Working Papers

www.cesifo.org/wp

Heterogeneous Workers, Trade, and Migration

Inga Heiland
Wilhelm Kohler

CESIFO WORKING PAPER NO. 4387
CATEGORY 4: LABOUR MARKETS
SEPTEMBER 2013

CESifo *Working Paper No. 4387*

Heterogeneous Workers, Trade, and Migration

Abstract

We argue that the narrative of variety-induced gains from trade in differentiated goods needs revision. If producing differentiated varieties of a good requires differentiated skills and if the work force is heterogeneous in these skills, then firms are likely to have monopsony power. We show that trade then has adverse labor market effects: It increases the monopsony power of firms and worsens the average quality of matches between firms and workers. We also show that international migration has the opposite beneficial effects. Our model can explain two-way migration among similar countries, a pattern that features prominently in migration data.

JEL-Code: F120, F160, F220, J240.

Keywords: two-way migration, gains from trade, heterogeneous workers.

Inga Heiland
Ifo Institute – Leibniz-Institute for Economic Research at the University of Munich
Poschingerstraße 5
Germany - 81679 Munich
heiland@ifo.de

*Wilhelm Kohler**
University of Tuebingen
Mohlstrasse 36
72074 Tuebingen
Germany
wilhelm.kohler@uni-tuebingen.de

*corresponding author

September 5, 2013
This paper has greatly benefitted from helpful comments by Carsten Eckel, Assaf Razin, Slobodan Djajic, Chris Parsons, Christian Schwarz, Gabriel Felbermayr, Vitor Trindade and Gerhard Glomm. We also thank participants of the seminar of the University of Munich, the 27th Annual Meeting of the EEA, the Midwest International Trade Conference Spring 2013 and the 6th International Conference for Migration and Development for valuable comments and suggestions. Wilhelm Kohler gratefully acknowledges financial support from the Volkswagen Foundation under the project “Europe’s Global Linkages and the Impact of the Financial Crisis: Policies for Sustainable Trade, Capital Flows, and Migration.”

1 Introduction

Modern theory of international trade emphasizes four channels for welfare gains from trade based on product differentiation and increasing returns to scale. First and foremost, trade leads to a larger degree of product differentiation. If consumers love *variety*, they will be made better off as a result. A similar effect arises on the production side, where a larger variety of intermediate inputs may increase productivity. Secondly, as a result of trade domestic firms may become *larger*, and with economies of scale this reduces their average cost. Thirdly, trade is likely to enhance the degree of competition, lowering the *markups* charged by domestic firms, thus reducing the welfare cost of imperfect competition and potentially avoiding anti-competitive effects from larger firms. And finally, trade-induced exit of *heterogeneous firms* may increase aggregate productivity, if the least productive firms are driven out of the market.

There is a sizable body of literature highlighting various details of these effects. The model of monopolistic competition pioneered by Krugman (1980) isolates the variety effect from trade by ruling out all scale and pro-competitive effects through Dixit-Stiglitz-type CES consumer preferences. The same holds true for the standard model of trade in intermediates pioneered by Ethier (1982). Krugman (1979) proposes a preference structure featuring a variable demand elasticity, thereby allowing for the gains from enhanced variety to be accompanied by positive scale and pro-competitive effects.[1] Starting with Melitz (2003), a whole strand of literature explores conditions under which selective market entry and exit of heterogeneous firms gives rise to a positive average productivity effect from trade.[2] Feenstra (1994), Broda and Weinstein (2006) and Feenstra (2010) have put numbers on the gains that derive from variety effects. For the US, the cumulative effect of enhanced variety between 1972 and 2001 is estimated at more than 2 percent of GDP; see Broda and Weinstein (2006). For small countries, Feenstra (2010) estimates trade gains well above 10 percent of GDP.[3] The modern theoretical and empirical literature on international trade thus appears as a huge narrative of gains from trade in differentiated goods.

---

[1] More recent trade models with preferences allowing for variable markups are found in Behrens and Murata (2007) and Melitz and Ottaviano (2008).

[2] An elegant summary of this literature ist presented in Arkolakis et al. (2012).

[3] These numbers should be seen as upper bounds. They ignore the negative variety effect that arises from domestic firm exit. Empirical numbers for conventional scale and pro-competitive effects are somewhat harder to come by; see Feenstra (2010). Feenstra and Kee (2008) demonstrate that the pro-productivity effect from Melitz-type selection among heterogeneous firms may similarly be interpreted as a variety effect, having to do with the variety of markets that may be served by domestic firms.

The present paper argues that this narrative needs revision. Product differentiation very often requires customization of inputs, which implies input specificity. Existing literature has focused on specificity of *intermediate inputs*, or manufacturing components. It has produced two important insights. The first, deriving from two influential contributions by Antràs (2003) and Antràs and Helpman (2004), is that trade in such components is often plagued by a hold-up problem that may explain firms' choice of the organizational form of trade. The second insight, due to Grossman and Helpman (2005), is that input trade often requires costly search for suitable input suppliers, willing to customize inputs in the face of limited contractibility, and that the conditions for this search are likely to differ between countries, which may be an important explanatory factor for the location of input sourcing.

This paper aims at generating new insights for the case where specificity holds for the most important of all inputs, which is labor. We argue that product differentiation is the outcome of an economic mechanism matching worker skills with product requirements. Consider any given type of product from a set of differentiated goods. Workers with different types of skills are likely to constitute matches of different quality as regards the specific type of labor input required for producing this good. There will be a special type of worker skills which is best suited for producing this good. Other types of skills will be less suitable, but prove ideal for other types of goods. Looking at a heterogeneous labor force from the perspective of any one good, one might think of workers as being differently "far away" from the ideal specification of skills which would maximize the worker's productivity in the production of this good. Conversely, different types of goods are differently "far away" in terms of skill requirements from the specific skills embodied in a certain worker. It transpires from this line of argument that worker heterogeneity in skill types is conducive for product differentiation.[4] But the prevalence of fixed costs works against each type of worker producing her distinct, ideal product variety. If production requires a fixed cost, then each firm's employment will typically include workers that deviate from the ideal type of skills. The overall productivity of an economy then depends on how well workers with different types of skills are matched with firms producing different types of goods. Moreover, each firm will have monopsony power in the labor market in that it faces an upward sloping supply of labor. The intuition is that expanding employment means bidding away labor with

[4] Note that we focus on horizontal product differentiation and, thus, on different *types* of skills, as opposed to different skill *levels*. This view on heterogeneity in skills on one hand and job requirements on the other has, for example, been used to in the literature to analyze employment (Bhaskar and To, 1999; Marimon and Zilibotti, 1999; Thisse and Zenou, 2000; Fiorillo et al., 2000), wage determination (Hamilton et al., 2000; Bhaskar and To, 2003), and the spatial distribution of economic activity (Amiti and Pissarides, 2005).

ever less suitable skills for one's own variety, and ever more suitable skills for a competitor firm's variety.

Under these conditions, trade and migration have effects that have so far largely gone unnoticed in the literature. First, if trade reduces the number of firms in any one economy, this has an adverse effect on the average quality of matches between workers and firms, which runs counter to the economies of large scale production emphasized by modern trade literature.[5] And secondly, it aggravates the labor market distortion caused by monopsony power, meaning a larger markup between the marginal productivity of labor and the wage rate, which runs counter to the pro-competitive effects on goods markets highlighted in variable markup models of modern trade theory. The same logic leads us to expect that international migration of labor has the opposite effects of improving the average quality of labor market matches and reducing the monopsony power of firms on the labor market.

We develop a formal model that allows us to analyze these effects. Following Amiti and Pissarides (2005), we assume a pre-existing horizontal differentiation of skills. A certain specification (or type) of skill may be interpreted as a unique composition (or bundle) of different abilities that is embodied in a worker and cannot be unbundled. Firms can only use bundles as embodied in workers, and workers are remunerated depending on their idiosyncratic bundles of characteristics.[6] A certain variety of a good similarly represents a bundle of commodity characteristics.[7] A key tenet of our approach is that the differentiation of skills represented by a heterogeneous workforce is a natural prerequisite for, and at the same time a natural limit to, the production of differentiated goods.

We employ a circular representation of worker skills which is combined with a circular view of product differentiation.[8] An important aspect of this is that each specification of skills (or differentiated variety of a product), measured by a certain position on the circle, features the same average similarity to all other types of worker skills (or product variety). Moreover, for each variety on the product circle there is a unique corresponding point on the skill circle, which

---

5 Notable exceptions are the papers by Eckel (2009a,b) that highlight a negative effect on productivity deriving from a trade-induced decrease in the number of intermediate goods producers.

6 This notion of bundles of characteristics is familiar from models of sorting; see Mandelbrot (1962) and Welch (1969). For a recent application of this idea to trade, see Ohnsorge and Trefler (2007).

7 This view of product differentiation goes back to Lancaster (1966).

8 This goes back to the circular city model by Vickrey (1964), Vickrey et al. (1999) and Salop (1979). For trade applications, see Helpman (1981), Grossman and Helpman (2005) Eckel (2009a,b).

identifies a perfect match between worker skills and product characteristics. Workers located elsewhere on the skill circle may still be used in producing this product, but they will be less productive. The farther away their position from the ideal match, the lower their productivity. For simplicity, we assume a single sector of differentiated goods. Production of each variety is subject to a fixed cost.

Profit maximization and free entry of firms determine the equilibrium number of firms as well as the positions of the differentiated products that they produce along the circle that measures product differentiation. All worker characteristics are observable and employment contracts are perfect. Heterogeneity of workers in the neighborhood of a firm's perfect match on the circle of worker skills means that from a firm's perspective labor supply is finitely elastic with respect to the wage rate. Consequently, firms perceive a marginal cost of employment which is above the ongoing wage rate, which in turn is equal to the marginal opportunity cost of employment to the society. This adds to the markup distortion in firms' pricing behavior, increasing the wedge between prices and the true marginal cost. Our model highlights a further distortion which relates to the entry decision. When considering market entry or exit, firms take as given the observed average quality of matches between worker skills and product requirements. But any entry of a further firm lowers the distance between firms on the circle of differentiated worker skills, thus increasing in the average quality of matches on the labor market, which in turn lowers average cost. Thus, entry of firms involves a positive externality.[9]

Our model allows us to readdress the general equilibrium effects of trade and migration that derive from horizontal skill heterogeneity of the labor force. With several distortions deriving from such heterogeneity, the welfare effects of trade and migration seem ambiguous a priori. We show that trade-induced specialization brings about a loss in efficiency through a lower average matching quality in the labor market as well as through higher wage markups from lower competition. However, comparing free trade with autarky, we prove that the conventional gains from enhanced goods market competition and a higher degree of variety dominate the adverse effects on the labor market. The gains from trade theorem survives. We also show that in such a world firms will have an incentive for cross-border hiring, and workers will have an incentive to migrate. Migration is beneficial for all countries in that it essentially undoes the adverse labor market effects of trade. We prove that the trade and migration equilibrium will always be preferable in terms of aggregate welfare to an equilibrium with free trade alone. Piecemeal trade

[9] Such an externality is also present in Helsley and Strange (1990).

liberalization entails lower markups on account of a lower perceived price elasticity of demand for goods, but it comes with an increase in monopsony power leading to higher wage-markups in the labor market as the number of domestic firms falls. We demonstrate that even in our single sector model such piecemeal goods market liberalization need not be beneficial. In contrast, piecemeal integration of labor markets does turn out to be unambiguously welfare increasing for all countries.

Our work is related to existing literature in a number of different ways. First, it relates to a long strand of literature dealing with the relationship between trade and migration. Traditional theory of endowment-based comparative advantage holds that trade and migration are substitutes, highlighted in the strongest possible way through the famous factor price equalization theorem. At the same time, it stresses that trade may lead to factor price divergence if caused by non-endowment based comparative advantage, say technological differences or scale effects; see Markusen (1983). Then, if migration responds to wage differences, trade may be said to be complementary to, or even cause, migration. Empirical evidence strongly favors the view that trade and migration are "non-substitutes"; see Felbermayr et al. (2012). However explaining this complementarity primarily along the lines of trade causing international factor price divergence seems questionable. Our model highlights a different view of migration and identifies a novel cause of non-substitutability between trade and migration. Moreover in our model trade does not cause factor price divergence, and migration occurs even if the no-migration equilibrium involves no observed differences in wages.

Our work also adds to the literature on migration in contributing to a better understanding of certain characteristics in recent migration flows. In particular, recent migration flows feature high-skilled labor moving between industrialized counties, and such flows are often *two-way* in nature. According to Docquier and Marfouk (2006), 33.7% of the stock of skilled migrants residing in OECD countries stem from other high-income countries.[10] Figure 1 plots bilateral stocks of immigrants and emigrants by country pair in (or around) the year 2000.[11] We plot the original data and residuals (both are in logs) obtained from a gravity-type estimation often used in the literature to explain cross-country patterns of migration, see. e.g. Ortega and Peri (2012), Mayda (2010), Beine et al. (2011). The residual migration stocks are free of any country-specific influences as well as the effect of distance, contiguity, economic integration agreements

[10] Cp. Table 5.3. in Docquier and Marfouk (2006).

[11] The data stem from the *Database on Immigrants in OECD countries (DIOC)* provided by the OECD.

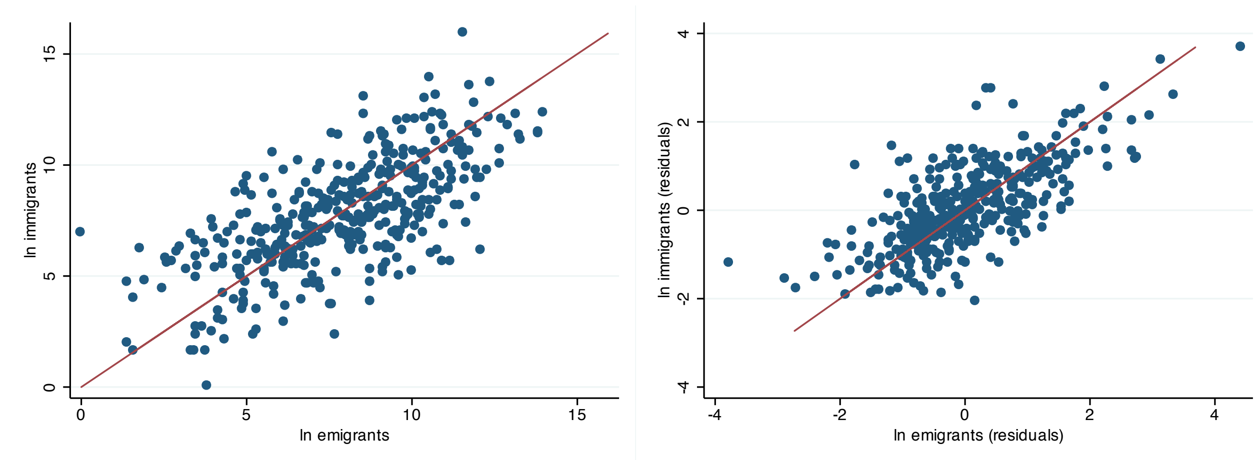

Figure 1: Correlation of bilateral stocks of immigrants and emigrants, OECD-DIOC 2000

and a common language. The correlation between original and residuals stocks is striking, with observations fairly closely aligned along the 45-degree line, indicating that a large number of immigrants from a certain country comes along with a large number of emigrants to that same country. A look at bilateral migration data from other sources reveals the exact same pattern and that the strong correlation is present not only in stocks but also in flows.[12] Existing theories of migration have a hard time explaining such two-way patterns of migration. They cannot be dismissed on empirical grounds as reflecting heterogeneity of observed migration in terms of skill-levels. Plotting residual stocks along the lines of figure 1 for different education levels reveals the same patterns.[13] Nor is the two-way element in migration flows entirely new, found only in recent migration data. It features prominently among the "Laws of Migration" that Ravenstein (1885) has carved out of a detailed account of 19th century flows of migration between different manufacturing districts in the UK. He calls it "counter-currents of migration," and he insists that it is driven by "business considerations."

The challenge, then, is to explain such business considerations that lead to two-way migration of labor with equal skill *levels* between countries facing similar endowments. Our model suggests a potentially powerful explanatory factor which is different from those hitherto proposed in

[12] We have looked at bilateral stocks from the *Global Bilateral Migration Database* recently issued by The World Bank, bilateral stocks of inter-US-state migrants from the *American Community Survey*, and bilateral inter-US-state migration flows from the *Statistics of Income Division* of the *Internal Revenue System*. Results confirming the picture conveyed by figure 1 may be obtained from the authors upon request.

[13] Also, using the DIOC data disaggregated by age groups, we do not find qualitative differences to the pattern in aggregate data. Again, detailed results may be obtained upon request.

the literature. The largest part of the literature that considers two-way migration rests on assumptions about heterogeneity in individuals' preferences or factor endowments. These models can explain different valuations of country-specific amenities, institutions, or economic outcomes, thus rationalizing the existence of migration incentives going both ways between *dissimilar* countries.[14] Schmitt and Soubeyran (2006) present a model that predicts two-way migration of individuals within *occupations*. But in their model, individuals with the *same level of skills* would never move in both directions, and even within-occupation movements are observed only between countries that differ in the relative endowment with skills.

The literature dealing with two-way migration between *similar* countries is scant. In Fan and Stark (2011), individuals migrate to escape the burden of social stigma arising from employment in an occupation of low social status. If individuals perceive the associated humiliation as less severe if working as a "foreigner" in the society of the immigration country than if working as a member of the "socially close" society in the home country, migration will occur even among similar countries. A further explanation is proposed by Kreickemeier and Wrona (2013) where two-way migration among similar countries emerges as a signaling device. Workers are vertically differentiated in their skill levels, technology requires formation of teams and it features complementarity between skill levels of different team members. Firms will thus want to hire teams through matching workers with similar skill levels. However, individual skill levels cannot be observed, hence the need of a signaling device. If migration is costly, then highly skilled workers can signal their above average skill through migration. Migration then promises to be matched to another above average skilled migrant thereby receiving a higher wage. As in our model, two-way migration does occur within groups of similarly skilled individuals and migration helps solving a labor market imperfection. However, in their model migration is driven by *vertical* skill differentiation among workers, whereas we focus on *horizontal* differentiation related to product differentiation, which, as argued in the introduction, is the hallmark of modern trade. This allows us to investigate the link between migration and diversity-driven trade.

Finally our work contributes to the small but growing literature that emphasizes the quality

[14] For example, Tabuchi and Thisse (2002) propose heterogeneity in the valuation of location-specific amenities as an explanation for two-way migration. In Galor (1986), individuals differ with respect to the rate of time preference and two-way migration emerges among countries with different interest rates. Gaumont and Mesnard (2000) show that differences in relative factor prices might lead to two-way migration when individuals are heterogeneous with respect to the degree of altruism. In Berninghaus and Seifert-Vogt (1991), two-way migration arises if there is heterogeneity in individuals' expectations about future living conditions in the home country and the foreign country.

of worker-firm matches as an important dimension, besides unemployment and relative wages, along which globalization affects labor markets; see Davidson et al. (2011, 2012). Davidson et al. (2008) develop a model with worker heterogeneity in talent and heterogeneity among firms in terms of productivity. They show that openness can enhance the quality of worker-firm matches by reallocating market shares to more productive firms, thereby increasing their ability to attract the highly talented workers and secure complementarities that would otherwise be left unexploited. We take a different look at worker heterogeneity, considering differences in the types of skills rather than levels of talent. On the firm side we similarly emphasize horizontal heterogeneity in terms of the required specification of skills deriving from horizontal product differentiation. Matching quality then reflects how well the specification of available worker skills on average fits with the specific skill requirements generated by firms that are engaged in horizontal product differentiation.

The remainder of the paper is organized as follows. In section 2 we describe the general model framework and characterize the autarky equilibrium. Then we discuss the effects of a transition to free trade and the scenario of piecemeal trade liberalization in section 3. In section 4 we introduce labor mobility and analyze the effects of migration. Section 5 concludes.

2 The Model Framework

We look at a one sector economy, in which firms produce differentiated varieties using specific types of skilled labor. Firms are assumed to be small compared to the market, hence they take aggregate parameters such as income, the total number of firms and the price index as given. We assume a preference structure where the perceived price elasticity of demand varies with firm size, which implies that the mark-up of prices over marginal cost is endogenous. The technology for producing a specific variety of the good is defined in terms of a unique ideal combination of worker skills. Given worker heterogeneity, expanding employment implies hiring workers of ever less suitable skills. We use a circular representation of varieties and worker heterogeneity, analogous to the Lancaster (1966) model. This "spatial structure" of the labor market implies strategic interaction among neighboring firms, as in the classical Hotelling model of spatial competition.

2.1 Utility and demand

Individual k derives utility from consumption of a bundle $C[c_k]$ of differentiated varieties $c_k = [c_{1k}, ..., c_{ik}, ..., c_{Nk}]$, where the N denotes the number of varieties available.[15] Throughout the paper, we use brackets $[\cdot]$ to collect arguments of a function and parentheses to collect algebraic expressions. We assume that $C[c_k]$ is homogeneous of degree one. The logarithmic indirect utility function is given by

$$\ln V_k = \ln y_k - \ln P[p], \tag{1}$$

where $P[p] = P[p_1, ...p_i, ..., p_N]$ is the minimum unit expenditure function for all varieties i, and y_k denotes income of individual k. Following Diewert (1974) and Bergin and Feenstra (2000), we assume that preferences are characterized by a symmetric translog expenditure system. The translog expenditure system implies a variable demand elasticity and a minimum unit expenditure function which is homogeneous of degree one.[16] The unit expenditure function is given by

$$\ln P[p] = \frac{1}{2\gamma N} + \frac{1}{N}\sum_{i=1}^{N} \ln p_i + \frac{\gamma}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N} \ln p_i(\ln p_j - \ln p_i). \tag{2}$$

The parameter $\gamma > 0$ controls the degree of substitutability between varieties, a larger γ implying higher substitutability.[17] Using Roy's identity, the Marshallian demand function for variety i can be derived from the indirect utility function according to

$$x_{ik}[p, y_k] = \frac{\partial \ln P[p]}{\partial \ln p_i}\frac{y_k}{p_i} = \delta_i \frac{y_k}{p_i}, \tag{3}$$

---

[15] We treat N as a continuous variable, assuming that the integer constraint is negligible.

[16] Recent applications of the symmetric translog expenditure system are Feenstra and Weinstein (2010), Arkolakis et al. (2010) and Ródriguez-López (2011). As Feenstra and Weinstein (2010) point out, another interesting feature of the this expenditure system is that it constitutes a second order Taylor approximation of any symmetric expenditure function.

[17] Feenstra (2003) shows that the translog expenditure function does not depend on reservation prices of varieties that are not available if it is *symmetric*. This property, which is shared for example with the CES demand system, is very useful in the context of models with an endogenous number of firms where the number of available varieties is smaller than the number of potential varieties. The specification that is used here and is borrowed from Arkolakis et al. (2010) assumes that the number of *potential* varieties is infinite. Then, the parameter a_0 in Feenstra's (2003) original notation simplifies to $\alpha_0 + \frac{\gamma}{2N}$, where N is the number of available varieties, and the number of potential varieties does not enter the expenditure function. Note also that α_0 has been normalized to zero.

where

$$\delta_i = \frac{1}{N} + \gamma \left(\frac{1}{N} \sum_{j=1}^{N} \ln p_j - \ln p_i \right) \tag{4}$$

is the expenditure share for variety i. Using Y to denote aggregate income and writing $\overline{\ln p} := \sum_i \ln p_i / N$, total demand for this variety i then follows as

$$q_i[p_i, \overline{\ln p}, N, Y] := \delta_i[p_i, \overline{\ln p}, N] \frac{Y}{p_i}. \tag{5}$$

Assuming that firm i treats the average log-price $\overline{\ln p}$ as given, the *perceived* price elasticity of demand for variety i emerges as

$$\varepsilon_i[p_i, \overline{\ln p}, N] := -\frac{\mathrm{d} \ln q_i}{\mathrm{d} \ln p_i} = 1 - \frac{\mathrm{d} \ln \delta_i}{\mathrm{d} \ln p_i} = 1 + \frac{\gamma}{\delta_i}. \tag{6}$$

With this perceived demand elasticity, the markup of prices over marginal cost will be an endogenous variable.

2.2 Skill heterogeneity and labor supply

The economy is endowed with a mass L of workers, which are differentiated by the types of skills they possess. We assume that the skill space is characterized by a circle with circumference $2H$, where each location on the circle reflects a skill type, and types that are more similar are located closer to each other on the circle. We may interpret H as the degree of differentiation of skills. Furthermore, we assume a uniform distribution of the labor force L over the entire circle. This implies that a mass of $\frac{L}{2H}\mathrm{d}s$ workers is located within an interval of length $\mathrm{d}s$ on the skill circle.

Firms produce differentiated varieties, whereby each variety has its distinct optimal type of skill $\bar{s}_i$ that corresponds to a specific location on the circle. Firms' locations on this cycle will be determined endogenously. Importantly, firms can also employ workers with skills different from their optimal types. However, the farther away a worker's skill type from the optimal type, the less productive this worker in the production of the firms' particular variety. Since skill types are ordered according to similarity, the productivity of a worker is falling in the distance of her skill type from the optimal type in the skill space. We assume that a worker with a skill that is a distance d away from the optimal type can supply $f[d]$ efficiency units of labor to this firm, where $f'[d] < 0$ and $f''[d] \leq 0$. In other words, the marginal "productivity penalty" on distance is increasing. This concavity will prove important below. Without loss of generality, we set the number of efficiency units supplied by the ideal worker to unity, $f[0] = 1$. We rule out the possibility that workers supply negative units of labor, which, given $f[\cdot]$ and $f[0] = 1$, requires a restriction on the degree of skill differentiation H, so that $f[H] > 0$.[18]

Skills are observable and wages are denoted as per efficiency unit of labor. A worker at distance d from firm i's ideal type obtains a reward of $w_i f[d]$ where w_i is the wage per efficiency unit set by the firm. Intuitively, with workers located uniformly across the entire skill circle, expanding employment always implies moving farther away from the ideal skill type, thus hiring workers with ever lower efficiency $f[d]$. At the same time, expanding employment implies bidding away workers from neighboring firms. A worker chooses to work with the firm where her income is maximized. Hence, the sorting of workers into firms depends on the worker-firm specific productivity *and* firms' wages per efficiency unit. Figure 2 depicts the income schedule of

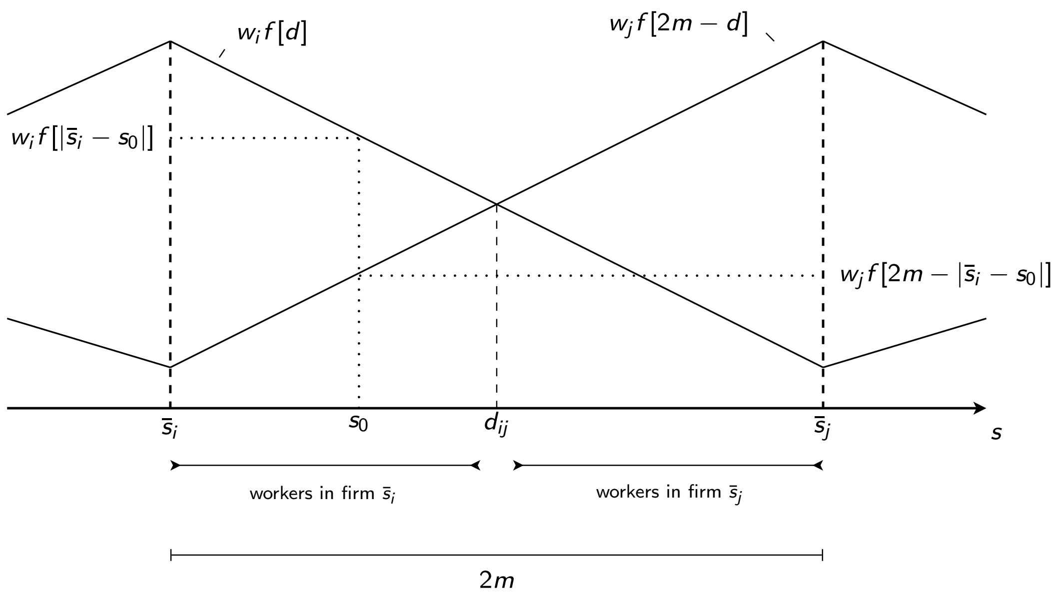

Figure 2: The wage schedules

workers on a fragment of the skill circle around the location of two firms with optimal skill types $\bar{s}_i$ and $\bar{s}_j$, which are at distance $2m$ from each other, if these firms set wages equal to w_i and

[18] Strictly speaking, the degree of skill heterogeneity measured as the productivity difference between the ideal worker and the worker at the farthest skill reach is determined by H in relation to the slope and the position of $f[\cdot]$. $f[H] > 0$ could alternatively be assured by restricting the slope of $f[\cdot]$ or by scaling $f[0]$.

w_j, respectively.[19] For easier drawing, but without loss of generality, the figure assumes that $f[\cdot]$ is linear. A worker with skill type s_0 obtains income $w_i f[|\bar{s}_i - s_0|]$ in firm i and income $w_j f[2m - |\bar{s}_i - s_0|]$ in firm j which is located at distance $2m$ from firm i. The worker at distance d_{ij} is indifferent to working for either firm, all workers to his left will want to work with firm i, and all workers to his right will prefer working for firm j. The same logic applies to the neighboring firm to the left of firm i. In the equilibrium described below, m will be a crucial variable to be determined endogenously. Obviously, it is closely related to N; the number of firms.

This type of worker sorting generates an upward sloping labor supply curve for firm i which depends on the distance to, and the wages set by, its "neighboring" firms. Since we shall describe an equilibrium where all firms are symmetric, we continue by describing firm i's labor supply assuming that both neighboring firms set the same wage $\bar{w}$. Then, the skill distance d_i from its ideal skill type that firm i is able to cover in either direction by setting a wage w_i is implicitly given by

$$w_i f[d_i] = \bar{w} f[2m - d_i]. \tag{7}$$

We use $d_i = d[w_i, \bar{w}, m]$ to denote the solution to this equation. Obviously, d_i is increasing in w_i and m, but falling in $\bar{w}$.

The entire amount of efficiency units that firm i is able to attract through setting a wage rate w_i can now be derived as twice the integral over all efficiency units $f[d]$ from distance zero up to distance d_i. Firm i's labor supply schedule thus emerges as

$$L_i^{ES}[w_i, \bar{w}, m, L, H] = 2\int_0^{d[w_i,\bar{w},m]} f[d]\frac{L}{2H}\mathrm{d}d = \frac{L}{H}\int_0^{d[w_i,\bar{w},m]} f[d]\mathrm{d}d. \tag{8}$$

Hence, each firm faces a labor supply function for efficiency units which is increasing in its own wage. We shall henceforth use $\eta_i = \eta_i[w_i, \bar{w}, m]$ to denote the own wage elasticity of L_i^{ES}. It is obvious that this elasticity is positive. Further properties of η_i will be explored below.

Referring to figure 2 we note that this view of the labor market implies a surplus of employ-

[19] Anticipating a symmetric equilibrium, the distance m is introduced as a variable which is uniform across all pairs of firms. Moreover, for simplicity the figure also assumes that $w_i = w_j$. Again, this equality will prevail in the symmetric general equilibrium described below, but for the *sorting* pattern described in the figure it is not essential.

ment that derives from diminishing worker efficiency. In the figure, this surplus is measured as the area between the wage curve of firms i and j and a horizontal line through $w_i f[d_{ij}]$ between $\bar{s}_i$ and d_{ij} and $\bar{s}_j$ and d_{ij}, respectively. This entire surplus is appropriated by workers. This begs the question of why firms shouldn't be able to wage-discriminate among workers who have different bargaining power, depending on their specific distance from the neighboring firm. However, since we assume free entry, each worker has the outside option of working for a potential entrant taking the exact same position of the wage discriminating firm, but undercutting it by accepting a lower share of the surplus in offering higher wages than a wage-discriminating firm. In short, zero profits also imply zero job surplus to the firm. As Amiti and Pissarides (2005) point out, this will lead firms to pay wages that are proportional to marginal products.[20]

2.3 Firm behavior

Firm behavior involves three subsequent steps: market entry, location of product specification and pricing. We assume, that the technology of production is symmetric across firms. Setting up production at a certain point on the circle requires a fixed labor input α, defined in terms of efficiency units of the corresponding ideal skill type on the skill circle. In addition, production requires β units of this input per unit of the good produced. Thus, the total labor input required by firm i in order to produce q_i units of its variety is equal to $\alpha + \beta q_i$.

Firm behavior is characterized by the following profit maximization problem:

$$\begin{aligned} & \max_{p_i} \{p_i q_i - w_i(\alpha + \beta q_i)\} \\ \text{s.t.:} \quad & q_i = q_i[p_i, \overline{\ln p}, N, Y] \quad \text{and} \\ & \alpha + \beta q_i[p_i, \overline{\ln p}, N, Y] = L_i^{ES}[w_i, \bar{w}, m, L, H] \end{aligned} \tag{9}$$

We assume that firms pursue Bertrand strategies on both the goods and the labor market, meaning that they take the prices and wages set by their competitors as given. Moreover, we assume small firms that take the number of firms N and their average log-price $\overline{\ln p}$, as well

[20] A different avenue is taken by Kim (1987, 1989), who assumes that the wage is given by the axiomatic Nash bargaining solution with equal bargaining power (before considering the outside option) instead of a take-it-or-leave-it rule. Both schedules yield the same result if the outside option is a potential entrant that threatens the wage discriminating firm with Bertrand wage competition. In the aforementioned papers, however, it is assumed that the outside option is the wage offer from the neighboring firm. This leads to a flat wage schedule, where the wage corresponds to the marginal product of the worker with the worst match.

as the location of their neighboring firms and income Y as given. The first order condition for profit maximization requires that perceived marginal revenue is equal to perceived marginal cost. It is relatively straightforward to show (see appendix A.1) that this condition reads as

$$p_i = \frac{\varepsilon_i}{\varepsilon_i - 1} \frac{\eta_i + 1}{\eta_i} w_i \beta. \tag{10}$$

Pricing thus involves a double markup. The second fraction represents the markup that derives from the firm's monopsony power on the labor market, whereby $\eta_i = \eta_i[w_i, \bar{w}, m]$ is the elasticity of labor supply as given in (8). Since η_i is unambiguously positive, this markup is larger than 1. We shall henceforth use $\tilde{w}_i[w_i, \bar{w}, m] := [(\eta_i + 1)/\eta_i] w_i \beta$ to denote perceived marginal cost. The first fraction in the pricing rule (10) represents the markup that derives from the firm's price setting power on the goods market, and it is similarly larger than 1, since $\varepsilon_i > 1$. From (6), we may write this markup as $\frac{\varepsilon_i}{\varepsilon_i - 1} = 1 + \frac{1}{\gamma N} + \overline{\ln p} - \ln p_i$. In appendix A.1 we show that using the *Lambert function*[21], as in Arkolakis et al. (2010), delivers the explicit solution for the optimizing price p_i as

$$p_i = \mathcal{W}\left[Z_i\right] \tilde{w}_i[w_i, \bar{w}, m] \qquad \text{with } Z_i = \frac{1}{\tilde{w}_i[w_i, \bar{w}, m]} \exp\left\{1 + \frac{1}{\gamma N} + \overline{\ln p}\right\}. \tag{11}$$

The term Z is a "summary measure" of the conditions that firm i faces on the labor market as well as the goods market. Given $\mathcal{W}_Z > 0$, a higher average log-price of the firm's competitors as well as a lower degree of substitutability (low value of γ) or a smaller number of firms leads to a higher markup over perceived marginal cost $\tilde{w}_i$, whereas the markup is falling in perceived marginal cost. The markup pricing described in (11) may be written as $p_i = p_i[w_i, \bar{w}, m, N, \overline{\ln p}]$. Inserting this price into (5), the firm-specific labor market clearing condition emerges as

$$\alpha + \beta q_i[p_i[w_i, \bar{w}, m, N, \overline{\ln p}], \overline{\ln p}, N, Y] = L_i^{ES}[w_i, \bar{w}, m, L, H]. \tag{12}$$

This condition implicitly determines the wage set by firm i, w_i, as a function of m, the half-distance to each of its neighboring firms on the skill circle, and $\bar{w}$, the symmetric wage set by these neighboring firms, as well as the number of firms N, aggregate income Y and the labor

[21] The Lambert function $\mathcal{W}[z]$ defines the implicit solution to $xe^x = z$ for $z > 0$. Furthermore, it satisfies $\mathcal{W}_z = \frac{\mathcal{W}[z]}{(\mathcal{W}[z]+1)z} > 0$, $\mathcal{W}_{zz} < 0$, $\mathcal{W}[0] = 0$ and $\mathcal{W}[e] = 1$. Here as elsewhere in the paper, we use a subscript index to indicate partial derivatives whenever this proves convenient without causing confusion.

force parameters L and H:

$$w_i = w_i[\bar{w}, m, \overline{\ln p}, N, Y, L, H] \tag{13}$$

Equations (11) and (13) together describe an individual firm's pricing behavior on the goods as well as the labor market. Returning to (9), the firm's profit may now be expressed as a function of these same variables: $\pi_i = \pi_i[\bar{w}, m, N, \overline{\ln p}, Y, L, H]$.

2.4 Symmetric autarky equilibrium with free entry

We have assumed above that preferences are symmetric, technology is symmetric across firms and the labor force is uniformly distributed around the circle. Under these assumptions, prices and wages will be the same across firms: $p_i = p$, with $\overline{\ln p} = \ln p$, as well as $w_i = w$ with $\bar{w} = w$. Moreover, firms will locate on the skill circle at equal distances from each other, anticipating the above pricing decision in the final stage of decision making. In a symmetric equilibrium the number of firms is inversely related to the distance between two firms on the circle, which is equal to $2m$. Given a circumference of the circle equal to $2H$, we may thus write $N[m] := H/m$.

We now proceed in simplifying description of firm behavior exploiting these symmetry assumptions, to be followed by the determination of m in a general equilibrium that features free entry of firms. Assuming an ex ante zero outside option and an infinite number of potential entrants, a free entry autarky equilibrium is characterized by zero profits. For ease of notation, we shall henceforth drop L and H in the collection of arguments for the various functions involved.

To pave the ground for discussing the existence and properties of this equilibrium, we first look at the elasticity of labor supply as given in (8), evaluated at $w_i = w$:

$$\eta[m] := \left. \frac{\partial L_i^{ES}}{\partial w_i} \frac{w_i}{L_i^{ES}} \right|_{w_i = w} = \frac{f[m]^2}{2F[m]} \frac{-1}{f'[m]} \tag{14}$$

where $F[m] := \int_0^m f[d]\mathrm{d}d$. Our assumption that $f''[m] \leq 0$ ensures that the labor supply elasticity falls in m.[22] Thus, firms' monopsony power in the labor market increases as firms become larger and the number of firms falls. This is an important result in view of Economides

[22] This follows directly from

$$\frac{\partial \eta[m]}{\partial m} = \frac{-f[m]}{F[m]} - \frac{f[m]^2}{2\left(F[m]f'[m]\right)^2}\left(-f[m]f'[m] - F[m]f''[m]\right).$$

(1989), who discusses the possibility of negative maximum profits in models of a circular economy with Bertrand behavior and sequential decision making, deriving from price-undercutting or wage-bidding, respectively. This is ruled out if competition intensifies as firms move closer together.

Symmetry implies $\overline{\ln p} = \ln p$ in equation (4), hence the expressions for ε and δ simplify considerably, and the profit maximizing price (11) may be written as

$$p[m] = \rho[m]\psi[m]\beta, \tag{15}$$

$$\text{where} \quad \rho[m] := 1 + \frac{1}{\gamma N[m]} \quad \text{and} \quad \psi[m] := \frac{\eta[m]+1}{\eta[m]}. \tag{16}$$

In (15) we have normalized the wage per efficiency unit to 1. We are free to do so, since our equilibrium is homogeneous of degree zero in nominal prices; see appendix A.1.

Next, we introduce θ to denote the average productivity of workers. Given a uniform distribution of the workforce around the circle, we have

$$\theta[m] = \frac{1}{m}\int_0^m f[d]\mathrm{d}d. \tag{17}$$

Notice that we have $\theta_m = (f[m] - \theta[m])/\, m < 0$ since $f'[m] < 0$. Given our wage normalization, $\theta[m]$ equals average income per worker, whence aggregate income emerges as $Y = L\theta[m]$, and output per firm is

$$q[m] = \frac{1}{N[m]}\frac{L\theta[m]}{p[m]}. \tag{18}$$

Given our normalization $w = 1$, the zero profit condition requires

$$p[m] = \frac{\alpha + \beta q[m]}{q[m]}. \tag{19}$$

Without loss of generality, we may now scale units, such that $\beta = 1$. The labor market clearing condition may then be written as $\alpha + q[m] = \frac{L}{N[m]}\theta[m]$, and aggregate variable labor input is $N[m]q[m] = L\theta[m] - \alpha N[m]$. Substituting these expressions in (19), we obtain the following representation of the zero profit condition:

$$p[m] = g[m] := \frac{L\theta[m]}{L\theta[m] - \alpha N[m]}. \tag{20}$$

Note that $g[m] > 1$ is the usual measure of the degree of economies of scale, i.e., the ratio of average to marginal cost, applied to the economy at large.[23] With zero profits, this ratio must be equal to the price relative to marginal cost. With $w\beta = 1$ from our scaling and normalization, this is exactly what we have in equation (20). Combining this with the Bertrand pricing equation in (15), we finally arrive at the following condition that determines m, the half-distance at which firms symmetrically locate on the skill circle in an autarky equilibrium:

$$g[m] = \rho[m]\psi[m]. \tag{21}$$

This is core condition that we use in the subsequent comparative static analysis. Note that $\rho_m > 0$ and $\psi_m > 0$, i.e. both markups increase in firm size and fall if the number of firms increases and competition in the respective market intensifies; see appendix A.2. The degree of economies of scale increases in firm size, hence it falls in the distance between neighboring firms, $g_m < 0$. As we discuss in appendix A.2, existence of a meaningful equilibrium with a unique value of $m > 0$ requires that, given the degree of skill heterogeneity H, the fixed cost α is not too large relative to L, the size of the labor force.

In the symmetric equilibrium described in (21), the utility of a worker who receives average income $\theta[m]$ is equal to

$$\ln V = \ln \theta[m] - \left(\frac{1}{2\gamma N[m]} + \ln p[m] \right). \tag{22}$$

Intuitively, welfare is rising in income and the number of firms in the market, and is falling in the price of a typical variety of goods. Note that all of these variables are determined by the equilibrium value of m. While we know from above that θ and N are both falling in m, the relationship between p and m is ambiguous at this stage of our analysis. More importantly, m is an endogenous variable, hence equation (22), while revealing, is no comparative static result. Before moving to a comparative static analysis in section 3 below, we address the question of whether a laissez faire equilibrium incorporates an optimal value of m. Given the multiple distortions present in this economy, the expected answer is “No.” In the next subsection, we discuss these distortions in more detail, establishing the conclusion that the laissez faire

[23] Given our scaling assumption $\beta = 1$ and the normalization $w = 1$, $L\theta[m]$ is the economy-wide total cost, while $\alpha N[m]$ is the aggregate use of labor for fixed cost, both expressed in efficiency units of labor. Hence, the right-hand side of (20) is the aggregate equivalent to the ratio of average to marginal cost.

equilibrium involves a sub-optimally large value of m, which implies excess firm entry.

2.5 Distortions

The equilibrium described above involves four distortions. (i) When considering market entry, firms fail to take into account the positive effect of their entry on welfare through a larger number of varieties. Following Dixit and Stiglitz (1977), this is often referred to as "consumer-surplus distortion." (ii) Moreover, potential entrants ignore the positive effect on average productivity, arising from a better quality of matches in the labor market. This is novel in the present model, relative to standard models of monopolistic competition, and we call it the "productivity distortion." Both, distortions (i) and (ii) constitute positive externalities, working towards insufficient entry in a laissez faire equilibrium. But entry also has negative externalities, having to do with markups on the goods and labor markets. More specifically, (iii) potential entrants anticipate both, a goods price markup as well as a wage markup, but fail to see that they will realize operating profits on such markups only at the expense of incumbent firms, due to the overall resource constraint. Following Mankiw and Whinston (1986), this may be called the "business-stealing" effect. And finally, (iv) potential entrants fail to anticipate that their entry will reduce the magnitudes of these same markups, due to enhanced competition. In a zero profit equilibrium, operating profits compensate for fixed cost, hence this "pro-competitive" effect, as well as the "business-stealing" effect, works towards excessive entry.

As is well known, in the standard CES version of the monopolistic competition model distorsions (i) and (iii) offset each other and firm entry is efficient. In appendix A.4 we show that in this model the net result of distortions (i)-(iv) is excess entry. Thus, the model inherits the "excess entry" result established by Salop (1979) for the circular city model.[24] Moreover the result is in line with Bilbiie et al. (2008), who find that in a monopolistic competition equilibrium with symmetric translog preferences the business-stealing effect dominates the consumer-surplus effect, giving rise to excess entry.[25] The excess-entry result plays a crucial role in the determination of the gains from globalization below, as those unfold partly through a mitigation of distortions.

Before we proceed to the analysis of trade and migration, it is instructive to look at the role

[24] As an example for circular labor markets, see Helsley and Strange (1990).

[25] A further case in point has been established for preferences of the constant absolute risk aversion by Behrens and Murata (2012).

of worker heterogeneity by comparing this model to the benchmark case of monopolistic competition with homogeneous workers. The presence of worker heterogeneity and labor specificity reduces the gains from specialization in differentiated varieties. When expanding production, firms must resort to workers whose skills are ever less suitable for their own variety and ever more suitable for their neighboring competitors. The degree of worker heterogeneity is measured by H. The more heterogeneous the work force, other things equal, the stronger the monopsony distortion on the labor market. We show in appendix A.3 that the model converges to the standard Krugman-model of monopolistic competition, as H converges to zero (homogeneous workers).

3 Symmetric trading equilibrium

3.1 Free trade

We first compare autarky to free trade. In the next subsequent section we shall then look at piecemeal trade liberalization. We assume that there are k symmetric countries and we denote the total number of firms worldwide by $N^T := k \cdot N$. Absent any trade barriers, prices for domestic and imported goods are equal, and given by

$$p[m] = \left(1 + \frac{1}{\gamma k N[m]}\right) \psi[m]. \tag{23}$$

Note that $\psi[m]$, the functional relationship between the wage markup and m, is the same as under autarky; see (16). We now define $\rho^T[m] := 1 + \frac{1}{k\gamma N[m]}$ as the goods price markup under free trade. This expression reflect the fact that firms now take into account foreign competitors, but it keeps the simplified form familiar from the autarky equilibrium as in (16). The reason is that with free trade as well as with autarky prices of imported and domestic varieties are fully symmetric, whence the price of any variety consumed is equal to the average price.

Total demand per variety remains unchanged, since the lower domestic demand is compensated by the larger number of countries:

$$q[m] = \frac{kL\theta[m]}{kN[m]p} = \frac{L\theta[m]}{N[m]p}. \tag{24}$$

The labor market clearing condition similarly remains unaffected. The equilibrium condition that determines m then follows as

$$g[m] = \rho^T[m]\psi[m]. \tag{25}$$

Note that $k = 1$ replicates the autarky equilibrium. The following proposition summarizes the comparison between autarky and free trade among $k > 1$ countries.

Proposition 1. *Opening up to free trade among k symmetric countries has the following effects relative to an autarky equilibrium: (i) There is exit of firms in each country, with an increase in the total number of varieties available to the consumer. (ii) There is a higher wage markup, coupled with a lower price markup, but goods prices are unambiguously lower. (iii) Each country's labor market suffers from a fall in the average matching quality, with lower average income. (iv) Each country enjoys a higher real income and higher aggregate welfare.*

Proof: A formal proof is relegated to appendix A.5.1.

The increase in variety (i) and the pro-competitive effect on the goods market (ii) are standard from international trade models with monopolistic competition and endogenous markups. The novel insight here relates to adverse labor market effects: A lower number of domestic firms lowers the degree of competition on labor markets, increasing the wage markup. But the pro-competitive effect dominates, leading to lower prices under free trade than under autarky (ii). In addition, the exit of firms makes it more difficult for workers to find firms matching well with their skills, causing a reduction in the productivity of the average worker (iii). However, the variety and pro-competitive effects more than compensate for this negative productivity effect, making the economy better off under free trade than under autarky (iv).

This positive welfare effect in this proposition reflects the excess entry property of the laissez faire equilibrium, whence an exit of firms entails a first order welfare gain. This holds true, whatever the cause of this exit. In the present scenario, this first order gain from $\mathrm{d}m > 0$ is driven by an opening up of borders, $\mathrm{d}k > 0$, which, in and of itself exerts a positive effect on household welfare through a larger number of product varieties available. Note, however, that workers are differently affected, depending on their location on the skill circle. While the maximum wage rate paid to a worker remains unchanged, the ideal workers in the trade equilibrium will be different from those of the autarky equilibrium. Moreover, given the increase in m, the lower bound of wages paid will be falling. Hence, some workers will even suffer a lower “nominal” wage rate. Statement (iv) of the proposition invokes the usual compensation argument in defining the aggregate welfare effect as the change in indirect utility of the worker who receives the average level of real income.

3.2 Costly trade and piecemeal trade liberalization

The superiority of free trade over autarky does not imply that a piecemeal liberalization in a world with costly trade is always beneficial. We stick to the symmetric case, but for simplicity reduce the number of countries to $k = 2$, using an asterisk to denote the foreign country. Suppose that firms face iceberg transport cost $\tau > 1$ for exports. A domestic firm that sells q_i units on the domestic market and q_i^* units on the export market then needs a labor input equal to $\alpha + q_i + \tau q_i^*$.[26] It solves the following profit maximization problem:

$$\max_{p_i, p_i^*} \{p_i q_i + p_i^* q_i^* - w_i(\alpha + q_i + \tau q_i^*)\} \tag{26}$$

$$\text{s.t.:} \qquad q_i = \delta_i \frac{Y}{p_i}, \quad q_i^* = \delta_i^* \frac{Y^*}{p_i^*} \quad \text{and} \tag{27}$$

$$\alpha + q_i + \tau q_i^* = L_i^{ES}[w_i, \bar{w}, m, L, H]$$

whereby

$$\delta_i = \frac{1}{N^T} + \gamma \left(\overline{\ln p} - \ln p_i \right) \qquad \text{and} \qquad \delta_i^* = \frac{1}{N^T} + \gamma \left(\overline{\ln p} - \ln p_i^* \right). \tag{28}$$

In these equations, $\overline{\ln p} = \frac{1}{N} \sum_{j=1}^{N} \ln p_j + \frac{1}{N^*} \sum_{j^*=1}^{N^*} \ln p_{j^*}$ denotes the log average price of competitors, where j and j^* index firm i's domestic and foreign competitors. Due to symmetry, this average log price is the same across markets.[27] Applying logic explained above and acknowledging symmetry, the first order conditions emerge as

$$p = \mathcal{W}[Z]\, \psi[m] \qquad \text{with } Z = \frac{1}{\psi[m]} \exp \left\{ 1 + \frac{1}{\gamma N^T[m]} + \overline{\ln p} \right\} \tag{29}$$

$$p^* = \mathcal{W}[Z^*]\, \psi[m]\tau \qquad \text{with } Z^* = \frac{1}{\psi[m]\tau} \exp \left\{ 1 + \frac{1}{\gamma N^T[m]} + \overline{\ln p} \right\}. \tag{30}$$

[26] Remember that we have scaled units such that the marginal production cost β equal unity.

[27] Due to symmetry, the expenditure functions are the same in both countries, but obviously, expenditure shares for domestic and imported goods are different. Expenditure shares are obtained by differentiation of the log expenditure function, i.e. $\delta_i := \frac{\partial \ln P}{\partial \ln p_i}$ and $\delta_i^* := \frac{\partial \ln P}{\partial \ln p_i^*}$.

The labor market clearing condition is

$$N[m]\left(\alpha + q[p,p^*,m] + \tau q^*[p,p^*,m]\right) = L\theta[m]. \tag{31}$$

In contrast to the autarky and the free trade case, the pricing conditions cannot be simplified further because individual firms' prices in (28) are not equal to average prices in the economy. The equilibrium skill reach of the representative firm, m, as well as domestic and export prices are determined by the system of equations (29), (30) and (31). It is the analogue to the free trade equilibrium condition (25) above.[28]

With Dixit-Stiglitz-type preferences households will always consume positive amounts of imported varieties, even for very large values of τ. In contrast, our preferences imply that a finite level of real trade cost might be prohibitive. We denote this prohibitive level of trade cost by $\bar{\tau}$, and it is determined implicitly by $\delta_i^* = 0$ in (28). Note that with $\delta_i^* = 0$ the price elasticity of demand for foreign goods becomes infinite; see (6). Note also that high values of γ imply low values of $\bar{\tau}$.

Proposition 2. *For two identical countries in a trading equilibrium, a decrease in trade cost τ within the non-prohibitive range, $\tau \in [1, \bar{\tau})$, has the following effects: (i) There is exit of firms in each country. (ii) The price of imported varieties falls, but the change in the price of domestically produced goods is ambiguous: it falls at low initial levels of τ, and it increases at high initial levels of τ. (iii) Aggregate welfare rises for sufficiently low initial levels of τ, and it falls for sufficiently high initial levels of τ.*

Proof: A formal proof is relegated to appendix A.5.2.

Part (iii) of this proposition may seem puzzling at first sight. According to standard results on piecemeal trade liberalization, in this fully symmetric economy a *uniform* proportional reduction of trade barriers across all varieties should be a welfare increasing "liberalization formula"; see Fukushima (1979). The key difference here arises from the labor market distortion. Liberalization involves two opposing effects. First, a lower price for imported varieties leads firms to lower their price markup on domestic goods; a straightforward pro-competitive effect that increases welfare. Note that this effect arises even at the prohibitive margin with $\tau = \bar{\tau}$ where

[28] As detailed in appendix A.5.2, for comparative statics it proves convenient to rewrite the system of equations in terms of the endogenous variables m, W, W^*.

no imports take place in the initial equilibrium.[29] At the same time, however, as firms in both countries ship more output to foreign markets, they use up more resources for transport, which bids up wage rates and causes firm exits in both countries.[30] Fewer domestic firms imply larger markups on the labor market as well as a lower average quality of matches between firms and worker skills. The magnitude of this effect clearly depends on the initial level of the trade cost; it is strongest at $\tau = \bar{\tau}$ and disappears for $\tau = 1$. The proof in the appendix demonstrates that for $\tau = \bar{\tau}$ initially, the adverse labor market effect of a marginal reduction of τ dominates, not just in terms of higher prices for domestic varieties, but also in terms of welfare so that liberalization is welfare reducing. Since we can also demonstrate that free trade, $\tau = 1$, is better than autarky, $\tau = \bar{\tau}$, there is a threshold value $\tilde{\tau}$, with $1 < \tilde{\tau} < \bar{\tau}$, such that piecemeal liberalization starting from $\tau < \tilde{\tau}$ is unambiguously welfare increasing. Proposition 1 implies that $\tilde{\tau} > 1$.

Invoking costless compensation, w use average income to evaluate aggregate welfare effects in an economy where heterogeneous workers are affected differently. Using the indirect utility function we see that welfare is affected by changes in prices of domestic and imported goods as well as by the change in m, which affects both average income and the number of available varieties. The change in welfare can be expressed as

$$\widehat{V} = \left(\frac{\partial \ln \theta}{\partial \ln m} - \frac{\partial \ln P}{\partial \ln m} \right) \widehat{m} - N\delta\hat{p} - N\delta^{*}\hat{p}^{*}. \tag{32}$$

The positive change in m as induced by a decrease in trade cost affects welfare negatively through a decrease in average income $\theta_m[m] < 0$ and a decrease in the number of firms, which implies an increase in the ideal price index $P_m[m] > 0$. The effects of changes in prices of domestic and imported goods are weighted by the expenditure share that consumers devote to the respective goods. For high initial levels of the trade cost the expenditure share for imported goods is small, whence consumers hardly benefit from the decrease in the price of imports, while being much affected by the the change in the price of domestic goods, which is positive for high initial levels of the trade cost. Hence, for high initial of τ the overall effect of a decrease in trade cost on welfare is negative. In contrast, for low initial values of the trade cost, the negative

[29] For $\tau = \bar{\tau}$ the trading equilibrium is quantitatively identical to the autarky equilibrium considered in section 2.4. This can be shown by inserting the implicit solution for $\bar{\tau}$, obtained by setting $\delta^{*} = 0$, into the pricing condition (29). Yet, the disciplinary effect of a decrease in import prices works through $\overline{\ln p}$ in equation (29), even if $\delta^{*} = 0$.

[30] This is the mechanism underlying the well-known home market effect for *asymmetric* countries, first noted by Krugman (1980). Of course, the home market effect as such does not arise here, since countries are assumed symmetric.

effect through a decrease in the number of firms becomes smaller, ultimately approaching zero. Furthermore, the higher the expenditure share for imported goods, the more consumers benefit from lower import prices, and the more important the competitive effect on domestic prices through the product market, ultimately overcompensating the anticompetitive effect through the labor market. Hence, we find a U-shaped relationship between welfare and the trade cost level.

4 Migration

So far, the effects of product market integration have been considered under the assumption that workers are immobile across countries. Allowing for migration, some workers in both countries will find their skills matching better with a firm in the foreign country. This constitutes an incentive for two-way migration. In this section, we show that trade cum migration delivers higher welfare than trade in goods alone. Furthermore, we show that under a slightly stronger assumption about the curvature of $f[d]$, piecemeal integration of labor markets, unlike piecemeal trade liberalization, unambiguously increases welfare in both countries. Our model highlights two welfare increasing effects of migration: First, integration of labor markets reduces monopsony power, as domestic firms now compete for workers, not only with other domestic firms, but also with foreign firms. This holds true even if the cost of migration is prohibitively large. Second, migration entails efficiency gains by improving the average matching quality, as workers in both countries are now able to find better skill matches for employment. Even though the number of available varieties might fall, compared to the free trade equilibrium, the efficiency gains and the pro-competitive gains on the labor market are always dominating, leading to a positive welfare effect.

4.1 Modeling migration

For simplicity, we consider the case of two symmetric countries, which implies the same number of firms in both countries, as well as equal prices and wages. This simplification allows us to focus on the part of migration that is related to the idea of skill mismatch. We deliberately ignore differences in average wages or in the cost of living that would clearly constitute migration incentives as well.[31] We model the cost of migration as reducing the productivity of a worker to a fraction $1-\lambda$ if moving to the other country. A domestic worker working for a domestic firm at distance d, delivering $f[d]$ efficiency units, thus delivers only $f[d](1-\lambda)$ efficiency units when working for a foreign firm at the same skill distance d.[32]

We analyze international migration as an entry/location/pricing game with Bertrand behavior analogous to sections 2 and 3, occurring simultaneously in both countries, whereby each firm takes into account the possibility of hiring workers from the other country. It is relatively easy to show that any equilibrium of such a game with positive migration cost involves an alternating pattern of firm locations on the skill circle. By alternating pattern, we mean any one firm facing two neighboring firms from the other country. To see this, consider a case where a domestic firm finds its closest competitor to be another domestic firm. Figure 3 depicts two such firms located at distance m on the skill-circle, with optimal skill-types equal to positions $\bar{s}_0$ and $\bar{s}_0+m$. They would obtain a firm-specific input of efficiency units equal to $f(\bar{s}_0+d)$ and $f(\bar{s}_0+m-d)$ from employing a domestic worker with skill type $\bar{s}_0+d$. The cut-off for equal productivity of domestic workers in both firms would be $\bar{s}_0+m/2$. For employment of foreign workers with the same positions on the skill circle the productivity lines would be scaled down by $1-\lambda$, symmetrically for both firms (dashed lines), leading to the same cut-off position for equal productivity of a foreign worker in both firms. With both firms setting equal wage rates per efficiency unit, domestic as well as foreign workers located to the left (right) of $\bar{s}_0+m/2$ would sort themselves into employment with the firm located at $\bar{s}_0$ (at $\bar{s}_0+m$).

Now compare this with a case where the domestic firm at point $\bar{s}_0$ finds a foreign firm at $\bar{s}_0+m$ as its closest competitor. Foreign workers located in the interval $(\bar{s}_0+d^m, \bar{s}_0+m/2)$ would now prefer to work for the foreign firm as they could save on migration cost, which implies an output loss to the domestic firm equal to the dark-shaded area. Conversely, domestic workers located in the interval $(\bar{s}_0+m/2, \bar{s}_0+d^n)$ would now prefer to work for the domestic firm, where they generate an additional output equal to the light-shaded area. It is apparent from the figure that, other things equal, the case with two neighboring firms being from different

[31] Labor mobility and free entry imply that there is also the possibility of an agglomeration equilibrium, where all workers and firms work and produce in the same country. This is ruled out, if trade cost is sufficiently low, compared to the cost of migration. Throughout this section, we assume free trade, hence our results are not impaired by instability. In the case of zero trade and zero migration cost, the equilibrium outcomes in the dispersed and the agglomeration equilibrium are the same in terms of prices and welfare.

[32] The proportionality assumption is convenient for modeling, yet it is not crucial. A general characterization of the specifications generating the results derived in this section is found in appendix A.7.

countries involves a larger output for both firms.[33] What we have said for neighboring firms locating to the right of $\bar{s}_0$ holds true, mutatis mutandis, for firms locating to the left of $\bar{s}_0$. This proves that for symmetric countries any equilibrium with positive migration cost must involve an alternating pattern of firm locations on the skill circle.

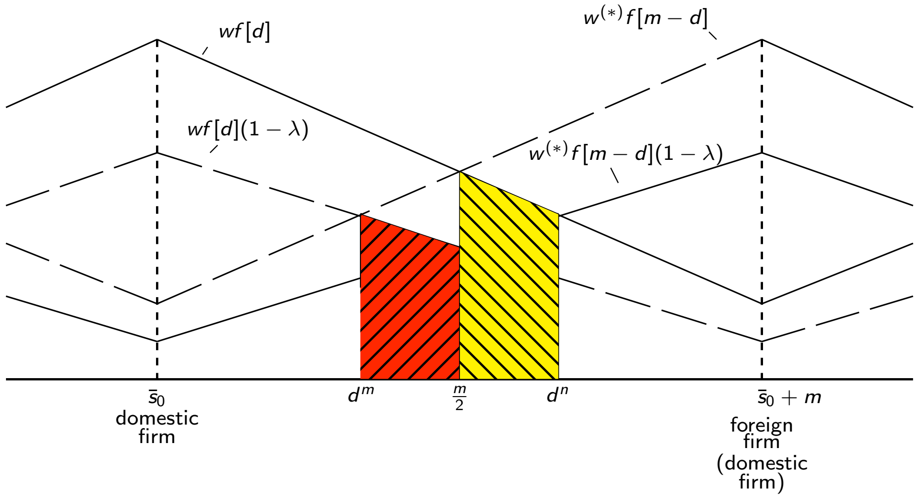

Figure 3: The "supply bonus" from locating next to a foreign neighbor

4.2 Worker sorting and labor supply with integrated labor markets

We continue using $2m$ to denote the skill distance between two firms located in the same country.[34] In the alternating equilibrium the firm's direct competitor on the labor market, which is located in the foreign country, is then found at distance m in the skill space. The sorting cutoffs, i.e., the maximum distances of native workers d^n and migrant workers d^m from their firms, are derived as follows. For a domestic firm i, taking the foreign wage w^* as given, the cutoff for native workers, $d_i^n = d_i^n[w_i, \bar{w}^*, m, \lambda]$, is determined by

$$w_i f[d_i^n] = \bar{w}^* f[m - d_i^n](1 - \lambda). \tag{33}$$

[33] The effect works through migration cost savings. A formal proof is provided in appendix A.6.1.

[34] Note that without migration $2m$ also measures the distance to the nearest competitor.

Analogously, the cutoff for migrant workers, $d_i^m = d_i^m[w_i, \bar{w}^*, m, \lambda]$, is determined by

$$w_i f[d_i^m] = \bar{w}^* f[m - d_i^m] \frac{1}{1-\lambda}. \tag{34}$$

As the migration cost falls, the cutoffs converge. At $\lambda = 0$ they coincide at $m/2$.

Under symmetry, the employment and migration pattern will be as follows: The domestic firm employs domestic workers with skill-types in the interval $(\bar{s}_0 - d^n, \bar{s}_0 + d^n)$, and foreign workers (migrants) located in the interval $(\bar{s}_0 - d^m, \bar{s}_0 + d^m)$, while the foreign firm located at $\bar{s}_0 + m$ will employ foreign workers located in the interval $(\bar{s}_0 + m - d^n, \bar{s}_0 + m + d^n)$ and domestic workers (migrants) with skill types in the interval $(\bar{s}_0 + m - d^m, \bar{s}_0 + m + d^m)$. Notice that $d^n + d^m = m$.

The supply of efficiency units as a function of the firm's wage now emerges as

$$L_i^{ES,M}[w_i, \bar{w}^*, m, \lambda, L, H] = \frac{L}{H} \int_0^{d_i^n[w_i, \bar{w}^*, m, \lambda]} f[d]\mathrm{d}d + \frac{L}{H} \int_0^{d_i^m[w_i, \bar{w}^*, m, \lambda]} f[d](1-\lambda)\mathrm{d}d \tag{35}$$

where $d_i^n[w_i, \bar{w}^*, m, \lambda]$ and $d_i^m[w_i, \bar{w}^*, m, \lambda]$ are given by (33), (34), respectively. Let $d^n = d^n[m, \lambda]$ and $d^m = d^m[m, \lambda] := m - d^n[m, \lambda]$ denote the cutoffs in the symmetric equilibrium. These two variables measure the skill reach of a representative firm for domestic and foreign workers, respectively. With $\lambda > 0$, we have $d^m < d^n$. As before, m may be interpreted as a mismatch indicator, but the average distance between worker skills and a firm's ideal type across employment of domestic and foreign workers is now equal to $m/2$, whereas without migration it was equal to m. In a symmetric equilibrium, average productivity then emerges as

$$\theta^M[m, \lambda] := \frac{1}{m} \left(\int_0^{d^n} f[d]\mathrm{d}d + \int_0^{d^m} f[d](1-\lambda)\mathrm{d}d \right). \tag{36}$$

By complete analogy to (14), the perceived elasticity of effective labor supply, evaluated at the symmetric equilibrium, can be derived as[35]

$$\eta^M[m, \lambda] := \left. \frac{\frac{\partial L_i^{ES,M}}{\partial w_i}}{\frac{L_i^{ES,M}}{w_i}} \right|_{w_i = \bar{w}^*} = \frac{2 f[d^n]^2}{f'[d^n] + (1-\lambda) f'[m - d^n]} \frac{-1}{\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda) \int_0^{d^m} f[d]\mathrm{d}d}. \tag{37}$$

Note that the labor supply function is subject to the constraint $d^m[m, \lambda] \geq 0$, which ensures that

[35] For details of the derivation see appendix A.6.2

both cutoffs lie in between the positions of the domestic and the foreign firm. This condition is equivalent to the condition that the migration cost λ is not prohibitive.[36] As the migration cost approaches the prohibitive level, the supply of efficiency units of labor becomes equal to the supply under autarky. This is readily verified by inserting $d^n = m$ and $d^m = 0$ into (35).

Interestingly, even if migration cost is prohibitive, firm behavior is influenced by the mere potential of migration through the perceived elasticity of labor supply.[37] The possibility of attracting migrants by setting higher wages and thus increasing the supply of efficiency units implies that firms perceive a higher elasticity of supply, even if they do not employ any migrant in equilibrium. Let $\bar{\lambda}$ denote the prohibitive level of migration cost, determined by setting $d^m[m, \lambda] = 0$. The perceived wage elasticity of labor supply evaluated at $\bar{\lambda}$ is given by

$$\eta^M[m, \bar{\lambda}] = \frac{2f[m]^2}{f'[m] + (1 - \bar{\lambda})f'[0]} \cdot \frac{-1}{F[m]}. \tag{38}$$

Note that concavity of $f[d]$ is sufficient to ensure that $\eta^M[m, \bar{\lambda}]$ is larger than the elasticity of supply under autarky as given in (14). Furthermore, it can be shown that $\eta^M[m, \lambda]$ falls as λ increases, provided that $f'''[d]$ is not too large. In what follows, we assume that this condition holds.[38] By analogy to (16), we now use $\psi^M[m, \lambda] := \left(\eta^M[m, \lambda] + 1\right) / \, \eta^M[m, \lambda]$ to denote the wage distortion under migration. For a given level of m, the magnitude of this distortion is unambiguously lower with migration and $\lambda \in [0, \bar{\lambda}]$ than without.

In addition to the wage distortion, migration also affects the average quality of skill matches between workers and firms. It is obvious that for prohibitively high migration cost, $\lambda = \bar{\lambda}$, the average matching quality, as given in equation (36), is the same as under autarky, as given in (17): $\theta^M[m, \bar{\lambda}] = \theta[m]$. Moreover, as we prove in the appendix, $\theta^M_\lambda < 0$. In other words, the matching quality unambiguously increases as λ falls, reaching $\theta^M[m, 0] = \theta[m/2]$ for frictionless migration where $\lambda = 0$. It is instructive to see how effective labor supply to a representative firm is affected by the cost of migration. Under frictionless migration, $\lambda = 0$, labor supply as

[36] Otherwise, if migration cost is too large relative to firm size, firms cannot attract any migrants in the first place and the supply curve looks different since they then compete again only with firms from the same country.

[37] We thank Vitor Trindade for pointing this out to us.

[38] The reasoning behind this condition is as follows: A higher λ leads firms to increase the share of migrants employed by shifting d^n outwards and d^m inwards. If the curvature of $f[d]$ falls (in absolute terms) as the cutoffs move to the right, an increase in λ helps firms to avoid competition by employing more native workers in the range where the curvature of $f[d]$ is lower and fewer migrants in the range where the curvature of $f[d]$ is strong. We rule this out by assuming that the curvature does not decrease too much (in absolute terms) as the cutoff moves to the right.

given in (35) emerges as

$$L^{ES,M} = 2\frac{L}{H}\int_0^{\frac{m}{2}} f[d]\mathrm{d}d = \frac{2L}{N^M}\theta^M[m,0] = \frac{L}{N}\theta\left[m/2\right]. \tag{39}$$

Note that $N^M = \frac{2H}{m} = 2N$, where N is the number of firms in each country. Comparing this to the autarky case, both the number of firms and workers are doubled. However, we know from above that for $\lambda < \bar{\lambda}$ we have $\theta^M > \theta$. Hence, with migration firms face a larger supply of efficiency units of labor than under autarky. The reason is that while employing the same number of workers as under autarky each firm now finds workers with skills closer to its optimal type. Importantly, all of this is conditional upon a given level of m, which is determined by the firm entry condition. As we shall see below, equilibrium adjustment of the number of firms after opening up to migration, driven by a lower wage markup, might bring about firm exit which has a countervailing, negative effect on average productivity.

4.3 The trade cum migration equilibrium

We complete the description of a trade cum migration equilibrium by a free entry (zero profit) condition that determines the number of firms or, equivalently, the distance between a neighboring domestic and foreign firm on the skill circle, which is now equal to m. We look at the case of free trade. As in (25), we formulate this condition as stating that the double markup is equal to the inverse of the degree of economies of scale:

$$g^M[m,\lambda] = \rho^T[m]\psi^M[m,\lambda]. \tag{40}$$

In this equation, $\rho^T[m]$ denotes the free trade price markup over perceived marginal cost obtaining in a free trade equilibrium without migration. Under free trade, this markup simplifies to $1 + 1/(\gamma N^M)$, where N^M is the number of firms world-wide; see equation (23). Unlike the wage markup, the price markup is not affected by allowing for migration. The term $\psi^M[m,\lambda]$ denotes the wage markup in a migration equilibrium, as introduced above. The term $g^M[m,\lambda]$ on the left measures the degree of scale economies, taking into account the labor market clearing condition, which now reads as $\alpha + q = (m/H)L\theta^M[m,\lambda]$, as well as goods market clearing, which requires $q = L\theta^M[m,\lambda]/(pN)$. This measure then reads as

$$g^M[m,\lambda] := \frac{L\theta^M[m,\lambda]}{L\theta^M[m,\lambda] - \alpha H/m}. \tag{41}$$

In order to understand the effects of labor market integration, we now proceed in two steps. We first look at the case where migration is allowed in principle, but where the cost of migration is prohibitively large, $\lambda = \bar{\lambda}$, and compare this case with the equilibrium under national labor markets. In the second step we then look at the effects of successively lowering the cost of migration below the prohibitive level, eventually leading to perfect labor market integration where $\lambda = 0$.

Proposition 3. *Compared to a free trade equilibrium with national labor markets, a trade cum migration equilibrium with two symmetric countries and a prohibitively high level of the cost of migration involves a lower number of firms, whereas the level of welfare is unambiguously higher in both countries.*

Proof: The analytical details of the proof are relegated to appendix A.6.3.

A key point to understand this proposition is that the excess entry property of the autarky equilibrium demonstrated in section 2.5 is inherited by the migration equilibrium for *any* $\lambda \in [0, \bar{\lambda}]$. While the productivity distortion is not affected as long as no one migrates, the wage markup is affected because firms perceive a larger elasticity of labor supply. By lowering the wage markup, opening up labor markets to migration leads to firm exit, even if the cost of migration is prohibitively high. And given that the free trade equilibrium involves excessive firm entry, this entails a positive welfare effect. With a lower wage markup distortion relative to the productivity distortion, the equilibrium is now closer to the social optimum. Referring to our discussion subsequent to propositions 1 and 2, we repeat that individual households are affected differently, due to skill heterogeneity. Speaking of an aggregate welfare effect implies the existence of a costless (lump-sum) redistribution mechanism.

Proposition 4. *In a trade cum migration equilibrium with two symmetric countries, piecemeal integration of labor markets through a marginal reduction in the cost of migration has an ambiguous effect on the number of firms. However, it unambiguously leads to lower prices and an increase in welfare in both countries, irrespective of the initial level of migration cost $\lambda \in [0, \bar{\lambda}]$.*

Proof: The analytical details of the proof are relegated to appendix A.6.4.

The intuition for this proposition is best grasped from figure 4, which depicts the schedules $g^M[m, \lambda]$ and $\rho^T[m]\psi^M[m, \lambda]$, identifying the equilibrium value of m at the intersection, in line with the zero profit equilibrium condition (41). The vertical axis of figure 4 may be interpreted as measuring goods prices. Remember that $g^M[m, \lambda]$ measures the inverse degree of scale economies, which is equivalent to the markup required for zero profits. An increase in m makes firms larger, but it also lowers the productivity of the average worker. The appendix shows that the size effect always dominates, whence the g^M-line is downward-sloping. The $\rho^T[m]\psi^M[m, \lambda]$-line depicts the double markup, reflecting monopoly power on the goods market and monopsony power on the labor market, respectively. This line is unambiguously upward-sloping, as a lower number of firms (higher m) reduces both the perceived price elasticity of goods demand as well the perceived labor supply elasticity with respect to the wage rate. We know from proposition 3 above that the intersection point for $\lambda = \bar{\lambda}$ involves a lower value of m than in the free trade equilibrium with national labor markets, which is determined by $g[m] = \rho^T[m]\psi[m]$.[39]

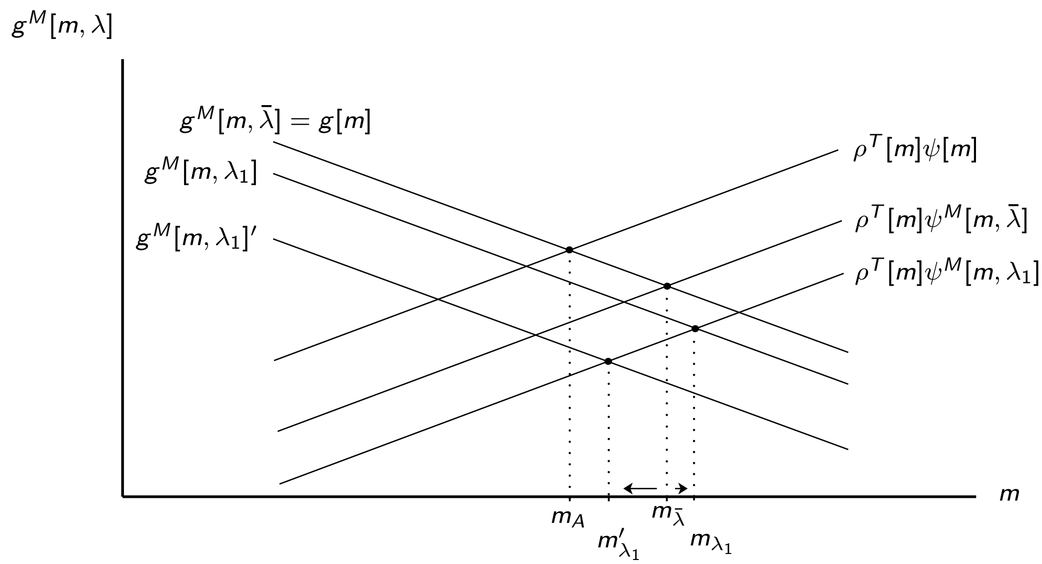

Figure 4: Comparative statics of the skill reach m

Now consider a reduction in λ from $\bar{\lambda}$ to $\lambda_1 \in [0, \bar{\lambda})$. For a notionally unchanged value of m, this improves the productivity of the average worker through a higher inframarginal surplus on

[39] Moving from an equilibrium with national labor markets to a trade cum migration equilibrium with $\lambda = \bar{\lambda}$ leaves g unaffected, $g[m] = g^M[m, \bar{\lambda}]$, while reducing the wage markup, $\psi[m] > \psi^M[m, \bar{\lambda}]$.

migrant labor as well as through a resorting of workers from native employment into migration.[40] This means that the g^M-line is shifted down by a reduction in λ. As regards the markup schedule $\rho^T[m]\psi^M[m,\lambda]$, we have shown above that the perceived elasticity of labor supply increases with a lower cost of migration, meaning that for a notionally unchanged m firms charge a lower wage markup $\psi^M[m,\lambda]$. Thus, the markup schedule shifts down as well, rendering an ambiguous effect on m. In the figure, the case $g^M[m,\lambda_1]$ ($g^M[m,\lambda_1]'$) depicts a relatively weak (strong) shift in the g^M-line, leading to an increase (a decrease) in m. However, the equilibrium unambiguously moves down on the vertical axis, which implies lower goods prices.

The welfare effect is determined by the change in real income and the number of varieties. Real income is given by $\theta^M[m,\lambda]/\,p[m]$, where average "nominal" income is measured by $\theta^M[m,\lambda]$, the productivity of the average worker. Invoking the indirect utility function, the welfare effect of our scenario may be described as

$$\hat{V} = \frac{\partial \ln\left[\theta^M/p\right]}{\partial \lambda} \cdot \mathrm{d}\lambda + \frac{\partial \ln\left[\theta^M/p\right]}{\partial m} \cdot \mathrm{d}m - \frac{1}{4\gamma H} \cdot \mathrm{d}m \tag{42}$$

The first term describes the direct effect of a lower migration cost, $\mathrm{d}\lambda < 0$, on real income. From the above we know that this term is unambiguously positive. The remaining terms involving $\mathrm{d}m$ are ambiguous in their entirety, because $\mathrm{d}m$ as caused by $\mathrm{d}\lambda < 0$ is ambiguous. However, we know from the above discussion of the distortions present in this economy that the autarky equilibrium involves excess firm entry, and from the proof of proposition 3 we know that any trade cum migration equilibrium inherits this excess entry property. Therefore, the positive real income effect of firm exit in the second term must dominate the negative variety effect in the third term. In other words, if the equilibrium adjustment depicted in figure 4 leads to $\mathrm{d}m > 0$, then the overall effect of $\mathrm{d}\lambda < 0$ on welfare is positive. If $\mathrm{d}m < 0$, then the welfare effect is less straightforward. While the final term of this expression is then unambiguously positive, the first two terms seem ambiguous. However, we show in the appendix that the first two terms of (42) are unambiguously positive for any initial $\lambda \in [0, \bar{\lambda}]$, if we insert $\mathrm{d}m = (\partial m/\partial \lambda) \cdot \mathrm{d}\lambda$.

[40] Note that for a constant m the average skill distance between workers and their firm's ideal type remains constant, equal to $m/2$. The productivity gain arises from savings in migration cost.

5 Conclusion

In this paper, we propose an important qualification to the common narrative of of variety-based gains from trade. Traditional models of monopolistic competition stress the importance of a large resource base for a large degree of product differentiation, if production is subject to a non-convex technology. By opening up to trade, even small countries may enjoy the benefits of a large resource base. Domestic firms may be driven out of the market, but this has no adverse effect. If anything, it increases the average productivity level through a positive selection effect.

This view neglects an important fact of modern manufacturing: Product differentiation relies on the availability of differentiated inputs, including non-traded inputs like labor. If producing a specific variety of a good requires a specific bundle of skills, then the skill-diversity of the labor force, rather than its size, determines the degree of product differentiation supplied by the market. In this paper, we have shown that trade is a somewhat less benign force in an environment where product differentiation is based on worker heterogeneity than portrayed in conventional models of monopolistic competition. In particular, trade-induced firm exit worsens the average quality of matches between the type of skills that workers bring to their firms and the specific skill requirements of the goods produced by these firms. In addition, product differentiation implies that firms have monopsony power in the labor market, whence trade-induced exit of firms increases the resulting distortion between the marginal productivity of labor and the wage rate. This works against the conventional pro-competitive effect of trade on the goods markets where trade lowers the markup between marginal cost and prices. Labor market integration gives rise to a migration incentive, whereby firms engage in cross-border hiring even under complete symmetry between countries. Migration essentially has effects that are opposite to those of trade.

We have developed a model which allows us to rigorously pin down these effects and to weigh them against the effects familiar from conventional models of monopolistic competition. In our model product differentiation is rooted in preferences represented by a translog expenditure function. When entering the market, firms decide upon which type of good to produce, based on a circular representation of skill heterogeneity among the work force, where each worker has the potential to serve as an "ideal" worker for a specific type of good. A non-convex technology implies a finite number of firms. A worker's supply of efficiency units is inversely related to the distance between her skill-position on the circle and the ideal skill position of the firm she works for. Having positioned themselves on the circle upon entry, firms engage in Bertrand competition on goods and labor markets, setting a double markup.

Using this model, we have explored both trade and migration scenarios. Comparing free trade with autarky in a symmetric many-country-world, we find that the variety and pro-competitive effects on goods markets unambiguously dominate the adverse effects from a lower average quality of matches between firms and workers and from higher markups on the labor market. Looking at piecemeal trade liberalization between two symmetric countries, we find an ambiguity: If liberalization takes place from a high initial level of trade cost, then it causes a lowering of aggregate welfare, whereas it increases aggregate welfare, if the initial level of trade cost is already below a certain threshold.

Starting from a free trade equilibrium in a symmetric two-country-world, integrating labor markets leads to two-way migration. Firms and workers in both countries face an incentive for cross-border hiring, even though the initial equilibrium features international wage equalization. Thus, our view of product differentiation based on worker heterogeneity generates a novel force of migration, contributing to an improved understanding of two-way migration, which looms large in the data but has so far lacked convincing explanation in standard models of migration. Interestingly, potential migration exerts a positive welfare effect on both countries, even if the migration cost is prohibitively large. Contrary to piecemeal trade liberalization, a piecemeal reduction in the cost of migration is unambiguously welfare increasing. The reason is that it improves the quality of matches while at the same time lowering firms' monopsony power on labor markets. From the simple fact that trade and migration have opposite effects it also follows that trade and migration are complements, rather than substitutes. The model clearly advocates opening up labor markets simultaneously with trade liberalization.

References

Amiti, M. and Pissarides, C. A. (2005). Trade and Industrial Location with Heterogeneous Labor. Journal of International Economics, 67(2):392–412.

Antràs, P. (2003). Firms, Contracts, and Trade Structure. The Quarterly Journal of Economics, 118(4):1375–1418.

Antràs, P. and Helpman, E. (2004). Global Sourcing. Journal of Political Economy, 112(3):552–80.

Arkolakis, C., Costinot, A., and Rodríguez-Claire, A. (2010). Gains From Trade under Monopolistic Competition: A Simple Example with Translog Expenditure Functions and Pareto Distributions of Firm-Level Productivity. mimeo.

Arkolakis, C., Costinot, A., and Rodriguez-Clare, A. (2012). New Trade Models, Same Old Gains? American Economic Review, 102 (1):94–130.

Behrens, K. and Murata, Y. (2007). General Equilibrium Models of Monopolistic Competition: A New Approach. Journal of Economic Theory, 136(1):776–787.

Behrens, K. and Murata, Y. (2012). Trade, Competition, and Efficiency. Journal of International Economics, 87(1):1–17.

Beine, M., Docquier, F., and Özden, C. (2011). Diasporas. Journal of Development Economics, 95(1):30–41.

Bergin, P. R. and Feenstra, R. C. (2000). Staggered Price Setting and Endogenous Persistence. Journal of Monetary Economics, 45(3):657–680.

Berninghaus, S. and Seifert-Vogt, H. G. (1991). A Temporary Equilibrium Model for International Migration. Journal of Population Economics, 4(1):pp. 13–36.

Bhaskar, V. and To, T. (1999). Minimum Wages for Ronald McDonald Monopsonies: A Theory of Monopsonistic Competition. The Economic Journal, 109(455):190–203.

Bhaskar, V. and To, T. (2003). Oligopsony and the Distribution of Wages. European Economic Review, 47(2):371 – 399.

Bilbiie, F. O., Ghironi, F., and Melitz, M. J. (2008). Monopoly Power and Endogenous Product Variety: Distortions and Remedies. NBER Working Paper 14383.

Broda, C. and Weinstein, D. E. (2006). Globalization and the Gains from Variety. The Quarterly Journal of Economics, 121(2)(2):541–585.

Davidson, C., Heyman, F., Matusz, S., Sjöholm, F., and Zhu, S. C. (2011). Globalization and Imperfect Labor Market Sorting. Working Paper 2011:25, Lund University.

Davidson, C., Heyman, F., Matusz, S., Sjöholm, F., and Zhu, S. C. (2012). Liberalized Trade and Worker-Firm Matching. American Economic Review, 102(3):429–34.

Davidson, C., Matusz, S. J., and Shevchenko, A. (2008). Globalization and Firm Level Adjustment with Imperfect Labor Markets. Journal of International Economics, 75(2):295–309.

Diewert, W. E. (1974). Applications of Duality Theory. In Intriligator, M. and Kendrick, D., editors, Frontiers of Quantitative Economics, volume 2, pages 106–171. North-Holland.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. American Economic Review, 67(3):297–308.

Docquier, F. and Marfouk, A. (2006). International Migration by Educational Attainment (1990-2000) - Release 1.1. In Özden, C. and Schiff, M., editors, International Migration, Remittances and Development, volume 1, chapter 25. Palgrave Macmillan, New York.

Eckel, C. (2009a). Endogenous Sunk Costs, Flexible Manufacturing and the Productivity Effects of International Trade. Scandinavian Journal of Economics, 111(2):369–386.

Eckel, C. (2009b). International Trade, Flexible Manufacturing, and Outsourcing. Canadian Journal of Economics, 42(4):1449–1472.

Economides, N. (1989). Symmetric Equilibrium Existence and Optimality in Differentiated Product Markets. Journal of Economic Theory, 47(1):178–194.

Ethier, W. J. (1982). National and International Returns to Scale in the Modern Theory of International Trade. American Economic Review, 72 (1)(3):389–405.

Fan, C. S. and Stark, O. (2011). A Theory of Migration as a Response to Occupational Stigma. International Economic Review, 52(2):549–571.

Feenstra, R. C. (1994). New Product Varieties and the Measurement of International Prices. The American Economic Review, 84(1):157–177.

Feenstra, R. C. (2003). A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity. Economics Letters, 78(1):79–86.

Feenstra, R. C. (2010). Product Variety and the Gains from International Trade. Cambridge, Mass: MIT-Press.

Feenstra, R. C. and Kee, H. L. (2008). Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity. Journal of International Economics, 74:500–518.

Feenstra, R. C. and Weinstein, D. E. (2010). Globalization, Markups, and the U.S. Price Level. NBER Working Paper 15749.

Felbermayr, G., Grossmann, V., and Kohler, W. (2012). Migration, International Trade and Capital Formation: Cause or Effect? IZA Discussion Paper 6975.

Fiorillo, F., Santacroce, S., and Staffolani, S. (2000). Monopsonistic Competition for the "Best" Workers. Labour Economics, 7(3):313–334.

Fukushima, T. (1979). Tariff Structure, Nontraded Goods and Theory of Piecemeal Policy Recommendations. International Economic Review, 20(2):427–35.

Galor, O. (1986). Time Preference and International Labor Migration. Journal of Economic Theory, 38(1):1 – 20.

Gaumont, D. and Mesnard, A. (2000). Altruism and International Labour Migration. Journal of Population Economics, 13(1):pp. 113–126.

Grossman, G. M. and Helpman, E. (2005). Outsourcing in a Global Economy. Review of Economic Studies, 72(1):135–159.

Hamilton, J., Thisse, J.-F., and Zenou, Y. (2000). Wage Competition with Heterogeneous Workers and Firms. Journal of Labor Economics, 18(3):453–72.

Helpman, E. (1981). International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition : A Chamberlin-Heckscher-Ohlin Approach. Journal of International Economics, 11(3):305–340.

Helsley, R. W. and Strange, W. C. (1990). Matching and Agglomeration Economies in a System of Cities. Regional Science and Urban Economics, 20(2):189–212.

Kim, S. (1987). A Product Differentiation Model with Outside Goods and Price Bargaining. Economics Letters, 24(4):305 – 309.

Kim, S. (1989). Labor Specialization and the Extent of the Market. The Journal of Political Economy, 97(3):692–705.

Kreickemeier, U. and Wrona, J. (2013). Twoway Migration between Similar Countries. Working paper, University of Tuebingen.

Krugman, P. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. American Economic Review, 70(5):950–59.

Krugman, P. R. (1979). Increasing Returns, Monopolistic Competition, and International Trade. Journal of International Economics, 9(4):469–479.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. The Journal of Political Economy, 74(2):132–157.

Mandelbrot, B. (1962). Paretian Distributions and Income Maximization. The Quarterly Journal of Economics, 76(1):57–85.

Mankiw, N. G. and Whinston, M. D. (1986). Free Entry and Social Inefficiency. RAND Journal of Economics, 17(1):48–58.

Marimon, R. and Zilibotti, F. (1999). Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. Economic Journal, 109(455):266–91.

Markusen, J. R. (1983). Factor Movements and Commodity Trade as Complements. Journal of International Economics, 14(3-4):341–356.

Mayda, A. M. (2010). International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows. Journal of Population Economics, 23(4):1249–1274.

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

Melitz, M. J. and Ottaviano, G. I. P. (2008). Market Size, Trade, and Productivity. Review of Economic Studies, 75(1):295–316.

Ohnsorge, F. and Trefler, D. (2007). Sorting It Out: International Trade and Protection with Heterogeneous Workers. Journal of Political Economy, 115(5):868–892.

Ortega, F. and Peri, G. (2012). The Effect of Trade and Migration on Income. NBER Working Paper 18193.

Ravenstein, E. G. (1885). The Laws of Migration. Journal of the Statistical Society of London, 48(2):167–235.

Ródriguez-López, J. A. (2011). Prices and Exchange Rates: A Theory of Disconnect. The Review of Economic Studies, 78(3):1135–1177.

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. The Bell Journal of Economics, 10(1):141–156.

Schmitt, N. and Soubeyran, A. (2006). A Simple Model of Brain Circulation. Journal of International Economics, 69(2):296–309.

Tabuchi, T. and Thisse, J.-F. (2002). Taste Heterogeneity, Labor Mobility and Economic Geography. Journal of Development Economics, 69(1):155 – 177.

Thisse, J.-F. and Zenou, Y. (2000). Skill Mismatch and Unemployment. Economics Letters, 69(3):415 – 420.

Vickrey, W. (1964). Microstatics. Harcourt, Brace and World, New York.

Vickrey, W. S., Anderson, S. P., and Braid, R. M. (1999). Spatial Competition, Monopolistic Competition, and Optimum Product Diversity. International Journal of Industrial Organization, 17(7):953–963.

Welch, F. (1969). Linear Synthesis of Skill Distribution. The Journal of Human Resources, 4(3):311–327.

Appendix

A.1. Pricing in the Bertrand equilibrium

Firms maximize profits π as defined in (9) w.r.t. to p_i, subject to goods demand (5) and labor supply (8). The first order condition is

$$\frac{\partial \pi_i}{\partial p_i} = \frac{\partial (p_i q_i)}{\partial p_i} - \frac{\partial w_i}{\partial L_i^E} \frac{\partial L_i^E}{\partial q_i} \frac{\partial q_i}{\partial p_i} (\alpha + \beta q_i) + w_i \beta \frac{\partial q_i}{\partial p_i} \stackrel{!}{=} 0$$

Setting $L_i^{ES}[w_i, \bar{w}, m] = \alpha + \beta q_i$ and using the definitions of η_i for labor supply and ε_i for goods demand, we arrive at the pricing equation (10). In view of (6) and (4), and observing the definition of perceived marginal cost, $\tilde{w}_i := [(\eta_i + 1)/\eta_i] w_i \beta$, this equation may be written as

$$\frac{p_i}{\tilde{w}_i} + \ln p_i = 1 + \frac{1}{\gamma N} + \overline{\ln p} \tag{A.1}$$

The left-hand side is an implicit function of the profit maximizing price p_i. Following Arkolakis et al. (2010), we use the *Lambert function* in order to derive an explicit solution. In general terms, this function is defined as the solution $\mathcal{W}[z]$ to $xe^x = z$ or, equivalently, to $\ln x + x = \ln z$. Rewriting (A.1) as

$$\frac{p_i}{\tilde{w}_i} + \ln p_i - \ln \tilde{w}_i = 1 + \frac{1}{\gamma N} + \overline{\ln p} - \ln \tilde{w}_i \tag{A.2}$$

and applying the Lambert function to the left-hand side, we obtain the following explicit solution for p_i

$$p_i = \mathcal{W}\left[\tilde{w}_i^{-1} \exp\left\{1 + \frac{1}{\gamma N} + \overline{\ln p}\right\}\right] \tilde{w}_i. \tag{A.3}$$

This is the same as equation (11). From (A.2), we see that the equilibrium is invariant to uniform scaling of both prices, p and w.

A.2. Existence and uniqueness of the symmetric autarky equilibrium

Rewriting (21), we find an equilibrium as a root of the function

$$G[m] := g[m] - \rho[m]\psi[m].$$

Remember that $g[m] > 1$ is the inverse of an aggregate version of the familiar measure of the degree of scale economies, i.e., the ratio of average to marginal cost. We expect this to be falling in m: The larger firm size m, and the smaller the number of firms, the closer average cost to marginal cost. In turn, $\rho[m] := 1 + \frac{1}{\gamma N[m]}$ and $\psi[m] := \frac{\eta[m]+1}{\eta[m]}$ are the two markups on the goods and the labor market, respectively. Given that a symmetric equilibrium has $N = H/m$, we have $\rho_m = 1/(\gamma H) > 0$. Moreover, in subsection 2.4 we have shown that $\eta_m < 0$, whence we now have $\psi_m = -\eta_m/\eta[m]^2 > 0$. As expected from intuition, both markups are falling in the number of firms and thus rising in the half-distance between two neighboring firms, m.

Note that $G[m] > 0$ implies that firms make losses, while $G[m] < 0$ implies positive profits. The key question is whether there exists an economically meaningful root m^* satisfying $G[m] = 0$. We first explore the properties of $g[m] := \frac{L\theta[m]}{L\theta[m]-\alpha N[m]}$. Implicitly defining $\tilde{m}$ by

$$\frac{L}{N[\tilde{m}]}\theta[\tilde{m}] = \alpha, \tag{A.4}$$

we see that $g[m]$ approaches $+\infty$ as m approaches $\tilde{m}$ from above and $-\infty$ as m approaches $\tilde{m}$ from below. We therefore have

$$\lim_{m\to\tilde{m}^+} G[m] = \infty \tag{A.5}$$

And finally, $g[m]$ approaches unity as m approaches ∞.

The value $\tilde{m}$ gives us the *minimum* size (in terms of "skill-reach") that symmetric firms populating an economy with a labor force equal to L must achieve, in order to be able to produce a positive amount of output, given the fixed cost α. The *maximum* firm size is reached with a single firm which is of size H, since $N = H/m$. It now becomes obvious that a *necessary* condition for the existence of a meaningful equilibrium, i.e. an equilibrium with positive firm-output, requires that $\tilde{m} < H$. Rewriting (A.4) as $F[m]/H = \frac{\alpha}{L}$ shows that $\tilde{m} < H$ is equivalent to $F[H]/H > \frac{\alpha}{L}$. Observing that the left-hand side falls in H, $\tilde{m} < H$ thus requires that, given the degree of worker heterogeneity, fixed cost is not too large relative to the size of the labor force. However, this condition is not sufficient for an economically meaningful equilibrium to exist. The following conditions are *sufficient* for such an equilibrium to *exist* and to be *unique*: i) $\tilde{m} < H$, ii) $G[\tilde{m}] > 0$, iii) $G[H] < 0$ and iv) $G_m > 0$ and $G[m]$ is continuous in the interval $(\tilde{m}, H]$. Condition i) has been addressed in the previous paragraph, and ii) follows from the above limit properties of $g[m]$. Condition iii) requires

$$\frac{L\theta[H]}{L\theta[H]-\alpha}\Big/ \psi[H] \leq \left(1+\frac{1}{\gamma}\right). \tag{A.6}$$

This is a separate condition, requiring that a single firm in the market makes at least zero profits. Observing (A.4), we can rewrite this as $\frac{\alpha}{L}\frac{\psi[H]\left(1+\frac{1}{\gamma}\right)}{\psi[H]\left(1+\frac{1}{\gamma}\right)-1} \leq \frac{F[H]}{H}$ which shows again that small fixed cost in relation to the size of labor force are conducive to existence of an equilibrium. Regarding the degree of worker heterogeneity, we have that the left-hand side, as well as the right-hand side, increases in H, with the net effect being ambiguous without any further restrictions on $f[\cdot]$. However, setting $\psi[H]$ to its minimum level of unity, we obtain $\frac{\alpha}{L}(1+\gamma) \leq \frac{F[H]}{H}$ as a sufficient condition for (A.6) to hold a fortiori as H falls. It shows, that given α, L and H, the degree of substitutability of goods in the utility function γ, that governs the price elasticity of demand, must not be too large. Relating back to (A.6) in its original form, these restrictions imply that the price markup over marginal cost that a single firm can choose exceeds its average cost.[41]

And finally, since we know from above that $\rho_m > 0$ as well as $\psi_m > 0$, condition iv) is satisfied if $g_m < 0$. It is straightforward to show that

$$g_m[m] = \frac{\frac{L}{H}f[m]}{\frac{mL}{H}\theta[m]-\alpha}\left(1-\frac{\frac{mL}{H}\theta[m]}{\frac{mL}{H}\theta[m]-\alpha}\right) < 0, \tag{A.7}$$

which completes the proof.

[41] This condition is well known from the standard New Trade Theory model with homogeneous workers (cp. Equation (10) in Krugman, 1980). Note that a homogeneous worker version with $\theta = 1$ of (A.4) must of course also hold in the standard model.

A.3. The limiting case of $H \to 0$

As we let the degree of skill heterogeneity approach zero, our equilibrium converges to the equilibrium of a monopolistic competition model with translog preferences. From the previous appendix it follows that if an equilibrium exists with some $\bar{H}$, it also exists for $H < \bar{H}$. In all of these equilibria, m will be smaller than $\bar{H}$, ensuring $H/m = N > 1$. Consider an exogenous decrease in the degree of skill differentiation $\hat{H} < 0$ within the interval $(0, \bar{H}]$. A smaller circumference means that the mass of labor on any interval of the skill circle increases. Holding m constant for a moment, this would allow firms to expand output without having to rely on workers with less suitable types of skills, thus increasing the degree of scale economies and decreasing $g[m]$. Moreover, from $N = H/m$ a smaller H means a lower number of firms, which implies a higher goods price markup. But this, together with the size effect, implies positive profits. Hence, $\hat{N} = \hat{H}$ with $\hat{m} = 0$ is not an equilibrium adjustment. Totally differentiating (21), we obtain

$$\hat{m} = \frac{g_H - \psi[m]\rho_H}{-g_m + \psi[m]\rho_m + \rho[m]\psi_m}\frac{H}{m}\hat{H} = \frac{g[m](g[m]-1) + \psi[m]\frac{m}{\gamma H}}{\frac{f[m]}{\theta[m]}g[m](g[m]-1) + \psi[m]\frac{m}{\gamma H} + \frac{\psi_m m}{\psi[m]}}\hat{H}.$$

The "multiplier" in front of $\hat{H}$ is positive, meaning that m falls as H decreases, but $f[m]/\theta[m] < 1$ and $\psi_m m/\psi[m] \geq 0$ imply that the multiplier can be greater or smaller one. Thus, the net effect on $N = H/m$ is generally ambiguous. Now, let $H \to 0$, whence $m = H/N$ must approach zero as well. Therefore, $f[m]/\theta[m]$ goes to unity and $\psi_m m/\psi[m] \geq 0$ goes to zero, so that the multiplier approaches unity and N converges to a constant $\underline{N}$. Returning to the equilibrium condition (21) and letting $m \to 0$ ($\theta[m] \to 1, \psi[m] \to 1$) and $H/m = N \to \underline{N}$. We finally obtain that $\underline{N}$ must satisfy

$$\frac{L}{L - \alpha\underline{N}} = 1 + \frac{1}{\gamma\underline{N}}$$

which is the equilibrium condition for the number of firms in a Krugman (1979)-type model with homogeneous workers and translog preferences.

A.4. The constrained social optimum

The social planner maximizes log utility with respect to m and subject to the condition that price equals average cost (AC) and the endowment constraint which we can combine to $p = \frac{L\theta[m]}{L\theta[m] - \alpha N[m]}$:

$$\max_m \ \ln V = \ln\theta[m] - \left(\frac{1}{2\gamma N[m]} + \ln p[m]\right) \qquad \text{s.t.} \qquad p[m] = \frac{L\theta[m]}{L\theta[m] - \alpha N[m]} \tag{A.8}$$

The first order condition results as

$$\frac{Lf[m]}{L\theta[m] - \frac{\alpha H}{m}} = 1 + \frac{m}{2\gamma H}. \tag{A.9}$$

The second order condition for a maximum holds since, as we can show, the welfare function is globally concave, i.e. $\frac{\mathrm{d}^2 \ln V}{\mathrm{d}m^2} = -\frac{\left(L\theta_m[m] + \frac{\alpha H}{m^2}\right)^2}{\left(L\theta[m] - \frac{\alpha H}{m}\right)^2} + \frac{L\theta_{mm}[m] - 2\frac{\alpha H}{m^3}}{L\theta[m] - \frac{\alpha H}{m}} < 0$. A sufficient condition for this to hold is $\theta_{mm}[m] := \frac{\partial^2\theta[m]}{\partial m^2} = \frac{1}{m}\left(f'[m] - \frac{2}{m}f[m] + \frac{2}{m}\theta[m]\right) \leq 0$ which requires $f[m] \geq \theta[m] + \frac{m}{2}f'[m]$. Since concavity of $f[\cdot]$ implies $f[m] \geq f\left[\frac{m}{2}\right] + \frac{m}{2}f'[m]$ and (by Jensen's inequality) $f\left[\frac{m}{2}\right] \geq \theta[m]$, it follows that $f[m] \geq f\left[\frac{m}{2}\right] + \frac{m}{2}f'[m] \geq \theta[m] + \frac{m}{2}f'[m]$ and therefore $\theta_{mm}[m] \leq 0$ and $\frac{\partial^2 \ln V}{\partial m^2} < 0$ always hold.

To compare the planer's solution with the laissez faire equilibrium determined by (21) we rewrite

(A.9) as $g[m] = \frac{\theta[m]}{f[m]} \frac{1}{\psi[m]} \psi[m]\rho[m/2]$. The difference between the two conditions appears on the right-hand side of this equation. Since $g_m < 0$, the social planer's solution implies a larger m than the market equilibrium, if the right-hand side is smaller than $\psi[m]\rho[m]$ for all values of m. Since $\rho_m > 0$,

$$\frac{\theta[m]}{f[m]} \frac{1}{\psi[m]} < 1 \tag{A.10}$$

is a sufficient condition for this to hold. We show next that concavity of $f[\cdot]$ suffices to establish this result. Rearranging (A.10) and inserting $\psi[m] = \frac{f[m]^2 - 2f'[m]m\theta[m]}{f[m]^2}$ yields $\frac{1+\frac{2}{f[m]}f'[m]m}{f[m]} < \frac{1}{\theta}$ which holds a fortiori because concavity of $f[\cdot]$ implies that $\frac{1+f'[m]m}{f[m]} < 1$. Hence, condition (A.10) is fulfilled and it follows that the market equilibrium firm size is too small compared to the socially optimal allocation.

A.5. Further details of the trading equilibrium

A.5. Proof of proposition 1

(i) Log-differentiating the equilibrium condition (25) and setting $k = 1$, we obtain

$$\hat{m} = A \cdot \hat{k} \qquad \text{with} \qquad A := \frac{\psi[m]\frac{1}{\gamma H}}{-g_m[m] + \psi[m]\frac{1}{\gamma H} + \rho^T[m]\psi_m[m]}. \tag{A.11}$$

Since $g_m < 0$ and $\psi_m > 0$,[42] we find that $0 < A < 1$ which implies $0 < \hat{m} = A \cdot \hat{k} < \hat{k}$. Hence, m increases and the number of firms in each country falls. However, $A < 1$ implies that the total number of available varieties $N^T = k \cdot N > N^A$ is still larger with trade than under autarky.

(ii) As the price markup depends negatively on the number of available varieties $k \cdot N$, it follows directly from the previous result that it must fall. Furthermore, we know from above that the wage markup increases. Log-differentiating(23) and again setting $k = 1$ yields

$$\hat{p} = B \cdot \hat{k} \qquad \text{with} \qquad B = \frac{\frac{m}{\gamma H}}{\left(1 + \frac{m}{\gamma H}\right)} \frac{g_m[m]}{\left(-g_m[m] + \psi[m]\frac{1}{\gamma H} + \rho^T[m]\psi_m[m]\right)}. \tag{A.12}$$

Since $-1 < B < 0$, it follows that $\hat{p} < 0$.

(iii) This follows from $\theta_m = \frac{1}{m}(f[m] - \theta[m]) < 0$.

(vi) Real income, $\theta[m]/p[m]$, must increase by virtue of the excess entry result demonstrated in A.4. With higher real income and a larger variety available for consumption as established in (i), it follows from (22) that welfare of the worker earning average income increases.

[42] see appendix A.2 for details

A.5. Proof of proposition 2

In the symmetric equilibrium with identical countries the average price in the domestic and the foreign market is the same and given by $\overline{\ln p} = \overline{\ln p}^* = 1/2 \ln p + 1/2 \ln p^*$. Inserting $\overline{\ln p}$ and $\overline{\ln p}^*$ into the Z-terms in (29), (30), we can use the same logic as in A.1 to obtain explicit solutions for p and p^*, where the price markups no longer depend on the own price, but only on the respective other price and the number of firms: $p = \frac{\mathcal{W}[\tilde{Z}]}{2}\psi$ with $\tilde{Z} = \frac{2}{\psi}\exp\left\{2 + \frac{m}{\gamma H} + \ln p^*\right\}$ and $p^* = \frac{\mathcal{W}[\tilde{Z}^*]}{2}\psi\tau$ with $\tilde{Z}^* = \frac{2}{\psi\tau}\exp\left\{2 + \frac{m}{\gamma H} + \ln p\right\}$. Inserting $p = \frac{\mathcal{W}[\tilde{Z}]}{2}\psi$ and $p^* = \frac{\mathcal{W}[\tilde{Z}^*]}{2}\psi\tau$ into the $\tilde{Z}$-terms, we obtain

$$p = \mathcal{W}\left[\mathcal{W}[\tilde{Z}^*]\tau \exp\left\{2 + \frac{m}{\gamma H}\right\}\right]\frac{\psi}{2} \qquad \text{and} \qquad p^* = \mathcal{W}\left[\frac{\mathcal{W}[\tilde{Z}]}{\tau}\exp\left\{2 + \frac{m}{\gamma H}\right\}\right]\frac{\psi}{2}\tau. \qquad \text{(A.13)}$$

It proves convenient to focus on the price markup values $W = \mathcal{W}[\tilde{Z}]$ and $W^* = \mathcal{W}[\tilde{Z}^*]$ instead of prices. The corresponding system of equations determining these values emerges as

$$W = W[W^*, m] = \mathcal{W}\left[W^*\tau \exp\left\{2 + \frac{m}{\gamma H}\right\}\right] \qquad \text{(A.14)}$$

$$W^* = W^*[W, m] = \mathcal{W}\left[\frac{W}{\tau}\exp\left\{2 + \frac{m}{\gamma H}\right\}\right]. \qquad \text{(A.15)}$$

Note that at zero trade cost ($\tau = 1$) the price markups are identical. While the markup on domestic varieties increases in τ, the markup on foreign varieties falls in the level of trade cost. For any $\tau > 1$, it must therefore be true that $W > W^*$. Note that the two country version of (A.1) can be written as $p = \left(1 + \frac{1}{\gamma N^T} + \frac{1}{2}\ln p^* - \frac{1}{2}\ln p\right)\tilde{w}$ and analogously for p^*. In view of (A.13) it follows that $\frac{W}{2} = 1 + \frac{1}{\gamma N^T} + \frac{1}{2}\ln p^* - \frac{1}{2}\ln p$ and $\frac{W^*}{2} = 1 + \frac{1}{\gamma N^T} + \frac{1}{2}\ln p - \frac{1}{2}\ln p^*$. The expenditure shares in (28) can therefore be written as

$$\delta = \left(\frac{W}{2} - 1\right)\gamma \qquad \text{and} \qquad \delta^* = \left(\frac{W^*}{2} - 1\right)\gamma. \qquad \text{(A.16)}$$

Rewriting the demand function for foreign varieties (27) in terms of W^* leads to $q^* = \left(1 - \frac{2}{W^*}\right)\frac{\gamma Y}{\psi}$. This implies that the prohibitive level of trade cost $\bar{\tau}$ for which $q^* = 0$ satisfies $\mathcal{W}\left[\frac{W}{\bar{\tau}}\exp\left\{2 + \frac{2}{\gamma N^T}\right\}\right] \equiv 2$. It follows that for non-prohibitive trade cost $W \geq W^* \geq 2$. Inserting demand (27), expenditure shares (A.16), prices $p = \frac{W}{2}\psi$, $p^* = \frac{W^*}{2}\psi\tau$ as well as income $Y = L\theta$ into the labor market clearing condition (31), and rearranging terms gives

$$\gamma\left(2 - \frac{2}{W} - \frac{2}{W^*}\right) = \frac{\frac{L\theta[m]}{N[m]} - \alpha}{L\theta[m]}\psi[m] \qquad \Leftrightarrow \qquad \gamma h[W, W^*] = \frac{\psi[m]}{g[m]N[m]}. \qquad \text{(A.17)}$$

For easier reference the second line introduces $h[W, W^*] := \left(2 - \frac{2}{W} - \frac{2}{W^*}\right)$. (A.17), (A.14) and (A.15) form our system of equations in W, W^* and m.

(i) Comparative statics of firm size and markups. The proof of proposition 2 requires that we solve this system for an exogenous change in τ. Doing so by log-linearization, we write the solution as $\widehat{W} = \omega \cdot \hat{\tau}$, $\widehat{W}^* = \omega^* \cdot \hat{\tau}$ and $\widehat{m} = \mu \cdot \hat{\tau}$. We next explore the sign of the elasticities ω, ω^* and μ. For notational convenience we suppress the functional dependence of N and ψ on m in the following, whenever it is not crucial. Log- differentiating (A.17), (A.14), (A.15) leads to

$$\begin{bmatrix} -\frac{\partial \ln h}{\partial \ln W} & -\frac{\partial \ln h}{\partial \ln W^*} & \frac{\partial \ln \psi}{\partial \ln m} - \frac{\partial \ln g}{\partial \ln m} - \frac{\partial \ln N}{\partial \ln m} \\ -1 & \frac{\partial \ln W}{\partial \ln W^*} & \frac{\partial \ln W}{\partial \ln m} \\ \frac{\partial \ln W^*}{\partial \ln W} & -1 & \frac{\partial \ln W^*}{\partial \ln m} \end{bmatrix} \begin{bmatrix} \widehat{W} \\ \widehat{W}^* \\ \widehat{m} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{\partial \ln W}{\partial \ln \tau} \cdot \hat{\tau} \\ -\frac{\partial \ln W^*}{\partial \ln \tau} \cdot \hat{\tau} \end{bmatrix}$$

$$\begin{bmatrix} -\frac{1}{W-1-\frac{W}{W^*}} & -\frac{1}{W^*-1-\frac{W^*}{W}} & \frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N}-\alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \\ -1 & \frac{1}{W+1} & \frac{1}{\gamma N}\frac{1}{W+1} \\ \frac{1}{W^*+1} & -1 & \frac{1}{\gamma N}\frac{1}{W^*+1} \end{bmatrix} \begin{bmatrix} \widehat{W} \\ \widehat{W}^* \\ \widehat{m} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{W+1} \cdot \hat{\tau} \\ \frac{1}{W^*+1} \cdot \hat{\tau} \end{bmatrix}. \tag{A.18}$$

Denoting the 3×3-matrix of derivatives by D, it follows that

$$\omega = \frac{1}{(W+1)(W^*+1)} \left[\left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \right) W^* - \frac{1}{\gamma N h[W,W^*]} \frac{4}{W^*} \right] \frac{1}{\det[D]} \tag{A.19}$$

$$\omega^* = \frac{1}{(W+1)(W^*+1)} \left[- \left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \right) W + \frac{1}{\gamma N h[W,W^*]} \frac{4}{W} \right] \frac{1}{\det[D]} \tag{A.20}$$

$$\mu = \frac{2W^*/W - 2W/W^*}{h[W,W^*](W+1)(W^*+1)} \frac{1}{\det[D]}. \tag{A.21}$$

The signs of the elasticities hinge upon the sign of the determinant which is given by

$$\det[D] = \left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \right) \frac{WW^* + W + W^*}{(W+1)(W^*+1)} - \frac{1}{\gamma N h[W,W^*]} \frac{(2+W^*)\frac{2}{W} + (2+W)\frac{2}{W^*}}{(W+1)(W^*+1)}.$$

Since $WW^* > 2$ and $W \geq W^*$, we have $WW^* + W + W^* > (2+W^*)\frac{2}{W} + (2+W)\frac{2}{W^*}$. This implies that $\det[D] > 0$ if

$$\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} > \frac{1}{\gamma N h[W,W^*]}. \tag{A.22}$$

We know from above that $\frac{f[m]}{\theta[m]} < 1$ and $\frac{\partial \ln \psi}{\partial \ln m} > 0$, and therefore, inequality (A.22) holds if

$$\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} > \frac{1}{\gamma N h[W,W^*]}. \tag{A.23}$$

Using the equilibrium condition (A.17), we can rewrite this as $\psi[m] \geq \theta[m]/f[m]$. We have proven in appendix A.4 that this inequality always holds. Hence, it follows that $\det[D] > 0$.

Returning to our elasticity ω, we note that $W^* \geq \frac{4}{W^*}$, $\det[D] > 0$ and (A.22) imply $\omega > 0$. By analogy, it follows that $\omega^* < 0$. And finally, $W \geq W^*$ implies that $\mu \leq 0$. For reasons pointed out in the text, μ is monotonic in the initial level of trade cost, converging to zero as τ approaches one. Looking at A.21, the level of τ enters through W and W^*. The lower the trade cost, the smaller the difference between W and W^*. At $\tau = 1$, price markups are identical and $m = 0$. This proves part (i) of the proposition.

(ii) Changes in prices. The proposition states that for $\hat{\tau} < 0$, $\hat{p}^* < 0$ while $\hat{p}$ is ambiguous. The price of imported varieties is affected by the change in τ and the changes in both markups

$$\hat{p}^* = \left(\omega^* + \frac{\partial \ln \psi}{\partial \ln m}\mu + 1\right)\hat{\tau} \tag{A.24}$$

where $\frac{\partial \ln \psi}{\partial \ln m} = \frac{-2mf''[m]F[m]}{f[m]^2\psi[m]} - \frac{2mf'[m]}{f[m]} > 0$. Inserting (A.20) and (A.21) shows that $\hat{p}^*$ is positive if and only if

$$-\frac{d_{13}W - \frac{2}{W}\frac{2}{\gamma hN} + \frac{\partial \ln \psi}{\partial \ln m}\frac{1}{h[W,W^*]}\left(\frac{2W}{W^*} - \frac{2W^*}{W}\right)}{d_{13}(WW^* + W + W^*) - \frac{2}{\gamma h[W,W^*]N}\left(\frac{2+W^*}{W} + \frac{2+W}{W^*}\right)} + 1 > 0 \tag{A.25}$$

where d_{13} is the element in row 1 and column 3 of D. Canceling identical terms in the denominator and the numerator shows that this is true if $\frac{\frac{\partial \ln \psi}{\partial \ln m}\frac{1}{h[W,W^*]}\left(\frac{2W}{W^*} - \frac{2W^*}{W}\right)}{d_{13}(WW^*+W^*) - \frac{2}{\gamma h[W,W^*]N}\left(\frac{W^*}{W} + \frac{2+W}{W^*}\right)} < 1$. Noting that $d_{13} = \frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m}$ and observing the inequality in (A.23), it follows that $WW^* + W^* \geq \frac{2W^*}{W} + \frac{4+2W}{W^*}$ and $WW^* + W^* \geq \frac{1}{h[W,W^*]}\left(\frac{2W}{W^*} - \frac{2W^*}{W}\right)$ is sufficient for the inequality in (A.25) to hold. Using from above $W \geq W^* \geq 2$, it is straightforward to show that these two conditions are fulfilled.

The change in the domestic price obtains as

$$\hat{p} = \left(\omega + \frac{\partial \ln \psi}{\partial \ln m}\mu\right)\hat{\tau}. \tag{A.26}$$

We know from above that $\omega > 0$; the pro-competitive effect of lower trade cost on the goods market. This is potentially offset by an increase in the wage markup. For τ close to one, the goods market effect clearly dominates as μ is close to zero.

Conversely, at $\bar{\tau}$ (prohibitive trade cost), the labor market effect dominates. Inserting (A.19) and (A.21) gives

$$\hat{p} = \left[W^*\left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m}\right) - \frac{2}{\gamma N h[W,W^*]}\frac{2}{W^*} - \frac{\partial \ln \psi}{\partial \ln m}\frac{1}{h[W,W^*]}\left(\frac{2W}{W^*} - \frac{2W^*}{W}\right)\right]$$
$$\times \frac{1}{(W+1)(W^*+1)}\frac{\hat{\tau}}{\det[D]}. \tag{A.27}$$

Remember that prohibitive trade cost imply an infinite price elasticity and therefore a price markup of zero, whence $W^* = 2$. To see if $\hat{p} > 0$ for $\tau = \bar{\tau}$, as stated in proposition 2, we must therefore evaluate the bracketed term at $W^* = 2$. We obtain $-2\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} + 2\frac{f[m]}{\theta[m]} - 2 - 2\frac{\partial \ln \psi}{\partial \ln m} + \frac{2}{\gamma N h[W,W^*]}\frac{2}{W^*} + \frac{\partial \ln \psi}{\partial \ln m}(W+2)$. Inserting the equilibrium condition (A.17), which reduces to $\gamma h[W,W^*] = \frac{L\theta[m]/N - \alpha}{L\theta[m]}\psi = \frac{2}{W}\frac{1}{N}$ at $\tau = \bar{\tau}$, shows that the expression is negative, if $\psi W\frac{f[m]}{\theta[m]} < 2\frac{f[m]}{\theta[m]} + W - 2 + W\frac{\partial \ln \psi}{\partial \ln m}$. Inserting the explicit expressions for ψ and $\frac{\mathrm{d} \ln \psi}{\mathrm{d} \ln m}$ leads to $\frac{W}{\theta[m]}\frac{f[m]^2 - 2f'[m]F[m]}{f[m]} < W - 2 + \frac{2f[m]}{\theta[m]} + W\left(\frac{-2f''[m]\theta}{f[m]^2\psi} - \frac{2mf'[m]}{f[m]}\right)$. Since $f''[m] \leq 0$, the inequality holds if $\frac{W}{\theta[m]}\frac{f[m]^2 - 2f'[m]F[m]}{f[m]} < W - 2 + \frac{2f[m]}{\theta[m]} - W\frac{2mf'[m]}{f[m]}$. Rearranging terms shows that this inequality holds if $f[m] < \theta[m]$, which is true given $f'[m] < 0$. This completes the proof of part (ii) of proposition 2.

(iii) Welfare. Indirect utility of the worker receiving average income in the equilibrium with trade cost is given by $\ln V = \ln\theta[m] - \ln P^T[p,p^*,m]$, where $\ln P^T[p,p^*,m] = \frac{1}{2\gamma N^T} + \frac{1}{N^T}\sum_{i=1}^{N^T}\ln p_i + \frac{\gamma}{2N^T}\sum_{i=1}^{N^T}\sum_{j=1}^{N^T}\ln p_i(\ln p_j - \ln p_i)$ with $N^T = N + N^*$ and i,j indexing domestic *and* foreign varieties. Under symmetry, which implies $N^* = N = N^T/2$, the price index simplifies to $\ln P^T[p,p^*,m] = \frac{1}{4\gamma N} + \frac{1}{2}\ln p + \frac{1}{2}\ln p^* - \frac{\gamma N}{4}(\ln p - \ln p^*)^2$. The change in indirect utility is then

$$\widehat{V} = \left(\frac{\partial\ln\theta}{\partial\ln m} - \frac{\partial\ln P}{\partial\ln m}\right)\hat{m} - \frac{\partial\ln P}{\partial\ln p}\hat{p} - \frac{\partial\ln P}{\partial\ln p^*}\hat{p}^* \tag{A.28}$$

with $\frac{\partial\ln\theta}{\partial\ln m} = \frac{f[m]-\theta[m]}{\theta[m]} < 0$, $\frac{\partial\ln P}{\partial\ln m} = \frac{1}{4\gamma N} + \frac{\gamma N}{4}(\ln p - \ln p^*)^2 > 0$, $\frac{\partial\ln P}{\partial\ln p} = \frac{1}{2} + \frac{\gamma N}{2}(\ln p - \ln p^*) \geq 0$ and $\frac{\partial\ln P}{\partial\ln p^*} = N\delta^* \geq 0$. Inserting yields equation (32). Using the results that at the prohibitive level of trade cost $\delta^* = 0$, $\hat{p} > 0$ and $\hat{m} > 0$, it follows from (32) that $\widehat{V} < 0$ at $\tau = \bar{\tau}$. Since at $\tau = 1$ it holds that $\hat{m} = 0$, $\hat{p} < 0$ and $\hat{p}^* < 0$, it follows that $\widehat{V} > 0$ at $\tau = 1$.

A.6. Additional details of the trade and migration equilibrium

A.6. The supply bonus from locating next to a foreign neighbor

The labor supply in the case where the neighbors are in the same country, wages are equal, the distance between any two firms is $d[\bar{w}] = m/2$ and $\lambda \in [0,\bar{\lambda}]$, is given by $LE^D = \frac{L}{H}(2-\lambda)\int_0^{m/2} f[d]\mathrm{d}d$. The labor supply for the case with foreign neighbors under the otherwise same conditions where $0 \leq d^m[\bar{w}] \leq \frac{m}{2}$ and $\frac{m}{2} \leq d^n[\bar{w}] \leq m$ is $LE^F = \frac{L}{H}\int_0^{d^n[\bar{w}]} f[d]\mathrm{d}d + \frac{L}{H}(1-\lambda)\int_0^{d^m[\bar{w}]} f[d]\mathrm{d}d$. Then, the difference in supply of efficiency units for a given wage results as $LE^F - LE^D = \frac{L}{H}\int_{m/2}^{d^n[\bar{w}]} f[d]\mathrm{d}d - (1-\lambda)\frac{L}{H}\int_{m/2}^{d^m[\bar{w}]} f[d]\mathrm{d}d$. Using the fact that in the symmetric equilibrium $\int_{d^m[\bar{w}]}^{m/2} f[d]\mathrm{d}d = \int_{m/2}^{d^n[\bar{w}]} f[m-d]\mathrm{d}d$ this can be rewritten as $LE^F - LE^D = \frac{L}{H}\int_{m/2}^{d^n[\bar{w}]}(f[d] - (1-\lambda)f[m-d])\,\mathrm{d}d \geq 0$. The inequality follows from $f[d] - (1-\lambda)f[m-d] \geq 0 \;\forall\; 0 \leq d \leq d^n[\bar{w}]$. Hence, in the symmetric equilibrium the labor supply for a given wage is (weakly) larger if the neighbors are foreign. If $\lambda = 0$, labor supply is identical in both cases.

A.6. The elasticity of labor supply

The elasticity of labor supply is defined as $\frac{\partial L_i^{ES,M}}{\partial w_i}\frac{w_i}{L_i^{ES,M}}$. From (35), (33), and (34), we obtain $\frac{\partial L_i^{ES,M}}{\partial w_i} = \frac{L}{H}\frac{\mathrm{d}d_i^n}{\mathrm{d}w_i}f[d_i^n] + (1-\lambda)\frac{L}{H}\frac{\mathrm{d}d_i^m}{\mathrm{d}w_i}f[d_i^m]$ with $\frac{\mathrm{d}d_i^n}{\mathrm{d}w_i} = \frac{f[d_i^n]}{-w_i f'[d_i^n] - w^*(1-\lambda)f'[m-d_i^n]}$ and $\frac{\mathrm{d}d_i^m}{\mathrm{d}w_i} = \frac{(1-\lambda)f[d_i^m]}{-w_i(1-\lambda)f'[d_i^m] - w^* f'[m-d_i^m]}$. Evaluating $\frac{\partial L_i^{ES,M}}{\partial w_i}\frac{w_i}{L_i^{ES,M}}$ at the symmetric equilibrium, where it holds that $w_i = w^* \equiv 1$, $d_i^n = d^n$, $d_i^m = d^m = m - d^n$ and $f[d^n] = (1-\lambda)f[d^m]$, we obtain

$$\begin{aligned}\eta^M = \frac{\partial L_i^{ES,M}}{\partial w_i}\frac{w_i}{L_i^{ES}}\bigg|_{w_i=w} =& \frac{L}{H}\left(\frac{f[d^n]^2}{-f'[d^n] - (1-\lambda)f'[m-d^n]} + \frac{(1-\lambda)^2 f[d^m]^2}{-(1-\lambda)f'[d^m] - f'[m-d^m]}\right)\\ &\times \frac{1}{\frac{L}{H}\left(\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\int_0^{d^m} f[d]\mathrm{d}d\right)}\\ =& \frac{2f[d^n]^2}{f'[d^n] + (1-\lambda)f'[d^m]}\cdot\frac{-1}{\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\int_0^{d^m} f[d]\mathrm{d}d}\end{aligned}$$

as displayed in (37). The elasticity of labor supply decreases in m:

$$\eta_m^M = \eta^M \left[\frac{2f'[d^n]}{f[d^n]} \frac{\mathrm{d}d^n}{\mathrm{d}m} - \frac{-f''[d^n]\frac{\mathrm{d}d^n}{\mathrm{d}m} - (1-\lambda)f''[d^m]\frac{\mathrm{d}d^m}{\mathrm{d}m}}{-f'[d^n] - (1-\lambda)f'[d^m]} - \frac{f[d^n]}{m\theta^M} \right] < 0, \tag{A.29}$$

where $\frac{\mathrm{d}d^n}{\mathrm{d}m} = \frac{(1-\lambda)f'[d^m]}{f'[d^n]+(1-\lambda)f'[d^m]} > 0$ and $\frac{\mathrm{d}d^m}{\mathrm{d}m} = \frac{f'[d^n]}{f'[d^n]+(1-\lambda)f'[d^m]} > 0$. Furthermore, provided that $f'''[\cdot]$ is not too positive, η^M decreases in λ:

$$\eta_\lambda^M = \eta^M \left[\underbrace{\frac{2f'[d^n]\frac{\mathrm{d}d^n}{\mathrm{d}\lambda}}{f[d^n]}}_{<0} + \underbrace{\frac{f''[d^n]\frac{\mathrm{d}d^n}{\mathrm{d}\lambda} + (1-\lambda)f''[d^m]\frac{\mathrm{d}d^m}{\mathrm{d}\lambda} + f'[d^m]}{-f'[d^n] - (1-\lambda)f'[d^m]}}_{<0} + \underbrace{\frac{F[d^m]}{F[d^n] + (1-\lambda)F[d^m]}}_{>0} \right] < 0 \tag{A.30}$$

with $\frac{\mathrm{d}d^n}{\mathrm{d}\lambda} = \frac{f[d^m]}{-f'[d^m]} - f'[d^n] - (1-\lambda)f'[d^m] > 0$ and $\frac{\mathrm{d}d^m}{\mathrm{d}\lambda} = -\frac{\mathrm{d}d^n}{\mathrm{d}\lambda} < 0$. $\eta_\lambda^M < 0$ follows from the fact that the first term in the brackets (in absolute terms) exceeds the third, since

$$\frac{2f'[d^n]}{f[d^n]} \frac{\mathrm{d}d^n}{\mathrm{d}\lambda} = \underbrace{2\frac{f[d^m]}{f[d^n]} \cdot \frac{f'[d^n]}{f'[d^n] + (1-\lambda)f'[d^m]}}_{\geq 1} \geq \underbrace{\frac{F[d^m]}{F[d^n] + (1-\lambda)F[d^m]}}_{<1}.$$

A.6. Analytical details of the proof of proposition 3

The number of firms is too large in the migration equilibrium. The social planner solves the same maximization problem as in appendix A.4, additionally taking into account the integrated labor market.[43] The first order condition of the planner then obtains as

$$\frac{Lf\,[d^n]}{L\theta^M - \frac{\alpha H}{m}} = 1 + \frac{m}{4\gamma H}. \tag{A.31}$$

where d^n, θ^M are shorthands for $d^n[m,\lambda], \theta^M[m,\lambda]$, respectively. A comparison with the market solution (40) shows that, as before, the number of firms in the market equilibrium is too large if the markup distortion is larger than the productivity distortion. We can show that this is the case in the migration equilibrium with non-prohibitive λ. The relevant condition is $\psi^M > \frac{\theta^M}{f[d^n]}$. Inserting for ψ^M this is equivalent to $1 - \frac{m\theta^M\left(f'[d^n]+(1-\lambda)f'[d^m]\right)}{2f[d^n]^2} > \frac{\theta^M}{f[d^n]}$. This, in turn, holds if $1 - \frac{m\theta^M f'[d^n]}{2f[d^n]^2} > \frac{\theta^M}{f[d^n]}$, since $-f'[d^m](1-\lambda)/(2f[d^n]^2) \geq 0$. Rewriting the condition leads to $f[d^n] > \theta^M + \frac{m}{2}\frac{f'[d^n]}{f[d^n]}\theta^M$. We will show below that $f\left[\frac{d^n}{2}\right] \geq \theta^M$. Then, this inequality holds if

$$f[d^n] > f\left[\frac{d^n}{2}\right] + \frac{m}{2}\frac{f'[d^n]}{f[d^n]}\theta^M. \tag{A.32}$$

Concavity of $f[\cdot]$ implies that $f[d^n] \geq f\left[\frac{d^n}{2}\right] + f'[d^n]\frac{d^n}{2}$. Moreover, we have that $f\left[\frac{d^n}{2}\right] + f'[d^n]\frac{d^n}{2} > f\left[\frac{d^n}{2}\right] + f'[d^n]\frac{m}{2}\frac{\theta^M}{f[d^n]}$ because $m \geq d^n$ and $\theta^M > f[d^n]$. Therefore, (A.32) holds a fortiori. Hence, the markup distortion exceeds the productivity distortion and consequently, the number of firms in the

[43] Note that this assumes that either the planner maximizes welfare for both countries or takes as given that a planner in the foreign country solves the exact same problem.

market equilibrium with migration is too large.[44]

Proof that $\theta^M \leq f\left[\frac{d^n}{2}\right]$. Using the expression for θ^M in (36) and Jensen's inequality which states that $f\left[E[x]\right] \geq E\left[f[x]\right]$ for concave functions $f[x]$, we can state

$$\theta^M = \frac{1}{m}\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d \leq \frac{d^n}{m} f\left[\frac{d^n}{2}\right] + (1-\lambda)\frac{d^m}{m} f\left[\frac{d^m}{2}\right]$$

Since $d^n + d^m = m$, we have that $\theta^M \leq \frac{d^n}{m} f\left[\frac{d^n}{2}\right] + (1-\lambda)\frac{d^m}{m} f\left[\frac{d^m}{2}\right]$. This reduces to $\theta^M \leq f\left[\frac{d^n}{2}\right]$ provided that $(1-\lambda) f\left[\frac{d^m}{2}\right] \leq f\left[\frac{d^n}{2}\right]$. From (33) and (34) it follows that a symmetric equilibrium is characterized by $(1-\lambda) = f[d^n]/f[d^m]$, so the condition becomes $\frac{f\left[\frac{d^m}{2}\right]}{f\left[\frac{d^n}{2}\right]} \leq \frac{f[d^m]}{f[d^n]}$, which is implied by $d^m \leq d^n$ and $f''[\cdot] \leq 0$. This completes the proof.

A.6. Proof of proposition 4

Totally differentiating the equilibrium condition (40) yields $\hat{m} = C \cdot \hat{\lambda}$ where C is given by[45]

$$\begin{aligned}
C &= \frac{g_\lambda^M - \rho^T \psi_\lambda^M}{-g_m^M + \rho^T \psi_m^M + \psi^M \rho_m^T} \frac{\lambda}{m} \lesseqgtr 0 \qquad \text{with} \qquad\qquad (\text{A.33})\\
g_\lambda^M &= \frac{L\theta_\lambda^M}{L\theta^M - \alpha N} - \frac{L\theta^M}{(L\theta^M - \alpha N)^2} L\theta_\lambda^M > 0 \quad \text{and} \quad \theta_\lambda^M = -\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d < 0\\
g_m^M &= \frac{L\theta_m^M}{L\theta^M - \alpha N} - \frac{L\theta^M}{(L\theta^M - \alpha N)^2}\left(L\theta_m^M + \frac{\alpha N}{m}\right) < 0 \quad \text{and} \quad \theta_m^M = \frac{1}{m}\left(f[d^n] - \theta^M\right) < 0\\
\psi_\lambda^M &= -\frac{1}{(\eta^M)^2} \cdot \eta_\lambda^M > 0 \text{ with } \eta_\lambda^M \text{ as in (A.30)} \qquad \psi_m^M = -\frac{1}{(\eta^M)^2} \cdot \eta_m^M > 0 \text{ with } \eta_m^M \text{ as in (A.29)}\\
\rho_m^T &= \frac{1}{2\gamma H} > 0.
\end{aligned}$$

While the denominator of C is always positive (a larger firm size m decreases the markup needed for zero profits g^M and increases both the price markup and the wage markup), the sign of the numerator depends on whether the effect of λ on g^M (which is positive) is stronger than the effect on the wage markup (which is also positive). In either case, prices fall as migration cost fall.

The effect on average income is ambiguous. While the partial effect of lower migration cost is positive, there is a countervailing effect when the general equilibrium adjustments lead to firm exit. In either case, however, real income increases when migration cost falls, as the decrease in prices overcompensates the potential decrease in average income. We show this by log-differentiating real income $\frac{\theta^M}{p} = \frac{L\theta^M - \frac{\alpha H}{m}}{L}$ as

[44] There is a subtle point to this proof in that $\theta^M[m, \lambda]$ is not necessarily concave in m, if there is migration. As a result, the social welfare function is not globally concave. However, it can be shown that the first oder condition in A.31 still describes a global maximum and that the social welfare function is monotonously increasing in the relevant range. Details of the proof are available upon request.

[45] Note that for notational convenience here and in the following we omit the functional dependence of $g^M, \psi^M, \rho^M, \theta^M, d^n$ on m and, where relevant, on λ.

obtained by rewriting (40):

$$\mathrm{d}\ln\left[\frac{\theta^M}{p}\right] = \frac{\partial \ln\left[\frac{\theta^M}{p}\right]\lambda}{\partial\lambda}\cdot\hat{\lambda} + \frac{\partial \ln\left[\frac{\theta^M}{p}\right]m}{\partial m}\cdot\hat{m} \tag{A.34}$$

$$\frac{\partial \ln\left[\frac{\theta^M}{p}\right]}{\partial\lambda} = \frac{L\theta^M_\lambda}{L\theta^M - \frac{\alpha H}{m}} < 0 \qquad \text{and} \qquad \frac{\partial \ln\left[\frac{\theta^M}{p}\right]}{\partial m} = \frac{L\theta^M_m + \frac{\alpha H}{m^2}}{L\theta^M - \frac{\alpha H}{m}} > 0. \tag{A.35}$$

In these equations $\theta^M_\lambda = -\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d < 0$ and $\theta^M_m = \frac{1}{m}\left(f[d^n] - \theta^M\right) < 0$. Note that $\frac{\partial \ln\left[\frac{\theta^M}{p}\right]}{\partial m} > 0$ in the relevant range follows from (A.31). Hence, the log-change in real income induced by a decrease in λ is clearly positive, if $\hat{m}$ is also positive. To show that real income also increases if $\hat{m}$ is negative we use (A.33) as well as (A.35) to rewrite (A.34) as

$$\begin{aligned}\mathrm{d}\ln\left[\frac{\theta^M}{p}\right] =& \frac{\lambda}{(L\theta^M - \alpha N)\left(-g^M_m + \rho\psi^M_m + \psi^M\rho_m\right)} \\ &\times\left[\left(L\theta^M_m + \frac{\alpha N}{m}\right)\left(g^M_\lambda - \rho\psi^M_\lambda\right) + \left(-g^M_m + \rho\psi^M_m + \psi^M\rho_m\right)L\theta^M_\lambda\right]\hat{\lambda}.\end{aligned} \tag{A.36}$$

We know that the first fraction on the right-hand side above is positive, hence we must show that the square-bracketed term is negative. Using $\left(L\theta^M_m + \frac{\alpha N}{m}\right)g^M_\lambda = \left[\frac{L\theta^M_m + \frac{\alpha N}{m}}{L\theta^M - \alpha N} - \frac{L\theta^M\left(L\theta^M_m + \frac{\alpha N}{m}\right)}{(L\theta^M - \alpha N)^2}\right]\cdot L\theta^M_\lambda$ and $L\theta^M_\lambda g^M_m = \left[\frac{L\theta^M_m}{L\theta^M - \alpha N} - \frac{L\theta^M\left(L\theta^M_m + \frac{\alpha N}{m}\right)}{(L\theta^M - \alpha N)^2}\right]\cdot L\theta^M_\lambda$ we can reduce the expression in squared brackets on the right-hand side of (A.36) to $L\theta^M_\lambda\left(\frac{\frac{\alpha N}{m}}{L\theta^M - \alpha N} + \psi^M_m\rho + \rho_m\psi^M\right) - \left(L\theta^M_m + \frac{\alpha N}{m}\right)\rho\psi^M_\lambda$. This is negative since $\theta^M_\lambda < 0$ and $\psi^M_\lambda > 0$. Hence, a decrease in λ raises real income also if it leads to exit of firms. This completes the proof of proposition 4.

A.7. Robustness with respect to the specification of migration cost

The proofs of proposition 3 and 4 reveal that our results are valid for more general specifications of the migration cost. The positive welfare effect of the potential of migration established in proposition 3 stems from a first-order welfare gain due the reduction of the markup distortion. Hence, the validity of proposition 3 is maintained, provided that the excess-entry property of the autarky equilibrium is preserved. The proof of proposition 4 shows that positive welfare gains from lower migration cost occur, provided that $\theta^M_\lambda < 0$ and $\eta^M_\lambda < 0$, and that the excess-entry result holds. It is relatively straightforward that this holds for a wide range of migration cost specifications.