}(k^*_{I,NL}+\epsilon)) = \pi_{I,NL}(k^*_{I,NL}+\epsilon)$$

Therefore $k^*_{I,NL} + \epsilon$ is strictly worse-off.

Hence, for all the 3 cases (a), (b) and (c), we have shown that atleast one individual is worse-off. Since these cases are exhaustive, the result follows. ■

Proof of Proposition 6. We proceed as follows - First we shall find the condition under which

$\overline{k}_H$ is better-off under Integration. Since $\overline{k}_H$ is a manager under both Autarky and Integration, we have to show that $V_{M,I}(\overline{k}_H) > V_{M,A}(\overline{k}_H)$. Since $\overline{k}_H$ is matched with k_A^* under Autarky and $\widetilde{k}$ under Integation, this implies that

$$[\sigma f(\overline{k}_H) - w_A(k_A^*)]n(k_A^*) < [\sigma f(\overline{k}_H) - w_I(\widetilde{k})]n(\widetilde{k})$$

Re-arranging, we have

$$\sigma f(\overline{k}_H)[n(k_A^*) - n(\widetilde{k})] < w_A(k_A^*)n(k_A^*) - w_I(\widetilde{k})n(\widetilde{k})$$

Now,

$$\begin{aligned} w_A(k_A^*)n(k_A^*) - w_I(\widetilde{k})n(\widetilde{k}) = & \frac{\delta + (1-\theta)(1-\delta)}{\delta}[\sigma f(k_A^*)n(\underline{k})\frac{n(k_A^*)}{1+n(k_A^*)} - \sigma f(k_I^*)n(\underline{k})\frac{n(k_I^*)}{1+n(k_I^*)}] \\ & + \frac{(1-\theta)(1-\delta)}{\delta}[f(\overline{k}_F)n(k_I^*)\frac{n(k_I^*)}{1+n(k_I^*)} - f(\overline{k}_H)n(k_A^*)\frac{n(k_A^*)}{1+n(k_A^*)}] \\ & + \frac{\delta + (1-\theta)(1-\delta)}{\delta}[\frac{n(k_A^*)}{1+n(k_A^*)}\int_{\underline{k}}^{k_A^*} f(m_A(k))n'(k)dk \end{aligned}$$

$$+\frac{n(k_I^*)}{1+n(k_I^*)}\int_{\widetilde{k}}^{k_I^*} f(m_I(k))n'(k)dk] - \int_{\underline{k}}^{\widetilde{k}} f(m_I(k))n'(k)dk$$

Let us consider each term on the RHS.

$$\sigma f(k_A^*)n(\underline{k})\frac{n(k_A^*)}{1+n(k_A^*)} - \sigma f(k_I^*)n(\underline{k})\frac{n(k_I^*)}{1+n(k_I^*)} > \sigma f(\underline{k})n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - \sigma f(\overline{k}_H)n(\underline{k})\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$$

$$f(\overline{k}_F)n(k_I^*)\frac{n(k_I^*)}{1+n(k_I^*)} - f(\overline{k}_H)n(k_A^*)\frac{n(k_A^*)}{1+n(k_A^*)} > f(\overline{k}_F)n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - f(\overline{k}_H)n(\overline{k}_H)\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$$

$$-\int_{\underline{k}}^{\widetilde{k}} f(m_I(k))n'(k)dk > -f(\overline{k}_H)[n(\overline{k}_H) - n(\underline{k})]$$

$$\frac{n(k_A^*)}{1+n(k_A^*)}\int_{\underline{k}}^{k_A^*} f(m_A(k))n'(k)dk > 0, \quad \frac{n(k_I^*)}{1+n(k_I^*)}\int_{\widetilde{k}}^{k_I^*} f(m_I(k))n'(k)dk > 0$$

Replacing them in the above equation,

$$
\begin{aligned}
w_A(k_A^*)n(k_A^*) - w_I(\widetilde{k})n(\widetilde{k}) >& \frac{\delta+(1-\theta)(1-\delta)}{\delta}[\sigma f(\underline{k})n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - \sigma f(\overline{k}_H)n(\underline{k})\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}] \\
&+\frac{(1-\theta)(1-\delta)}{\delta}[f(\overline{k}_F)n(\underline{k})\frac{n(\underline{k})}{1+n(\underline{k})} - f(\overline{k}_H)n(\overline{k}_H)\frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}] \\
&-f(\overline{k}_H)[n(\overline{k}_H)-n(\underline{k})] \\
=& A(say)
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
f(\overline{k}_H)[n(k_A^*) - n(\widetilde{k})] &< f(\overline{k}_H)[n(\overline{k}_H)-n(\underline{k})] \\
&= B(say)
\end{aligned}
$$

Hence the sufficient condition for $\overline{k}_H$ to be strictly better-off uder Integration is that $A > B$. After a bit of algebra, this condition reduces to

$$
(1-\theta)(\frac{1-\delta}{\delta}) > \frac{2f(\overline{k}_H)[n(\overline{k}_H)-n(\underline{k})]+n(\underline{k})[\mu(\overline{k}_H)f(\overline{k}_H)-\mu(\underline{k})f(\underline{k})]}{\mu(\underline{k})[f(\overline{k}_F)+f(\underline{k})]n(\underline{k})-\mu(\overline{k}_H)[n(\underline{k})+n(\overline{k}_H)]f(\overline{k}_H)}
$$

where $\mu(\underline{k}) = \frac{n(\underline{k})}{1+n(\underline{k})}$ and $\mu(\overline{k}_H) = \frac{n(\overline{k}_H)}{1+n(\overline{k}_H)}$. Of course, this only ensures that $\overline{k}_H$ is strictly better off. We need to show that every Home agent can be made better off.

Notice that for $k \in [\underline{k}, k_A^*]$, agents are workers under both regimes. For $k \in [k_A^*, k_I^*]$, agents are workers under Integration but managers under Autarky. Finally for $k \in [k_I^*, \overline{k}_H]$, agents are managers under both regimes. In the steady-state, $V_{M,i}(k) = \frac{1}{\delta}\pi_i(k)$, $i \in \{A, I\} \Rightarrow V'_{M,i}(k) = \frac{1}{\delta}\pi'_i(k) = \frac{1}{\delta}f'(k)n(m_i^{-1}(k))$. For $k \in [k_I^*, \overline{k}_H]$,

$$
\begin{aligned}
m_I^{-1}(k) < m_A^{-1}(k) &\Rightarrow \frac{1}{\delta}f'(k)n(m_I^{-1}(k)) < \frac{1}{\delta}f'(k)n(m_A^{-1}(k)) \\
&\Rightarrow V'_{M,I}(k) < V'_{M,A}(k)
\end{aligned}
$$

Suppose $V_{M,I}(\overline{k}_H) > V_{M,A}(\overline{k}_H)$. Since $V_{M,A}(.)$ is decreasing at a faster rate than $V_{M,I}(.)$ in the neighbourhood $[k_I^*, \overline{k}_H]$, this implies that $V_{M,I}(k) > V_{M,A}(k)$ for $k \in [k_I^*, \overline{k}_H]$. In particular, $V_{M,I}(k_I^*) > V_{M,A}(k_I^*)$. For $k \in [k_A^*, k_I^*]$,

$$
V_{W,I}(k) = \frac{1}{\delta+(1-\theta)(1-\delta)}w_I(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta+(1-\theta)(1-\delta))}\sigma f(m_I(k))n(k)
$$

$$\Rightarrow V'_{W,I}(k) = \frac{1}{\delta + (1-\theta)(1-\delta)}\sigma f(m_I(k))n'(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta + (1-\theta)(1-\delta))}\sigma f'(m_I(k))n(k)m_I^{'}(k)$$

Also, $V'_{M,A}(k) = \frac{1}{\delta}f'(k)n(m_A^{-1}(k))$. When $\delta = 1$, $V'_{W,I}(k) = f(m_I(k))n'(k)$ and $V'_{M,A}(k) = f'(k)n(m_A^{-1}(k))$. Now,

$$\frac{f'(k)}{f(m_I(k))} \geq \frac{f'(\underline{k})}{f(m_I(k))} > \frac{f'(\underline{k})}{f(\overline{k}_H)} > \frac{n'(\underline{k})}{n(\underline{k})} \geq \frac{n'(m_A^{-1}(k))}{n(m_A^{-1}(k))} > \frac{n'(k)}{n(m_A^{-1}(k))}$$

Hence $V'_{M,A}(k) > V'_{W,I}(k)$. Therefore, $\exists\ \delta_1$ s.t. $\forall \delta > \delta_1$, $V'_{M,A}(k) > V'_{W,I}(k)$ and hence $V_{W,I}(k) > V_{M,A}(k)$ for $k \in [k_A^*, k_I^*]$. In particular, $V_{W,I}(k_A^*) > V_{M,A}(k_A^*)$. For $k \in [\underline{k}, k_A^*]$,

$$V'_{W,A}(k) = \frac{1}{\delta + (1-\theta)(1-\delta)} f(m_A(k))n'(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta + (1-\theta)(1-\delta))} f'(m_A(k))n(k)m_A^{'}(k)$$

and

$$V'_{W,I}(k) = \frac{1}{\delta + (1-\theta)(1-\delta)} f(m_I(k))n'(k) + \frac{(1-\theta)(1-\delta)}{\delta(\delta + (1-\theta)(1-\delta))} f'(m_I(k))n(k)m_I^{'}(k)$$

When $\delta = 1$, $V'_{W,A}(k) = f(m_A(k))n'(k) > f(m_I(k))n'(k) = V'_{W,I}(k)$. Therefore, $\exists\ \delta_2$ s.t. $\forall \delta > \delta_2$, $V'_{W,A}(k) > V'_{W,I}(k)$ and hence $V_{W,I}(k) > V_{w,A}(k)$ for $k \in [k_A^*, k_I^*]$. Hence if we choose $\delta^* = \max\{\delta_1, \delta_2\}$, $\forall \delta > \delta^*$, $V_{W,I}(k) > V_{w,A}(k)$ for $k \in [\underline{k}, \overline{k}_H]$. ■

Derivation of the unit of time : In our paper, because of PAM, workers do not work for the same manager for two consecutive periods. But we can think of each period as consisting of several sub-periods. In each of these sub-periods, the worker learns a little bit and accordingly, his wage also increases gradually. This wage increase does not take place in our model. This is because the workers are implicitly tied to their managers for one period. We can think of the rising wages as what the managers would have paid if the workers could potentially leave at any point in time. We make some simplifying assumptions : Firstly, the wage growth takes place at a constant rate every year. Secondly, although wages are growing, the manager's earnings remain constant. And finally, the marginal increment to knowledge is constant over the sub-periods. The annual rate of growth of wage due to knowledge accumulation turns out to be 2% (Basically we are solving for g, where g satisfies $(1+g)^{10} = 1.25$). In our model, the difference between the earnings of the worker and the manager is only due to a difference in their level of knowledge. In the U.S., on an average, the manager's wage is a little more than twice the wage of the worker (Source : U.S. Bureau of Labor Statistics). Worker here refers to individuals engaged in all non-managerial occupations. Assuming that the wage of the worker continues to rise at the rate of 2% per year, this essentially means that the worker would catch up with the manager in around 34 years (Here we are using Assumption (ii)). With an uniform learning distribution, the expected knowledge accumulated by the worker in one period lies half way between the worker's initial knowledge and the knowledge of his manager. Assumption (iii) then implies that in order to catch up with the manager, the worker's knowledge has to rise by a constant amount every period for 34 years. Therefore, the worker will reach the half-way stage in 17 years. Hence the unit of time is 17 years.

Income groups : We have a total of 117 countries. The countries are classified according to their per capita GDP in 1970.

High-income : Switzerland, Luxembourg, United States, Denmark, Sweden, Australia, Canada, Netherlands, New Zealand, Norway, France, Germany, Belgium, United Kingdom, Japan, Finland, Austria, Italy, Iceland.

Middle-income (High) : Barbados, Argentina, Spain, Israel, Greece, Puerto Rico, Ireland, Gabon.

Middle-income (Low) : Hong Kong, South Africa, Uruguay, Portugal, Namibia, Hungary, Mexico, Cyprus, Venezuela, Trinidad &Tobago, Chile, Singapore, Peru, Poland, Costa Rica, Nicaragua, El Salvador, Panama, Mauritius, Brazil.

Low-income (High) : Turkey, Iran, Papua New Guinea, Jamaica, Guatemala, Angola, Seychelles, Zambia, Colombia, Cote d'Ivoire, Guyana, Equatorial Guinea, Fiji, Algeria, Taiwan, Guinea, Malaysia, Tunisia, Korea, Republic of, Bolivia, Paraguay, Zimbabwe, Ecuador, Philippines, Comoros, Mozambique, Romania, Dominican Republic, Syria, Egypt, Morocco, Central African Republic, Thailand, Jordan, Mauritania, Honduras.

Low-income (Low) : Ghana, Senegal, Togo, Gambia, The, Cameroon, Madagascar, Sri Lanka, Cape Verde, Sierra Leone, Botswana, Niger, Pakistan, Congo, Republic of, Chad, Benin, Bangladesh, India, Congo, Dem. Rep., Guinea-Bissau, Kenya, Haiti, Burundi, Lesotho, Nigeria, Indonesia, Rwanda, Nepal, China, Mali, Tanzania, Burkina Faso, Malawi, Ethiopia, Uganda.