Journal of International Economics 65 (2005) 1–20
www.elsevier.com/locate/econbase

Journal of INTERNATIONAL ECONOMICS

ELSEVIER

A simple model of firm heterogeneity, international trade, and wages

Stephen Ross Yeaple*

Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA

Received 25 June 2002; received in revised form 23 November 2003; accepted 26 January 2004

Abstract

This paper presents a general equilibrium trade model in which homogeneous firms choose a technology from a set of competing technologies and choose employees from a set of workers of heterogeneous skill. In equilibrium, the interaction between the characteristics of competing technologies, international trade costs, and the availability of workers of heterogeneous skill gives rise to firm heterogeneity. The model generates several of the stylized facts concerning the (apparent) superiority of firms that engage in international trade relative to those that do not and has implications for the effect of international trade on the skill premium and on observed industry-level productivity.

Keywords: Intra-industry trade; Technology adoption; Firm heterogeneity; Trade and wages

JEL classification: F0; F1

It has long been known that firms producing apparently similar products display considerable heterogeneity in terms of their size, productivity, and the wages they pay of their employees. Relatively recent empirical research has also uncovered a systematic link between the characteristics of firms and their propensity to engage in international trade. Using highly disaggregated firm and plant level data, a number of studies have shown that, even within narrowly defined industries, not all firms participate in export markets and those that do tend to be larger, tend to use more advanced technologies, and tend to pay higher wages than those that do

* Tel.: +1-215-898-1875; fax: +1-215-573-2057.
E-mail address: Snyeapl2@ssc.upenn.edu (S.R. Yeaple).

0022-1996/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jinteco.2004.01.001

not. Moreover, exporters are often found to be more productive than non-exporters.[1]

This paper contributes to the theoretical literature that seeks to explain the systematic differences between exporters and non-exporters and to understand the economic implications of international trade in the presence of such heterogeneity. In much of this literature, firm heterogeneity is generated by assigning productivity levels to firms randomly. Conditional on their productivity draw, firms then sort endogenously into exporter/non-exporter status.[2] My point of departure is different: Firms are identical when born, are free to produce with technologies that differ in their characteristics, and are free to hire workers who vary in their skill on a perfectly competitive labor market. Firm heterogeneity arises because firms endogenously choose to employ different technologies and then systematically hire different types of workers.

To fix ideas, suppose that a recent advance has led to the creation of a technology that allows production at a lower unit cost relative to an older technology once an additional fixed cost has been incurred. Now suppose that workers vary in their skill and that, given two workers, the more skilled worker has an absolute advantage in both technologies and a comparative advantage in the newer, low unit cost technology.[3] In some equilibria, firms that use the old technology and firms that use the new co-exist, and the number of each type is determined in large part by scarcity of skill in the labor market.

Now suppose that international trade is costly, incurring both a fixed cost and an iceberg transport cost per unit sold abroad. In the presence of such a fixed cost, only firms that use the low unit cost technology, and hence are able to sell a large quantity profitably, enter the export market, so that firms that export are larger, use more advanced technology, and pay higher wages than those that do not. They also exhibit greater sales per worker, but this apparent productivity advantage disappears when worker heterogeneity is properly controlled. A reduction in transport costs increases the incentive for firms to adopt the new, lower unit cost technology, causing workers to be reallocated from the old technology to the new. In addition, total employment in the industry falls as the least skilled workers leave the industry for employment elsewhere. Hence, a reduction in trade friction between even identical countries can raise the relative demand for skilled workers and the average skill level of workers across all industries as firms are induced to adopt technologies favoring the highly skilled.[4]

While the model presented in this paper is novel in that firm heterogeneity is generated by the interaction between trade costs, the characteristics of competing technologies, and the existence of worker skill heterogeneity, there are at least two papers that contain similar components. The first is Ekholm and Midelfart Knarvik (2001), who consider the effect of trade and technology choice in a reciprocal dumping setting with two homogeneous

[1] See for instance, Bernard and Jensen (1997, 1999). That this phenomenon appears in developing countries buttresses the argument that exporters are different from non-exporters. See Clerides et al. (1998).

[2] For recent models fitting this description, see Meltiz (2003) and Bernard et al. (2003a,b).

[3] See Bartel and Lichtenberg (1987) and Bartel and Sicherman (1999) for evidence supporting this assumption.

[4] This is a conjecture that has appeared in the literature. See, for instance, Wood (1994).

factors. Their paper differs from mine in that they do not consider factor heterogeneity and rely heavily on simulations to generate their results. The other closely related paper is Manasse and Turrini (2001), who consider a model in which entrepreneurs vary in their ability. Each entrepreneur produces a single variety of a differentiated good using a homogenous input (raw labor), and the quality of this variety increases in entrepreneurial ability. As in my paper, firms compete via monopolistic competition, incur a fixed cost if they export and a transport cost on each unit sold abroad. The key decision facing these heterogeneous entrepreneurs is whether or not to engage in international trade. In equilibrium, only the firms of the "superstar" entrepreneurs export. Hence, like mine, their model attributes the differences between exporters and non-exporters to differences in inputs (entrepreneurial skill). As transport cost fall between countries marginal entrepreneurs expand into export markets.

The key difference between my paper and that of Manasse and Turrini is that I consider the effect of trade costs on four (rather than one) firm-level decisions: (1) entry, (2) technology choice, (3) whether or not to export, and (4) the types of workers to employ. I show that the interaction of trade costs with the characteristics of competing technologies and with the scarcity of skilled workers in the labor market can explain the same stylized facts as that of Manasse and Turrini; i.e., exporters are larger and pay non-production workers higher wages than non-exporters. Further, falling transport costs affects all four decisions, changing the demand conditions in the market for heterogeneous workers and leading to the reallocation of workers across technologies within an industry and across industries as well. Our papers therefore offer very different and competing explanations for the differences between exporters and non-exporters as well as for the causes of the growing skill premium. The model presented here also makes two additional predictions that are consistent with the results of Bernard and Jensen (1999): (1) that exporters also pay a premia to production workers, and (2) that exporters employ more advanced technology than non-exporters. Finally, this paper also explores the effect of trade on observed productivity, a topic of much recent interest in both the theoretical and empirical literature. Indeed, I will demonstrate that many of the predictions of models of firm heterogeneity based on random technology shocks (i.e., Meltiz, 2003) can also be explained by a model based on inherently homogeneous firms.

The rest of the paper is organized into four sections. In the first, I present the model. In the second, I characterize the model's equilibrium in a closed economy setting. In the third, I characterized the open economy equilibrium, analyze the effect of a reduction in international trading costs on the allocation of factors across firms and industries, and derive the implications of expanding world trade for the wage distribution. The final section concludes.

1. Setup of the model

I begin by considering a closed economy version of the model to highlight the mechanisms through which firm heterogeneity emerges. The model will only require slight adjustments to consider trade between two identical economies.

1.1. Demand

The preferences of a representative consumer are Cobb–Douglas over a homogenous good, Y, and a composite differentiated good, X, and CES over a continuum of varieties of X. Specifically, they are

$$U = (1-\beta)\ln Y + \beta \ln X \text{ where}$$

$$X = \left[\int_0^N x(i)^{\alpha} \mathrm{d}i\right]^{\frac{1}{\alpha}} \text{ and } \sigma = \frac{1}{1-\alpha} > 1.$$

The elasticity of substitution across varieties of X is given by σ. As shown by Dixit and Stiglitz (1977), consumer behavior can be modeled by considering the set of varieties consumed as an aggregate good, X, with aggregate price:

$$P_X = \left[\int_0^N p(i)^{1-\sigma} \mathrm{d}i\right]^{\frac{1}{1-\sigma}}. \tag{1}$$

Total demand for any variety can then be written

$$x(i) = \left(\frac{\beta E}{P_X}\right)\left(\frac{p(i)}{P_X}\right)^{-\sigma}, \tag{2}$$

where E is aggregate expenditure and $\beta E/P_X$ is total spending on the composite good X.

1.2. Workers

In each country, there is a continuum of workers with mass M. Workers are differentiated by their skill level, which I index by Z. A larger value of Z corresponds to a more skilled worker. There are two interpretations of Z. The first is that workers differ in some observable characteristic, such as educational attainment. The second interpretation of Z is as a measure of worker quality or ability that can be observed by a firm but not by an econometrician. The distribution of skills in the population is given by $G(Z)$ with density $g(Z)$ and support $[0,\infty)$.[5]

1.3. Production

Each good is produced using only labor. Entry into either sector is free with each entrant into the X sector becoming the sole producer of a distinct variety. There is a single technology for producing Y, but there are two technologies for producing varieties of X.

[5] It should be understood that this interval can be quite large but is finite at some upper bound.

The amount of a good a worker of skill Z can produce is given by $\varphi_j(Z)$ where j is an index to indicate which of the three technologies a worker of skill Z is using. Let $j \in \{Y, H, L\}$, where $j = Y$ refers to the technology for producing Y, $j = H$ refers to the employment of technology H (the new or "high-tech" technology) for producing X, and $j = L$ refers to the employment of technology L (the old or "Low-tech" technology) for producing X.

A skilled worker is more productive than an unskilled worker, so $\varphi_j(Z)$ is continuous and increasing in Z for all activities. Further, I assume

$$\varphi_H(0) = \varphi_L(0) = \varphi_Y(0) = 1 \text{ and}$$
$$\frac{\partial \varphi_H(Z)}{\partial Z}\frac{1}{\varphi_H(Z)} > \frac{\partial \varphi_L(Z)}{\partial Z}\frac{1}{\varphi_L(Z)} > \frac{\partial \varphi_Y(Z)}{\partial Z}\frac{1}{\partial_Y(Z)} > 0. \tag{3}$$

These assumptions are consistent with worker comparative advantage based on skill. Highly skilled workers have a comparative advantage in high technology production of X relative to moderate and low skilled workers and moderately skilled workers have a comparative advantage producing X relative to low skilled workers. Also note that for all but the least skilled worker, the new technology (subscripted H) is lower unit cost than the old technology (subscripted L).

Firms are free to enter in both sectors, but to produce a variety of X, a firm must first bear a fixed cost. The size of this fixed cost depends on the technology employed: to produce with technology $j \in \{L, H\}$, a firm must incur fixed cost F_j. I assume that $F_H > F_L$. In the interest of analytic tractability, I assume that fixed costs take the form of output that must be produced in order to enter, but which ultimately cannot be sold.[6]

2. Closed economy equilibrium

I begin the derivation of the closed economy equilibrium by deriving the optimal allocation of workers to technologies. By first considering the closed economy equilibrium, I illustrate as transparently as possible the model's structure and shed light on the relationship between firm size and wages.

Our analysis at this point relies only on the fact that profit-maximizing firms set marginal revenue equal to the marginal cost of production. As is well known, for the monopolistically competitive firms producing X, the revenue earned by the producer of variety k is

$$R_k = (\beta E P_X^{\sigma-1}) p_k^{1-\sigma}$$

and the marginal revenue for a firm k is equal to αp_k. It follows immediately that all firms in the X sector charge a price equal to a constant mark-up over unit cost. In the perfectly competitive Y sector, firms charge a price equal to unit cost. Note that for firms producing

[6] This specification of fixed costs is chosen for the purpose of tractability. The model delivers very similar results under alternative assumptions, but the labor market clearing conditions become considerably more complicated.

with technology j, the unit cost of producing with a worker of skill Z is $W(Z)/\varphi_j(Z)$. In a perfectly competitive labor market, the wage distribution over Z adjusts to equalize the unit costs of all firms using the same technology. Firms minimize their costs given their technology and the equilibrium wage distribution. I define C_Y, C_L, and C_H as the unit cost of firms producing with each of the three technologies. The following lemma can easily be proven:

Lemma 1. *If a worker with skill $\bar{Z}$ works in the Y sector, then all workers with skill $Z<\bar{Z}$ will only be hired in the Y sector. If a worker with skill $\hat{Z}$ works for a firm using the H technology, then all workers with skill $Z>\hat{Z}$ will be hired by firms using the H technology.*

There must always be some workers who produce Y, and by Lemma 1, these workers are the least skilled. I define the most skilled worker using the Y technology as Z_1. If the distribution of skill is unbounded, then there must be some workers with the highest skill levels hired by H firms. I define the least skilled workers using the H technology as Z_2. If there are firms that use technology L, then these firms must hire workers with intermediate skills or $Z\in(Z_1, Z_2)$. It remains to be seen whether firms using technology L actually appear in equilibrium. For the time being, I simply assume that they do. Given this allocation of workers to technologies, it follows that the wage distribution in a competitive labor market is

$$W(Z) = \begin{cases} C_Y\varphi_Y(Z) & 0\leq Z\leq Z_1 \\ C_L\varphi_L(Z) & Z_1\leq Z\leq Z_2 \\ C_H\varphi_H(Z) & Z\geq Z_2 \end{cases}. \tag{4}$$

Each worker employed with a given technology must be paid a "technology" specific efficiency wage. Given that $\mathrm{MR}_Y=p_Y=1$ and that workers with skills Z_1 and Z_2 are indifferent between working in (respectively) Y or L and L or H, the unit costs are then immediately given by

$$\begin{aligned} C_Y &= 1 \\ C_L &= \frac{\varphi_Y(Z_1)}{\varphi_L(Z_1)} < 1 \\ C_H &= \frac{\varphi_Y(Z_1)}{\varphi_L(Z_1)}\frac{\varphi_L(Z_2)}{\varphi_H(Z_2)} < C_L. \end{aligned} \tag{5}$$

Expressions (4) and (5) completely characterize the wage distribution that must obtain if firms using all three technologies exist in equilibrium.

To illustrate the relationship between the two expressions, I present the resulting mapping of skill to log wages ($\log W(Z)$) in Fig. 1.[7] Fig. 1 shows each of the three expressions in Eq. (4). The outer hull (shown in bold) represents the market log wage paid by firms using each of the three different technologies. As the figure

[7] The linear representation is one of many relationships consistent with the assumptions in Eq. (3).

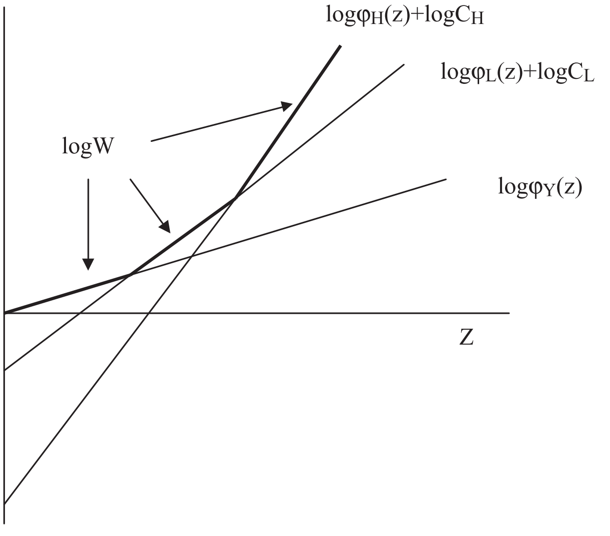

Fig. 1. The wage distribution.

shows, the wage at which a worker who is assigned to the "wrong" technology is strictly less than the wage that could be obtained if the worker were assigned to the "right" technology. Note also that a firm with a given technology is indifferent by construction to a high or low Z worker as long as both workers are in the appropriate interval.

The gradient of the wage distribution is increasing at the thresholds, Z_1 and Z_2 (the intersections), because the value of an additional unit of workers' skill is greater for firms using the technologies that are progressively more sensitive to skill. If one interprets Z as unobserved quality heterogeneity among workers, X as traded manufacturing goods, and Y as non-traded services (the natural interpretation of these two industries in the open economy version), then this wage distribution is consistent with the stylized fact that manufacturing pays a premium to workers relative to most service industries (see Katz and Summers, 1989). Finally, changes in C_L and C_H, brought about by changes in Z_1 and Z_2, have the effect of altering the relative wages across workers of different skill as can be seen in Eq. (5). An increase in Z_1 lowers the relative wage of the moderately skilled workers and an increase in either Z_1 and/or Z_2 lowers the relative wage of the most highly skilled workers. The following proposition is immediate:

Proposition 1. *If firms that employ the L technology and firms that employ the H technology appear in equilibrium, then relative to a firm that employs technology L, a firm that employs technology H pays higher wages and has greater revenues.*

The first part of Proposition 1 follows directly from our derivation of the wage-skill schedule in Eq. (4). The skill of any worker employed with technology H is greater than the skill of any worker employed with technology L and so must be paid a greater wage.

The second part follows directly from Eq. (5) and the fact that

$$\frac{R_H}{R_L} = \left(\frac{C_H}{C_L}\right)^{1-\sigma} > 1,$$

where R_H and R_L are the revenues of firms employing the H and L technologies respectively. The predictions of Proposition 1 are consistent with the empirical finding of Davis and Haltiwanger (1991) that within industries, large firms pay a higher wage than small firms.

Now consider other differences that might arise between firms that employ the H and firms that employ the L technology. Let R_j be the revenue of a firm using technology j. Free entry ensures that profits for both firm types are zero so that

$$R_j = C_j(x_j + F_j). \tag{6}$$

Eq. (6) states that the revenue earned by a firm using technology j must be equal to its total labor cost.[8] Hence, the sum of the revenue of all firms using technology j must be equal to the total wages paid to workers employed using technology j. Average revenue per firm for firms using technology j is then simply the average wage paid to workers using this technology. Finally, note that every worker that uses technology H is paid a higher wage than a worker using technology L. The following proposition is then immediate:

Proposition 2. *If in equilibrium there are firms that employ the L technology and other firms that produce using the H technology, average revenue per worker at the H technology firms will be greater than average revenue per worker at L technology firms.*

Propositions 1 and 2 together imply that large firms should pay higher wages and have greater observed worker productivity than small firms. These theoretical results are consistent with a conjecture made in the Labor literature that large firms pay higher wages because they employ technologies that are best implemented by relatively high skilled workers.[9] The focus of Proposition 2 on revenue per worker is natural given that a large number of studies (see, for instance, Idson and Oi, 1999; Bernard and Jensen, 1999; Bernard et al., 2003b) use value added per worker as their measure of productivity.[10] The difference in productivity between H and L technology firms needs to be interpreted with care. The model is built on the strong assumptions of perfectly competitive labor markets, homogenous firms, and free entry in the product market that revenues are always equal to total costs as in Eq. (6). As a result, productivity measures that perfectly accounted for differences in

[8] Recall that the fixed cost is in terms of a quantity of output that must be produced but cannot be sold. Also, the assumption that both firms exist implies that entrants make nonnegative profits.

[9] See, for instance, Idson and Oi (1999) who present evidence on the firm size–wage relationship highly consistent with the implications of Propositions 1 and 2. They conclude (p. 107), "that firms that achieve large size create jobs (technologies, equipment, and work organizations) that must be matched with more productive individuals."

[10] The difference in average output in quantity (generally unobserved given the lack of firm specific price information) between H and L technology firms would be even greater because H firms charge lower prices.

factor qualities would reveal no difference between large and small firms. This result need not hold under alternative market structures, which are outside the scope of this paper.

I now turn to the two equilibrium conditions that pin down Z_1 and Z_2. The first is derived from the zero profit conditions for firms producing X and the second is the market clearing condition for Y. All other endogenous variables of interest are functions of these thresholds.

Beginning with the zero profit condition, the revenues of H firms (and L firms if they exist) must exactly equal their costs. As is standard in the monopolistic competition setting with CES preferences, the revenue of a firm using technology j less its variable costs is a fixed multiple of its revenue or R_j/σ, which, in turn, given free entry, must be less than or equal to its fixed cost, or C_jF_j. If both H and L technology firms make zero profits then

$$\frac{R_H}{R_L} = \left(\frac{C_H}{C_L}\right)^{1-\sigma} = \frac{C_H F_H}{C_L F_L}.$$

Reorganizing this expression and substituting using the definition of C_H and C_L yields the following equilibrium condition for Z_2:

$$\frac{C_H}{C_L} = \frac{\varphi_L(Z_2)}{\varphi_H(Z_2)} = \left(\frac{F_H}{F_L}\right)^{-\frac{1}{\sigma}}. \tag{7}$$

The threshold Z_2 is pinned down by the zero profit conditions alone and so is invariant to country characteristics such as the size of the labor force or the distribution of skill in the population.

Since the ratio of labor productivities, $\varphi_L(Z)/\varphi_H(Z)$, is decreasing in Z, it follows from Eq. (7) that an increase in the relative fixed cost to using the H technology, F_H, would raise Z_2, which is consistent with a shift of labor out of firms using the H technology and a reduction in the premium paid to highly skilled labor (see Fig. 1). Given a higher fixed cost to adopting the H technology, the cost of employing highly skilled labor must fall to compensate H technology firms with greater revenues net of variable costs. An increase in σ, the elasticity of substitution across varieties, reduces Z_2 and therefore increases the share of the labor force employed at H firms. Greater product market competition increases the importance of having low unit cost.

Now consider market clearing in the Y sector. Since Y is the numeraire and first tier preferences are Cobb–Douglas, total expenditures on Y must be

$$Y = (1-\beta)E = (1-\beta)M\bar{W}, \tag{8}$$

where E is expenditure, M is the mass of workers, and $\bar{W}$ is the average wage per worker. From Eqs. (4) and (5), the average wage per worker is

$$\bar{W} = \int_0^{Z_1} \varphi_Y(Z)\mathrm{d}G(Z) + C_L \int_{Z_1}^{Z_2} \varphi_L(Z)\mathrm{d}G(Z) + C_H \int_{Z_2}^{\infty} \varphi_H(Z)\mathrm{d}G(Z) \tag{9}$$

The three integrals correspond to the earnings of labor employed using the Y, L, and H technologies, respectively. It will be convenient to define the following functions that make explicit the connection between the thresholds Z_1 and Z_2 to the unit costs C_L and C_H,

$$S(Z_1) \equiv \frac{\varphi_Y(Z_1)}{\varphi_L(Z_1)} = C_L, \text{ and } A(Z_2) \equiv \frac{\varphi_L(Z_2)}{\varphi_H(Z_2)} = \frac{C_H}{C_L}.$$

Note that $S(Z_1)$ and $A(Z_2)$ are strictly decreasing. Expenditure on Y is then a function of Z_1 and Z_2. Everything else equal, a decrease in either threshold raises average wages in terms of the numeraire and therefore increases demand for Y.

The demand for Y given by Eq. (8) must be equal to the total wages paid to workers employed in the Y sector, which is $M\int_0^{Z_1} \varphi_Y(Z)\mathrm{d}G(Z)$. Market clearing in Y then yields the second equilibrium condition,

$$\frac{\beta}{(1-\beta)} \frac{1}{S(Z_1)} \int_0^{Z_1} \varphi_Y(Z)\mathrm{d}G(Z) = \int_{Z_1}^{Z_2} \varphi_L(Z)\mathrm{d}G(Z) + A(Z_2) \int_{Z_2}^{\infty} \varphi_H(Z)\mathrm{d}G(Z). \quad (10)$$

Together, Eqs. (7) and (10) define the cutoffs Z_1 and Z_2. The left-hand side of Eq. (10) is strictly increasing in Z_1, while the right-hand side is strictly decreasing in Z_1 given Z_2 fixed by Eq. (7).

There are three features of this system worth mentioning. First, the size of the economy in terms of M plays no role in the allocation of workers across technologies and so has no effect on the wage schedule. Second, changes in the distribution of worker skill, given by the density function $g(Z)$, will have an impact on Z_1 given a fixed Z_2 and so does have an impact on the mapping of skill to wages, $W(Z)$. Finally, by totally differentiating Eq. (10), it can be established that Z_1 and Z_2 must move in opposite directions, i.e., $\mathrm{d}Z_1\mathrm{d}Z_2 < 0$. Intuitively, a decrease in Z_2 has the effect of increasing national expenditure, which drives up the demand for good Y. For the market in Y to clear, employment in the Y sector must rise so Z_1 must rise. Parameter changes that increase Z_2 via Eq. (7) decrease Z_1 through Eq. (10).

I have assumed in the discussion so far that L technology firms arise in equilibrium. The equilibrium conditions just derived suggest that L technology firms are more likely to exist if the fixed cost of using the H technology F_H is large. Put another way, if there is little fixed cost saving for using the L technology, it may not be used. Firms using the L technology are also more likely to be observed if the share of total expenditure on the composite X good β is large. To see this, note that by Eq. (7), $Z_2>0$. If $\beta = 1$, then there is no spending on good Y, and then $Z_1 = 0$ and L firms must exist. As β falls, good Y expenditure rises and Z_1 must rise for fixed Z_2, making L technology firms more likely to disappear.[11] For the remainder of the paper, I assume that F_H and β are sufficiently large to allow L firms to arise in equilibrium.

To complete the discussion of the closed economy, I derive the number of firms that adopt the H technology and the L technology as a function of Z_1 and Z_2. First, consider the number of firms in the X sector that adopt the H technology, N_H. To solve for N_H, notice that the amount of X produced by H firms is $M\int_{Z_2}^{\infty} \varphi_H(Z)\mathrm{d}G(Z)$. Since each H technology firm faces identical costs and uses the same technology, they each employ an identical

[11] This can be confirmed by totally differentiating Eq. (10), holding Z_2 fixed.

quantity of effective units of labor. The output of X goes to two ends: to the product market and to satisfy the fixed cost. Hence, total effective labor required by the representative H technology firm is x_H+F_H, where x_H represents the quantity sold of the representative variety. Since free entry requires that $R_H=\sigma C_H F_H$, it follows that $x_H=(\sigma-1)F_H$ so that the number of firms using the H technology is

$$N_H=\frac{M}{\sigma F_H}\int_{Z_2}^{\infty}\varphi_H(z)\mathrm{d}G(Z). \tag{11}$$

By the same procedure, we find that the number of firms using the L technology is

$$N_L=\frac{M}{\sigma F_L}\int_{Z_1}^{Z_2}\varphi_L(z)\mathrm{d}G(Z). \tag{12}$$

Finally, since each firm charges a markup of $1/\alpha$ over its unit cost, it can be shown using Eqs. (7), (10), (11), and (12) that the price index for the composite X good can be written as

$$P_X=\frac{1}{\alpha}\left[\frac{\beta}{1-\beta}\frac{M}{\sigma F_L}\right]^{\frac{1}{1-\sigma}}S(Z_1)^{\frac{\sigma}{\sigma-1}}\left(\int_0^{Z_1}\varphi_Y(Z)\mathrm{d}G(Z)\right)^{\frac{1}{1-\sigma}}. \tag{13}$$

3. The open economy

To keep the analysis tractable, I assume that the world is composed of two identical economies.[12] Trade in such an environment is entirely intra-industry in varieties of X. International trade is costly, forcing firms that engage in international trade to incur both variable and fixed costs. Variable costs take the form of iceberg transport costs so that for one unit of a good to arrive, $\tau>1$ units must be shipped. The fixed cost of participating in the export market is F_X, which, like the other fixed costs, takes the form of output that must be produced but which cannot be sold. The fixed cost to exporting is consistent with the need to obtain information about the foreign market, to alter a product's characteristics to suit a foreign market, or to create a distribution network in the foreign country.[13] Because countries are identical, I need only consider the equilibrium allocations and prices in one.

3.1. Open economy equilibrium

The extension to the open economy case requires only minor modification. The allocation of workers to technologies and the resulting relationship between Z and wages remains the same (although Z_1 and Z_2 are free to vary). Since trade is between identical

[12] By considering identical countries, I need only consider two thresholds, Z_1 and Z_2, which are common to each country. Country differences give rise to thresholds and equilibrium conditions that vary across countries.

[13] Roberts and Tybout (1997) highlight the importance of sunk costs of exporting. Since my framework is static, I consider fixed rather than sunk cost.

countries, Y is not traded so the market clearing for good Y continues to be given by Eq. (10). Only the zero profit conditions in the differentiated goods sector need adjustment. Firms must now decide whether or not to engage in international trade in addition to serving their home market. Relative to sales in the domestic market, revenues in the foreign market are reduced by the proportion $\tau^{1-\sigma}$, reflecting the higher marginal cost induced by transport cost. Firms that employ technology $j \in \{L, H\}$ realize revenues of R_j from the local market and $R_j(1+\tau^{1-\sigma})$ if they serve both the local and foreign market. Firms incur fixed costs of C_jF_j if they serve the local market and $C_j(F_j+F_X)$ if they serve both the local market and export. Assuming that in equilibrium that there are firms that choose technology L and other firms that choose technology H, then it follows that if $F_H < F_X\tau^{\sigma-1}$, then no firms export, while if $F_L > F_X\tau^{\sigma-1}$, then all firms export. The case of greatest interest is where

$$F_H > F_X\tau^{\sigma-1} > F_L. \tag{14}$$

When Eq. (14) holds, the gain in additional sales from exporting is sufficiently large that no H technology firm can enter and serve only the local market and make nonnegative profits while no L technology firm can enter and make nonnegative profits by serving both the local and foreign markets. The following proposition has been established.

Proposition 3. *If $F_H > F_X\tau^{\sigma-1} > F_L$, then H technology firms export and L technology firms do not.*

I assume for the remainder of the paper that Eq. (14) holds. Under this condition, Propositions 1–3 establish that firms that export are different from firms that do not. First, because exporting firms are those that have chosen a low unit cost, high fixed cost technology, their sales are greater in the domestic market than non-exporters. This result is consistent with the stylized fact that exporters tend to be larger than their non-exporting competitors. Second, because more highly skilled workers have a comparative advantage in the H technology, the firms that export pay higher wages than those that do not. This is consistent with the stylized fact that exporters pay higher wages than non-exporters even within narrowly defined industrial categories and across countries of substantial levels of development.

Finally, the value of output produced per worker is greater for exporters than for non-exporters for the simple reason that each worker employed by an exporter is more skilled and hence more productive than any worker employed at a non-exporter. As in the closed economy case, however, revenue must equal total factor cost by the free entry condition so that productivity comparisons calculated in such a way as to adjust perfectly for variation in labor quality would not reveal any difference between exporters and non-exporters. It is interesting to note that Bernard and Jensen (1999) report wage premiums paid to labor at exporting firms that are of a similar magnitude (although somewhat smaller) to exporter premiums in value added per worker as would be the case if exporters hire higher quality workers.

I now adapt Eq. (7), the free entry condition to the open economy case. Since revenues less variable costs must equal fixed costs for both H and L technology firms,

$$\frac{R_H(1+\tau^{1-\sigma})}{R_L} = \left(\frac{C_H}{C_L}\right)^{1-\sigma}(1+\tau^{1-\sigma}) = \frac{C_H(F_H+F_X)}{C_L F_L}.$$

Rearranging this expression yields

$$\frac{C_H}{C_L} = \frac{\varphi_L(Z_2)}{\varphi_H(Z_2)} \equiv A(Z_2) = \left(\frac{F_H+F_X}{F_L(1+\tau^{1-\sigma})}\right)^{-\frac{1}{\sigma}}. \tag{15}$$

In the open economy case with two identical countries, Eqs. (10) and (15) define Z_1 and Z_2. Given these thresholds, all other variables of interest, such as the exact shape of the wage schedule, the number of entrants using each technology, and the relative price of the composite X sector good are each pinned down. As in the case of the closed economy, the existence of L technology firms is not guaranteed but becomes more likely the larger the share of income spent on the composite X good, β, and the higher the fixed cost to adopting the H technology, F_H. In going from the closed to the open economy, it follows directly from Eq. (15) that Z_2 must fall. A drop in Z_2 must increase Z_1 through Eq. (10), making it possible that L technology firms might exist in the closed economy but not in the open economy. I assume for the remainder of the paper that F_H is sufficiently large to ensure that L technology firms exist.

Of the remaining variables of interest, only the number of firms using the H technology need be adjusted. The only change to Eq. (11) that must be made is that the new level of fixed costs being incurred by H technology firms is now F_H+F_X instead of just F_H. Using the same reasoning as above, it can be shown that the number of H technology firms must be given by[14]

$$N_H = \frac{M}{\sigma(F_H+F_X)}\int_{Z_2}^{\infty}\varphi_H(Z)\mathrm{d}G(Z). \tag{11$'$}$$

The number of L technology firms continues to be given by Eq. (12) and the price index for the composite X good, P_X, continues to be given by Eq. (13).

3.2. The effect of falling trade costs

A reduction in the costs of international trade can occur either because the marginal cost to exporting has fallen (τ) or because the fixed costs to exporting (F_X) have fallen. Of the two costs, the fixed cost to exporting is of central importance in generating an outcome in which some firms export while others do not. In this respect, the model is similar to Manasse and Turrini (2001) and Meltiz (2003). But, for most of the variables of interest, a small change in the fixed costs to exporting, as long as it does not result in the violation of Eq. (14), has economic implications that are similar to the effects of changes in τ. Given

[14] Note that labor must be used to ship goods internationally given the iceberg transport cost assumption.

this similarity, and the fact that τ is easier to quantify, I focus on changes in the marginal cost of exporting, τ.

The following proposition follows directly from Eqs. (10) and (15):

Proposition 4. *A small decrease in τ increases the share of the labor force working with technology H and technology Y and reduces the share of labor employed making good X. ($dZ_1 d\tau < 0$ and $dZ_2 d\tau > 0$).*

Intuitively, a reduction in trade costs is much like a technological improvement for H technology firms. For the zero profit condition (Eq. (15)) to continue to hold, the production cost of H technology firms must rise relative to L technology firms, which can only be accomplished by a decrease in Z_2. Note that from Eq. (11′), it follows immediately that a reduction in transport cost also increases the number of firms that adopt the H technology and enter the export market. A decrease in Z_2 raises the wage of the most highly skilled workers in terms of the numeraire, Y, leaving the wage of all other workers unchanged. As a result, total expenditure, E, must rise, raising the demand for Y and inducing an increase in the demand for labor from the Y sector. The implication of Proposition 4 is that an increase in international trade reduces employment in manufacturing, a prediction consistent with global employment trends.

Another implication of Proposition 4 is that falling transport costs induce a reallocation of workers from L technology jobs to H technology jobs within industries. Further, it follows immediately from Proposition 4 and Eq. (12) that the number of firms using the L technology must decrease as transport costs fall. It can also be shown, that as in Meltiz (2003), a reduction in transport costs reduces the total number of varieties produced in each country as the increase in N_H is less than the decrease in N_L. Hence, trade induces increased sales concentration.

I now consider changes in average revenue per worker, or observed labor productivity, at the industry level. Aggregate revenue in the Y sector is $(1-\beta)E = Y$. Hence, aggregate revenue per worker is given by

$$\frac{1}{G(Z_1)} \int_0^{Z_1} \varphi_Y(Z) \mathrm{d}G(Z),$$

which is increasing in Z_1. Proposition 5 follows immediately.

Proposition 5. *A reduction in τ raises revenue per worker in the Y sector.*

A reduction in trade costs pushes the least productive workers in the X sector into the Y sector, but these workers are of above average productivity in the Y sector, increasing the value of output per worker. Since firms earn zero profit, however, all revenue is paid to workers so that measures of productivity that account for worker heterogeneity would reveal no change.

Now consider revenue per worker in the X sector. Since all firms make zero profits, total revenue of all X firms must be equal to the total labor income of workers in the X sector or

$$M\left[C_L \int_{Z_1}^{Z_2} \varphi_L(Z) \mathrm{d}G(Z) + C_H \int_{Z_2}^{\infty} \varphi_H(Z) \mathrm{d}G(Z)\right].$$

Using Eq. (10) and the fact that employment in X is $M[1-G(Z_1)]$, revenue per worker can be shown to be

$$\frac{\beta}{1-\beta}\frac{1}{1-G(Z_1)}\int_0^{Z_1}\varphi_Y(Z)\mathrm{d}G(Z),$$

which is strictly increasing in Z_1. Since a reduction in τ increases Z_1, the next proposition follows immediately.

Proposition 6. *A reduction in τ raises revenue per worker in the X sector.*

This implication of the model is consistent with recent empirical evidence as presented in Bernard et al. (2003b), who show that in those industries in which trade is primarily intra-industry, a reduction in transport costs is associated with greater average sales revenue per labor input. Note also that average quantity of output per worker also rises as the average skill in the industry increases. As before, the caveat that revenues are equal to costs applies so that productivity calculations that control for input heterogeneity reveal no changes in productivity.

The next lemma (proof is in Appendix A) establishes the impact of reductions in the cost of international trade on the unit cost of production of L and H technology firms, or C_L and C_H, respectively.

Lemma 2. *A reduction in τ reduces C_L, the unit cost of production using the L technology and increases C_H, the unit cost of production using the H technology.*

Lemma 2 has many implications. First, using Lemma 2 and the zero profit condition for the representative L technology firm, $R_L=\sigma C_L F_L$, it follows immediately that the revenues of L technology firm must fall with a reduction in τ. Second, using Lemma 2 and the zero profit condition for the representative H firm, $R_H(1+\tau^{1-\sigma})=\sigma C_H(F_H+F_X)$, it follows that a reduction in τ must increase total revenues of an H firm. Third, it can be shown that domestic sales of an H firm, R_H, must fall with a decrease in trade costs, while export sales, $R_H\tau^{1-\sigma}$, must rise. Since the number of exporters has increased (N_H) and the export revenue of each exporter has increased, it follows immediately that the volume of trade between the two identical countries has also increased. These predictions of the model are similar to those of Meltiz (2003).

Lemma 2 can also be used to analyze the effect of shipping costs on the distribution of income. Propositions 5 and 6 establish that revenue per worker rises in both sectors. Since firms earn zero profits, these increases in labor productivity are ultimately passed onto workers so that it immediately follows that average wages (measured in terms of the numeraire) have increased in both industries. I now turn to income distribution issues that are masked by movement in aggregate industry wages.

The distribution of wages is given by Eq. (4), which shows that the function mapping skill Z into wages is a function of costs, C_L and C_H, which, in turn, are functions of the thresholds of Z_1 and Z_2. Since these thresholds are affected by the reduction in τ, there is a direct effect on the income distribution across Z. This effect is summarized in the following proposition:

Proposition 7. *A reduction in τ increases the wage of the most highly skilled members of society, $Z>Z_2$, and does not affect the wage of the least skilled workers, $Z<Z_1$. Moderately skilled workers, those that remain employed in the X sector producing with the L technology or those that have become employed in the Y sector, must see their wage fall.*

The original wage gradient is shown in Fig. 2 as the solid line. A reduction in transport cost reduces Z_2 to Z_2' and increases the Z_1 to Z_1'. This has the effect of changing the wage function from the solid to the broken line. Workers who were initially using the Y technology continue to use that technology after the reduction in transport costs and hence do not see their wage change relative to the least skilled member of society.

It is the moderately skilled people who see their status in society eroded. Workers who are thrown out of the X sector are less productive in the Y sector and hence earn less relative to the least skilled worker than they had before the increase in trade. The increase in Z_1 reduces the wages of those who remain in employed with non-exporting X sector firms that use the L technology because the marginal worker is paid what she is worth in sector Y.

While Proposition 7 establishes that the economic status of moderately skilled workers has eroded relative to both high and low skilled workers, it is not clear whether the real income of these workers has fallen. A first step in determining the implications for worker real income is to note that a reduction in trade cost must reduce the price of the composite X good, P_X, via Eq. (13). Since the wages of workers originally employed in Y have not changed relative to the numeraire and since the wages of workers originally employed in X with the H technology have increased relative to the numeraire, the real income of both

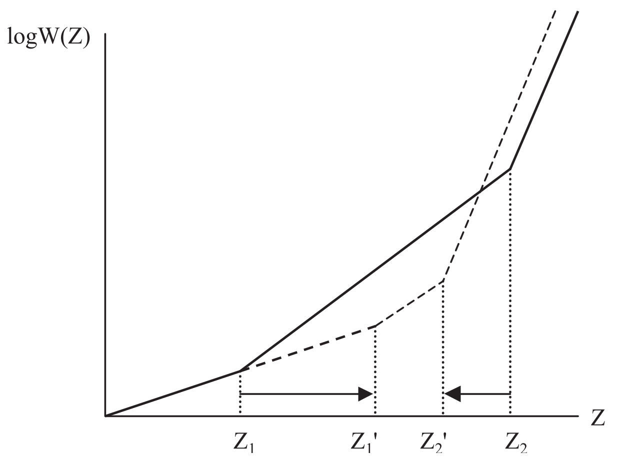

Fig. 2. Falling transport cost and the wage distribution.

sets of workers must rise. The wage of moderately skilled workers (those that have been forced out of the X sector or remain employed with the L technology) falls relative to the numeraire. Their wage relative to the composite X good has risen, however. To see this, note that the wage of one of these workers is $C_L\varphi_L(Z)$ so the wage of one of these workers in terms of the composite X good is

$$\frac{C_L\varphi_L(Z)}{P_X} = \alpha\left(\frac{\beta}{1-\beta}\frac{M}{\sigma F_L}\right)^{\frac{1}{\sigma-1}}\left(\frac{1}{S(Z_1)}\int_0^{Z_1}\varphi_Y(Z)\mathrm{d}G(Z)\right)^{\frac{1}{\sigma-1}}.$$

Since this expression is increasing in Z_1 and since a reduction in trade costs must increase Z_1, the wage of moderately skilled workers must rise relative to the P_X. Hence, the welfare impact on the moderately skilled workers in terms of their real income is ambiguous.

3.3. Discussion

I briefly discuss the empirical relevance of the implications of Proposition 7. As noted earlier, one interpretation of the variable Z is that it is an observable worker characteristic such as years of education. According to this interpretation, lower trade barriers between identical countries (1) increase the average level of education of workers in both the traded and non-traded sector, (2) increase the relative return of the most highly educated laborers, and (3) reduce the return to the moderately skilled workers. With respect to the facts as presented in Baldwin and Cain (2000), predictions (1) and (2) are consistent with movements in aggregate data, while (3) is inconsistent with the fact that the wages of the least educated workers have fallen most in the last several decades.

A problem with interpreting Z as education is that the production structure in the model is too simple to capture the input mix of actual firms. An alternative, and perhaps more appealing, interpretation of Z is as unobserved heterogeneity in worker quality. According to this interpretation, the model makes predictions over worker wages as a function of their unobserved ability after having controlled for differences in observed characteristics such as education. Inter-industry wage differentials that persist after controlling for observed worker characteristics are well documented (see, for instance, Katz and Summers, 1989). Bernard and Jensen (1999) also report a large within-industry wage premium paid to both production and non-production workers of exporting firms.

Given this interpretation, the model suggests that a decrease in transport costs alters the distribution in wages across workers of similar education by increasing the return to worker quality as observed by firms. It is workers of moderate quality rather than education that see their relative wage drop in response to falling trade barriers and the most able workers, or those originally employed at exporters, that see their wage rise in response to a reduction in trade costs. This drop is consistent with growing *intra-industry* wage differentials between workers employed at small, non-exporting plants, and workers employed at large, exporting plants. In fact, Davis and Haltiwanger (1991) report that between 1963 and 1985, the wage premium paid to production workers employed at very large plants relative to very small plants has increased by 79%. Finally, it can be shown in the model that the reduction in trade costs raises average real wages across both traded and non-traded sectors and so has an ambiguous impact on the *inter-industry* wage differential and yet might still lower the relative wages of those moderately able workers. It is possible in the context of the model that inter-industry wage differentials may remain relatively stable while a segment of workers, those that continue to be employed by small plants in manufacturing or those forced into the service industry, conclude that international trade is behind the disappearance of "good jobs, paying good wages."

4. Conclusion

In the model presented in this paper, homogeneous firms face four types of decisions: (1) entry, (2) technology choice, (3) whether or not to export, and (4) type of worker to employ. The interaction between the characteristics of competing technologies with trade costs and with worker heterogeneity gave rise to a type of firm heterogeneity that is consistent with some of the stylized facts: Exporting firms are larger, employ more advanced technologies, pay higher wages, and appear to be more productive than firms that do not export. Second, the model showed that a reduction in trade frictions can induce firms to switch technologies, leading to an expansion of trade volumes, an increase in the wage premium paid to the most highly skilled workers and a decrease in the wage premium paid to moderately skilled workers. Hence, the model offers a competing explanation for the size and non-production wage differentials between exporting and non-exporting firms as well as for the causes of the growing skill premium relative to other papers in this literature. Moreover, the model provides guidance to the empirical literature that seeks to explain the relationship between trade and observed industry productivity: Some of the observed productivity gains associated with falling transport costs may reflect changes in the allocation of heterogeneous workers across technologies rather than homogeneous workers across heterogeneous firms.

An appealing research agenda would be to combine firm and worker heterogeneity in a single framework. While some of the stylized facts concerning within industry heterogeneity can be captured in a framework that considers heterogeneous firms and representative workers, or representative firms and heterogeneous workers, many of the productivity dynamics observed in the data would better explained by model that allows for both dimensions of heterogeneity.

Acknowledgements

I thank the participants at the May 2001 Midwest International Trade conference and the March 2002 NBER ITI conference, and Yongsung Chang. I would also like to thank two anonymous referees and the Editor for their extremely valuable suggestions concerning the exposition.

Appendix A

Proof of Lemma 2. From Eqs. (4) and (5), it follows immediately from $\mathrm{d}C_L/\mathrm{d}Z_1 = S'(Z_1) < 0$. Proposition 4 established that a reduction in trade costs must increase Z_1. Totally differentiating Eq. (5), the function that determines C_H, yields

$$\mathrm{d}C_H = \left[S'(Z_1)A(Z_2)\frac{\mathrm{d}Z_1}{\mathrm{d}Z_2} + S(Z_1)A'(Z_2)\right]\mathrm{d}Z_2$$

where a prime indicates a partial derivative. It can then be shown via substitution that

$$\mathrm{d}C_H = \left[\frac{(1-\beta)S'(Z_1)\displaystyle\int_{Z_1}^{Z_2}\varphi_L(Z)\mathrm{d}G(Z) - \varphi_Y(Z_1)g(Z_1)}{\beta\dfrac{S'(Z_1)}{S(Z_1)}\displaystyle\int_0^{Z_1}\varphi_Y(Z)\mathrm{d}G(Z) - \varphi_Y(Z_1)g(Z_1)}\right][S(Z_1)A'(Z_2)]\mathrm{d}Z_2$$

The numerator first bracketed expression is positive, and the second bracketed expression is negative so $\mathrm{d}C_H \mathrm{d}Z_2 < 0$. Proposition 4 established that $\mathrm{d}Z_2/\mathrm{d}\tau$ is positive. □

References

Baldwin, R., Cain, G., 2000. Shifts in U.S. relative wages: the role of trade, technology, and factor endowments. Review of Economics and Statistics 92, 580–595.

Bartel, A., Lichtenberg, F., 1987. The comparative advantage of educated workers in implementing new technology. Review of Economics and Statistics 69 (1), 1–11.

Bartel, A., Sicherman, N., 1999. Technological change and wages: an interindustry analysis. Journal of Political Economy 107 (2), 285–325.

Bernard, A., Jensen, B., 1997. Exporters, skill upgrading, and the wage gap. Journal of International Economics 42 (1–2), 3–31.

Bernard, A., Jensen, B., 1999. Exceptional exporter performance: cause, effect, or both? Journal of International Economics 47 (1), 1–25.

Bernard, A., Eaton, J., Jensen, B., Kortum, S., 2003a. Plants and productivity in international trade. American Economic Review 94 (3), 1265–1290.

Bernard, A., Jensen, B., Schott, P., 2003. Falling trade costs, heterogenous firms, and industry dynamics. Mimeo.

Clerides, S., Lach, S., Tybout, J., 1998. Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morroco. Quarterly Journal of Economics 113 (3), 903–947.

Davis, S., Haltiwanger, J., 1991. Wage dispersion between and within U.S. manufacturing plants, 1963–1986. Brookings Papers on Economic Activity, 115–180.

Dixit, A., Stiglitz, J., 1977. Monopolistic competition and optimum product diversity. American Economic Review 67 (3), 297–308.

Ekholm, K., Midelfart Knarvik, K., 2001. Relative wages and trade-induced changes in technology. Mimeo, Stockholm School of Economics.

Idson, T., Oi, W., 1999. Workers are more productive in large firms. Papers and Proceedings-American Economic Review 89, 104–108.

Katz, L., Summers, L., 1989. Industry rents: evidence and implications. Brookings Papers on Economic, 209–275.

Manasse, P., Turrini, A., 2001. Trade, wages, and ‘superstars’. Journal of International Economics 54, 97–117.

Meltiz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6), 1695–1726.

Roberts, M., Tybout, J., 1997. The decision to export in colombia: an empirical model of entry with sunk costs. American Economic Review 87 (4), 545–564.

Wood, A., 1994. North–South Trade, Employment, and Inequality: Changing Fortunes in a Skill Driven World. Oxford Univ. Press, Oxford.