Heterogeneous Workers, Trade, and Migration*

Inga Heiland[a]
University of Oslo
SSB, CESifo, CEPR

Wilhelm Kohler[b]
University of Tuebingen
CESifo, GEP and IAW

February 2021

Abstract

We analyze the welfare effects of trade and migration, focusing on two-sided horizontal heterogeneity among workers and firms. Horizontal (skill-type) heterogeneity among workers generates monopsonistic labor markets as well as within-firm wage inequality and an endogenous quality of worker-firm matches. In a model combining horizontal worker heterogeneity with monopolistic competition on goods markets, trade liberalization causes firm exit which raises wage markups and worsens the average quality of worker-firm matches. It also increases the degree of income inequality. Yet, aggregate welfare is higher under free trade than under autarky. Integration of labor markets leads to two-way migration between symmetric countries. Liberalizing migration has an ambiguous effect on the quality of worker-firm matches and income inequality, but it leads to lower wage markups and lower goods prices and thus to higher welfare in both countries. Our model clearly advocates opening up labor markets simultaneously with trade liberalization.

JEL codes: F12, F16, F22, J24

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

---

[a]University of Oslo and Statistics Norway. Moltke Moes Vei 31, 0831 Oslo, Norway; inga.heiland@econ.uio.no. Corresponding author.

[b]University of Tuebingen, Nauklerstrasse 50, 72074 Tuebingen, Germany; wilhelm.kohler@uni-tuebingen.de.

*This is a revision of "Heterogeneous Workers, Trade, and Migration" published as CESifo Working Paper No. 7355 (2018).

This paper has greatly benefitted from helpful comments by Pol Antràs, Carsten Eckel, Gabriel Felbermayr, Gene Grossman, Elhanan Helpman, Marc Melitz, Gianmarco Ottaviano, Assaf Razin, Chris Parsons, Martin Richardson, Philip Sauré, and Vitor Trindade. We also thank participants of the 27th Annual Meeting of the EEA, the Midwest International Trade Spring Meeting 2013, the 6th International Conference for Migration and Development, the 6th Sinergia-Workshop (University of Zurich) on Economic Inequality and International Trade, the Harvard Graduate Workshop in International Economics, the 12th Australasian Trade Workshop 2017, the ANU Research School of Economics Seminar 2017, and the CESifo Area Conference on the Global Economy 2018 for valuable comments and suggestions. Wilhelm Kohler gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) under grant no. KO 1393/2-1. Inga Heiland gratefully acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement 715147).

1 Introduction

Economic well-being crucially depends on how well workers' abilities are matched with firms' skill requirements. In this paper, we use a general equilibrium model to show that the quality of worker-firm matches in an economy with horizontal skill heterogeneity among workers depends upon how open the economy is to trade and migration. By horizontal skill heterogeneity, we mean that, controlling for skill *levels*, workers have different *types* of skills. We show that under plausible conditions, this leads to monopsony power on the labor market and to imperfect matching between workers and firms. Our central contribution is to demonstrate that international trade and migration have important but distinct consequences on the quality of worker-firm matches as well as on the wage markups and the inequality of wage incomes. Trade increases firms' monopsony power on the labor market while worsening the average quality of worker-firm matches, yet the gains from trade theorem survives. Migration enhances the competitiveness of labor markets and tends to increase the average quality of matches in both countries. We also show that imperfect matching constitutes an incentive for two-way migration between symmetric countries.

Heterogeneity of workers has recently gained substantial attention in trade theory. Grossman and Maggi (2000) and Ohnsorge and Trefler (2007) were the first to analyze, in different model environments, how cross-country differences in the skill distribution among workers shape the pattern of trade in sorting models with skill heterogeneity of workers and submodularity or supermodularity of production techniques. Two-sided heterogeneity coupled with log-supermodularity assumptions are also at the core of a new, general formulation of the theory of comparative advantage developed by Costinot (2009). Costinot and Vogel (2010) develop an application of Costinot (2009) where workers with different skill levels are assigned to tasks (intermediate inputs) differing by skill intensity.[1]

These assignment models of international trade leave two important issues to be addressed. First, two-sided heterogeneity may not be amenable to ranking, in which case the concept of supermodularity cannot be applied. The obvious case in point is horizontal heterogeneity, meaning skill types rather than skill levels. Empirically, horizontal heterogeneity appears no less important than vertical heterogeneity. Consider an architect and a fashion designer who have undergone the same amount of training. They might

[1] A survey of assignment models of this kind in trade is found in Costinot and Vogel (2015)

be regarded as having the same level of skills, yet they obviously embody different types of skills. The same might be said when comparing a bridge building engineer with an architect or with a fashion designer, whereby it is equally obvious that the bridge building engineer embodies a skill type closer to that of an architect than that of a fashion designer. Moreover, the architect might be considered a skill type which is at an equal "distance" to the skill types, respectively, of the bridge building engineer and the fashion designer. These examples illustrate what we mean by horizontal, or skill-type heterogeneity of workers, and they clearly demonstrate that it is a pervasive phenomenon. Clearly, horizontal heterogeneity of workers defies assignment based on log-supermodularity, which is often described as reflecting comparative advantage of workers for different industries, occupations, or tasks.[2] We propose an assignment based on horizontal heterogeneity that may be seen as reflecting absolute advantage.

The second issue is that in assignment models featuring a continuum of tasks as well as a continuum of skills the matching is always *perfect* in that each task is assigned a unique skill level and each skill level is assigned to only one task (see, e.g., Ohnsorge and Trefler, 2007; Costinot and Vogel, 2010). But we almost always observe matches where any one worker is assigned to several tasks. Specifically, in this paper we view a firm as a collection of different tasks, with an associated ideal collection of skills on the part of the worker, which we call the ideal skill type for the firm. Given a continuum of worker heterogeneity in skill types, imperfect matching will then necessarily arise whenever the presence of fixed cost leads to a finite number of firms. By necessity, any firm will employ workers whose skill types deviate, to varying degrees, from its ideal type. However, the collection of tasks that makes a firm is not exogenous. We argue that when entering the market a firm has an incentive to organize its production process in such a way that the collection of tasks required optimally responds to the existing skill-type heterogeneity among workers and to the ideal skill types chosen by other firms.[3] We model the skill-type

[2] The above types of assignment models could, in principle, be reformulated for horizontal skill heterogeneity. In particular, with multi-dimensional heterogeneity it is possible to arrive at skill types that, while horizontal in one sense, may still be ranked in some other sense. This is the case in Ohnsorge and Trefler (2007) where workers who are ranked according to different *relative* skill levels might be considered as different skill types, conditional on having the same *absolute* level of any one of the two skills. But this is a special case. In many cases, horizontal skill heterogeneity defies a clear ranking of skill types.

[3] This is analogous to heterogeneous consumers prompting firms to choose idiosyncratic combination of product characteristics putting their respective products at a distance to products of competing firms; see Vogel (2008). Trade models employing a circular representation of heterogeneity relating to products are found in Lancaster (1980), Helpman (1981), Grossman and Helpman (2005), and Eckel (2009a,b).

matches arising in such an environment as the result of firms entering the market with specific skill-type requirements as well as through workers' self-assignment to firms. And we demonstrate that the quality of this matching depends on trade and migration.

Following Amiti and Pissarides (2005), we use a circular representation of skill-type heterogeneity among workers to model firms' skill-type choices as spatial competition. Thus, we model skill-type heterogeneity of workers in a single dimension, although the underlying abilities surely have many dimensions. The big advantage of this approach is that it allows us to analyze the extent to which a firm employs workers deviating from its ideal skill type; we speak of the "skill reach" of firms. The underlying assumption is that workers with skill types that differ from a certain firm's ideal type don't find all of their skills entirely useless for this firm, but will simply be less productive when working for this firm than workers with skill types closer to the firm's ideal type. Our model thus implies unequal wages within firms, consistent with the empirical results in Becker et al. (2019), who show that within-plant variation of residual wages (controlling for skill levels) accounts for a significant share of overall wage inequality in Germany[4]. Another implication of our labor market model is that firms wanting to increase iemployment will have to reach out to workers with skill types farther away from their ideal types, and closer to other firms' ideal types. This is in line with empirical evidence that workers' skills are transferrable, or portable, across jobs, but imperfectly so, as presented in Gathmann and Schönberg (2010). It is also consistent with evidence from German employer-employee data provided by Gulyas (2018), who finds that the worker-firm match quality is below average for newly hired workers in growing firms.

We derive the firm-specific labor supply function generated by self-sorting of workers based on horizontal skill differentiation, given a firm's ideal skill type. We employ this function in setting up a two-stage game. Stage one involves firm entry, and stage two involves wage- and price setting. Price setting takes place in an environment of monopolistic competition based on translog preferences featuring love-of-variety, which places our model in the Krugman (1979) tradition of modern trade theory.

In a nutshell, our contributions are as follows. First, we describe the general equilibrium of this economy, assuming free entry of firms. This equilibrium features endogenous wage and price markups and an endogenous number of firms. In addition, it describes how

[4] Becker et al. (2019) find that more than half of the observed variance of daily wages in Germany during the period 1996-2014 is due to (unexplained) variation within plants and occupations

welfare depends on the quality of worker-firm matches, which in turn depends on the equilibrium skill reach of firms. In this paper we focus on an equilibrium where firms are symmetrically spread over the skill circle. The working paper version (Heiland and Kohler, 2018) offers a more detailed analysis of firm entry allowing for asymmetric positioning on the skill circle and a proof of existence and uniqueness of the symmetric equilibrium studied here.

Secondly, we apply the model to identify two novel effects of trade that derive from skill-type heterogeneity of workers. The familiar trade-induced gains from scale that are accompanied by firm exit are countervailed by a decline in the average quality of worker-firm matches. Moreover, firm exit reduces competition in the labor market, leading to higher markups between wages and the marginal productivity of workers. However, relative to autarky, we prove that the conventional pro-competitive and variety effects of trade dominate these adverse labor market effects. Hence, the gains from trade theorem survives. But the trade-induced firm exit aggravates inequality, because workers at the bottom end of the income distribution will see their skill type becoming less suitable in production. Moreover, piecemeal trade liberalization involves a non-monotonicity: When the real trade cost gradually falls from a prohibitive level to zero, aggregate welfare is rising (falling) for low (high) initial levels of trade costs.

Thirdly, we demonstrate that any trade-cum-migration equilibrium delivers greater aggregate welfare than an equilibrium with free trade alone, because it lowers wage markups in all countries. But the effects of migration on income inequality and on the average quality of worker-firm matches are ambiguous. Moreover, in contrast to piecemeal integration of goods markets, piecemeal integration of labor markets is unambiguously welfare increasing for all countries.

Our paper provides a novel explanation for a number of salient patterns in global migration data. The first is two-way migration between countries of similar standards of living, which looms large in the data. Docquier and Marfouk (2006) report that in 2000 30% of all migrants residing in OECD countries were born in high-income countries. And to a large extent this migration is two-way in nature. Using bilateral stocks of immigrants and emigrants reported in the OECD's DIOC database, we calculate a correlation coefficient between these stocks (in logs) equal to .69, statistically significant at the 1% level. Admittedly, this correlation may be due to country size and symmetry in bilateral migration barriers. Yet, using residuals obtained from a regression equation explaining bilateral stocks of immigrants and emigrants by origin and destination fixed effects as well as the usual gravity-type covariates, the correlation coefficient remains high (.59) and statistically significant. We find correlation coefficients in the same vicinity also when accounting for the age or education of migrants; see Table A.1 in the appendix. In our model, this type of two-way migration arises because in each country some workers find some foreign firms better suited for their skill types than domestic firms, and there exists a subset of firms finding some foreign workers who are better suited for their skill requirements than domestic ones.

The second pattern relates to the relationship between trade and migration. In models of comparative advantage, trade and migration are usually portrayed as substitutes. Empirically, however, trade and migration flows often reveal patterns of complementarity. A sizeable body of empirical literature (surveyed in Felbermayr et al., 2015) attributes this finding to a trade-cost-reducing effect of migrant networks, and a demand-side effect deriving from migrants' preference for goods produced in their country of origin. Our model proposes a rationale for the opposite direction of causality, where trade-induced firm exit within each country enhances the scope for a mutual improvement of the quality of worker-firm matches through migration.

The remainder of the paper is organized as follows. In Section 2, we briefly discuss the strands of literature that we contribute to with this paper. In Section 3, we describe the general model framework and characterize the autarky equilibrium. In Section 4, we then discuss the effects of a transition from autarky to free trade and the scenario of piecemeal trade liberalization. In Section 5, we introduce labor mobility and analyze the effects of migration, comparing "trade cum migration" with trade alone. Section 6 concludes.

2 Related literature

2.1 Worker heterogeneity and trade

Our paper is related to Ohnsorge and Trefler (2007) and Costinot and Vogel (2010), who develop assignment models of trade based on continuous worker heterogeneity coupled with supermodularity. These models feature perfect competition, and assignment is perfect in that there is a unique, monotonic matching function mapping worker skills to sectors. Our paper expands this stream of literature by considering horizontal skill heterogeneity, which defies supermodularity and generates monopsony power in the labor market. In addition, we allow for increasing returns to scale and imperfect competition on goods markets. Assignment between workers and firms is driven by absolute, rather than comparative, advantage of workers based on their types of skill.[5] The number of firms as well as the quality of worker-firm matches is endogenous. Crucially, both trade and migration affect the average quality of matches as well as the degree of competition in the labor market.[6]

Grossman et al. (2017) analyze wage inequality across skill levels, occupations, and industries in a model featuring matching of heterogeneous workers to heterogeneous managers and sorting of these types into industries. Industries differ in labor intensity, that is, the optimal measure of workers of a given type that are paired with a manager of a given type in equilibrium. Wages vary between workers (managers) due to their inherent skill-level differences and due to the fact that they are matched with differently skilled managers (workers) and different measures of workers of the same type within a firm. Worker types are uniquely matched with manager types and each type is assigned to only one industry. Hence, there is no notion of mismatch and no wage inequality between workers with identical skill levels, which are the essential aspects of our model.

A recent paper by Becker et al. (2019) is notable here for two reasons. First, it provides evidence in favor of significant within-plant variation of residual wages. And second, their explanation for this type of wage variation is similar to our approach in that they use a circular representation of worker heterogeneity. Their circle represents a continuum of tasks, and a point on the circle marks the task a worker is best at, called her core ability. Production is organized in terms of occupations, whereby an occupation specifies a certain subrange of tasks, all of which are performed by each of the workers hired for that particular occupation, independently of their core ability. Wage inequality derives from workers' core abilities matching differently well with the range of tasks to be performed. Intuitively, the wider the range of tasks assigned to an occupation, the lower the quality of the poorest match. The larger the number of occupations chosen by a firm, the better the average occupation-worker match. The analogy to our distance between workers' skill

[5] There is a long strand of literature, originating in Roy (1951), that considers assignment of skill types defined as incorporating certain bundles of different skills that need to be matched with with multi-dimensional skill requirements of jobs; see Mandelbrot (1962); Rosen (1978); Moscarini (2001); Lazear (2009), and Lindenlaub (2017). Welch (1969) compares this approach to the ideal variety approach to consumption theory proposed by Lancaster (1966), and applied to trade in Lancaster (1980).

[6] Imperfect worker-firm matches also arise in Davidson et al. (2008) and Helpman et al. (2010). However, in those models heterogeneity of workers is vertical in nature although assignment is not based on supermodularity.

type and the skill requirements of firms is relatively obvious. A distinctive feature of our analysis is the explicit description of the competition for workers' skill types among firms with similar skill requirements, with trade liberalization unambiguously worsening the average quality of worker-firm matches and increasing the degree of within-firm inequality. Moreover, we expand our focus to the labor market effects of migration.

More broadly, our paper also contributes to a voluminous modern literature on gains from trade in the spirit of Krugman (1979) and Melitz (2003). Arkolakis et al. (2012) have reinvigorated the discussion of gains from trade in these new trade models. Our contribution here is to emphasize two additional welfare channels that derive from horizontal skill heterogeneity among workers through endogenous wage markups as well as through the endogenous quality of worker-firm matching. We assume a translog expenditure function, which is nested in Arkolakis et al. (2019) and implies sub-convex demand, as shown by Mrázová and Neary (2017). Thus, our model falls into the category of recent trade models delivering the familiar pro-competitive effects of trade.

Finally, our paper relates to a small literature dealing with trade in the presence of monopsony power on the labor markets. Generally, one would expect that a monopsonistic labor market distortion adds a potentially important first-order welfare effect of trade and trade policy, alongside the more familiar distortions deriving from monopolistic competition on goods markets. Moreover, firms wielding monopsony power on the labor market is a pervasive empirical phenomenon; see Manning (2003). Yet, there are only a few trade papers allowing for monopsonistic labor markets. In MacKenzie (2019), monopsony power derives from firm-worker-specific productivity draws in a random utility model to pin down the sorting of a continuum of workers into a finite set of heterogeneous firms. Similar to our setting, market power leads to suboptimally small firm size. However, firm heterogeneity adds another important twist: MacKenzie (2019) finds that the size distortion is particularly strong for more productive firms, which are too small in absolute terms and too small compared to the less productive firms. The set of firms, as well as their set of export markets are exogenously given. The welfare effects of trade liberalization unfold through a reallocation of resources towards larger firms, thus mitigating the relative size distortion. In contrast, in our model trade liberalization affects welfare through firm exit, which mitigates the absolute size distortion. Egger et al. (2019) use a similar microfoundation for monopsonistic labor markets that similarly relies on discrete choice based on random utility to explore the implications of monopsony power for exporting and offshoring. As in MacKenzie (2019), firms are heterogeneous, and each worker draws an idiosyncratic random utility component capturing the amenities that the worker draws from working for a specific firm. Upward-sloping firm-specific labor demand generates a novel incentive for offshoring, as this allows firms to separately exploit monopsony power on the domestic and the foreign labor market. Both, exporting and offshoring increase firm size, thus ameliorating the size distortion deriving from monopsonistic labor market, but whereas exporting increases the domestic wage, offshoring comes with a reduction of domestic wages. Similar to our explanation of two-way migration, Egger et al. (2019) provide an explanation of two-way offshoring. In contrast to these random utility models, our explanation of monopsony power by firms relies on horizontal skill differentiation among workers, which allows us to model endogenous skill-type location of entering firms.

2.2 Two-way migration

Trade models highlighting endowment-based comparative advantage imply that trade and migration are substitutes, but if trade is driven by other forces they may be complements, as first emphasized by Markusen (1983). Models featuring two-way migration typically rest on assumptions regarding differences in technologies or endowments, as in Iranzo and Peri (2009).[7] Two-way migration among similar countries has proven harder to explain. Much of the recent migration literature on sorting across destinations has invoked location-specific preferences to rationalize these flows; see, for example, Tabuchi and Thisse (2002) and the strand of empirical literature building on Grogger and Hanson (2011). Caliendo et al. (2019) develop a general equilibrium model with trade and migration involving exogenous location preferences. More specific theoretical models derive two-way migration incentives between similar countries from social stigma attached to employment in low social status occupations, as in Fan and Stark (2011), or from migration serving as signaling device for high skilled individuals when skills are unobservable, as in Kreickemeier and Wrona (2017). Our two-way migration incentive between symmetric countries derives from the firm-specificity of skills and firms' endogenous location choices in the skill space, and it generates migration flows that are complementary to the volume of trade.

[7] In Schmitt and Soubeyran (2006), two-way migration arises within occupations but across skill levels, and only if countries' skill distributions are sufficiently different. In Galor (1986), individuals differ with respect to the rate of time preference and two-way migration appears among countries with different interest rates. Gaumont and Mesnard (2000) show that differences in relative factor prices might also lead to two-way migration when individuals are heterogeneous with respect to the degree of altruism.

3 The modeling framework

Our model economy is endowed with a mass L of workers, differentiated by the *types* of skills they possess. Following Amiti and Pissarides (2005), we use a circle to characterize the relationship between skill types. Each location on the circle represents a skill type, and types that are more similar are located closer to each other.[8] The circumference of this circle measures the degree of horizontal skill differentiation present in the labor force. We use H to denote the half circumference and assume that workers are uniformly distributed over this circle. Firms decide in two stages. In *stage one*, they decide on whether to enter and set up production of a differentiated variety requiring one particular kind of skill corresponding to a specific location on the skill circle. In line with the vast majority of the related literature, we consider only symmetric location equilibria where all firms' optimal varieties are located at equal distances on the skill circle.[9] In *stage two*, firms pursue Bertrand strategies in setting goods prices as well as wage rates. The resulting Bertrand-Nash equilibrium is thus conditional on the number firms determined in stage one. When deciding on entry in stage one, firms anticipate the Bertrand-Nash equilibrium of stage two (subgame perfection). We assume free entry of an infinite number of potential entrepreneurs with zero outside options. Hence, equilibrium is characterized by zero profits.

3.1 Price and wage setting with worker heterogeneity

The more a worker's skills deviate from a firm's ideal skill type the less productive she is when working for this firm. The function $f[d]$ gives the number of efficiency units of labor delivered per physical unit of labor by a worker whose skills are at arc-distance d from the firm's ideal type.[10] We assume that $f'[d] < 0$, $f'[0] = 0$, $f''[d] < 0$, and $f[d] = f[-d]$. This last property states that distance in either direction on the circle has the same effect. Without loss of generality, we set $f[0] = 1$.

---

[8] Our model differs from Amiti and Pissarides (2005) by featuring a more general form of the relationship between skill distance and productivity. Moreover, we introduce endogenous price markups, and labor mobility across countries. These features allow us to study the diverse effects of labor market and product market integration.

[9] In Heiland and Kohler (2018) we prove existence and uniqueness of the symmetric location equilibrium in this model under a set of restrictions on parameters and beliefs. We do so by explicitly modeling an entry game where entry involves ideal-skill locations of firms, allowing for asymmetric location patterns.

[10] We use brackets $[\cdot]$ to collect arguments of a function and parentheses to collect algebraic expressions.

We assume enforceable contracts between firms and workers, specifying the quantity of, and price for, efficiency units of labor. Each worker inelastically supplies one unit of physical labor and knows her skill distance from all firms positioned on the skill circle as well as the productivity schedule $f[d]$. In contrast, firms are unable to observe individual workers' abilities, paying the same wage per efficiency unit for all workers. This informational assumption negates wage discrimination among workers, with the entire job surplus accruing to the worker. All workers sort themselves into different firms so as to maximize their individual incomes, given firm-specific wage offers as well as their skill distance to these firms. As a result, each firm faces an upward-sloping labor supply and, therefore, has wage setting power. This type of self selection, or assignment of workers to firms, is based on absolute advantage, as opposed to the comparative advantage assignments considered in Ohnsorge and Trefler (2007) and Costinot and Vogel (2010). Specifically, comparing any two workers working for different firms, it will always be true that either worker is more productive in the firm she is working for than the other worker would be if both were to work in the same firm.

When firms set wages and prices, the number of firms and their skill location on the circle is given. We use m to denote the half arc-distance between firm i's position and the nearest neighbor to its left and to its right. The distance between firms on the skill circle is inversely related to the number of firms through $N = \frac{H}{m}$. Firm i posting a wage rate w_i per efficiency unit of labor will attract workers up to a skill distance d_r to its right, with d_r defined by $w_i f[d_r] = w_{i+1} f[2m - d_r]$ where w_{i+1} denotes the wage rate posted by the nearest neighbor to the right, firm $i+1$. Obviously, for any given w_{i+1}, the distance d_r is increasing in w_i. For $w_i > w_{i+1}$, firm i attracts workers who would in fact be better matches for firm $i+1$; workers' absolute advantage is then overruled by wage asymmetries. Analogously, we use d_ℓ to define the distance of the worker to the left of firm i who is indifferent between working for firm i and firm $i-1$. Given a uniform distribution of workers along the skill circle, we obtain firm i's labor supply schedule as

$$L^S[w_i, w_{i-1}, w_{i+1}, m] = \int_0^{d_\ell[w_i, w_{i-1}, m]} f[d] \frac{L}{2H} \mathrm{d}d + \int_0^{d_r[w_i, w_{i+1}, m_i]} f[d] \frac{L}{2H} \mathrm{d}d. \tag{1}$$

This labor supply function is increasing in the firm's own wage.[11] In what follows, we shall

[11] For simplicity, we limit the description of the labor supply function to ranges of wage rates implying positive labor supply for every firm. For sufficiently large relative wage differences some firms' labor

use η_i to denote the wage elasticity of firm i's labor supply. Obviously, η_i is a function of w_i, the neighbors' wage rates w_{i-1}, w_{i+1}, and m.

We assume that preferences may be described by a symmetric translog expenditure function. This expenditure system corresponds to homothetic preferences and features a variable demand elasticity. Both properties are desirable for our study of the welfare effects of globalization on heterogenous workers.[12] Aggregate demand for variety i is

$$q_i[\boldsymbol{p}, Y] = \delta_i \frac{Y}{p_i} \quad \text{with} \quad \delta_i = \frac{1}{N} + \gamma \left(\overline{\ln p} - \ln p_i \right), \tag{2}$$

where p_i is the price of variety i, $\overline{\ln p}$ denotes the average log price across all firms, and Y denotes aggregate income. The parameter $\gamma > 0$ measures the degree of substitutability between varieties, a larger γ implying higher substitutability. Revenue r_i then follows as

$$r_i = \delta_i Y \qquad \text{with} \qquad \delta_i = \gamma \mathcal{W} \left[\exp \left\{ \frac{1}{\gamma N} + \overline{\ln p} \right\} \frac{q_i}{\gamma Y} \right], \tag{3}$$

where $\mathcal{W}[\cdot]$ denotes the *Lambert function*.[13] While (2) expresses the expenditure share as a function of $\ln p_i$, in (3) this share is expressed as a function of the quantity q_i; Appendix A.2 has the details. Given consumers' love for variety, no two firms will produce the same variety, so that we may use i to indicate both firms and varieties.

Firms' production technology is identical and characterized by a constant marginal cost β and a fixed cost α, defined in efficiency units of firm-specific labor. Armed with these representations of goods demand and labor supply, firm behavior in stage two may now be characterized by the following profit maximization problem:

$$\max_{w_i} \; r_i - w_i L_i \qquad \text{s.t.} \qquad q_i = \frac{L_i - \alpha}{\beta} \quad \text{and} \quad q_i \geq 0, \tag{4}$$

where r_i is given by (3) and $L_i = L^S [w_i, w_{i-1}, w_{i+1}, m]$. The restriction ensures that the

---

supply would fall to zero. Such strategies, however, cannot occur in the two-stage equilibrium with free entry. Off-equilibrium strategies in this game are described in Heiland and Kohler (2018).

[12] Neither constant elasticity of substitution utility nor quadratic utility, the two most commonly used preference assumption in the literature on gains from trade, exhibit both properties. 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.

[13] The Lambert function $\mathcal{W}[z]$ defines the implicit solution to $xe^x = z$ for $z > 0$. Furthermore, it satisfies $\mathcal{W}'[z] = \frac{\mathcal{W}[z]}{(\mathcal{W}[z]+1)z} > 0$, $\mathcal{W}''[z] < 0$, $\mathcal{W}[0] = 0$ and $\mathcal{W}[e] = 1$.

firm is on its labor supply function and produces a positive quantity. We proceed under the assumption that the non-negativity constraint is non-binding. The corresponding restrictions on the parameter space are discussed in Appendix A.4. Note that the problem (4) is conditional on m and thus on the number of firms, which is determined in stage one to be discussed below.

We assume that firms pursue Bertrand strategies on both the goods and the labor market, taking the average log price $\overline{\ln p}$ and aggregate income Y as being beyond their own influence. The perceived price elasticity of demand for variety i emerges as

$$\varepsilon_i[p_i, \overline{\ln p}, N] := -\frac{\mathrm{d} \ln q_i}{\mathrm{d} \ln p_i} = 1 - \frac{\mathrm{d} \ln \delta_i}{\mathrm{d} \ln p_i} = 1 + \frac{\gamma}{\delta_i} > 0, \tag{5}$$

where δ_i is given in (3). From the first-order conditions, pricing involves a double markup over marginal cost:[14]

$$p_i = \frac{\varepsilon_i}{\varepsilon_i - 1} \frac{\eta_i + 1}{\eta_i} w_i \beta. \tag{6}$$

From (5) and (2), we may write the first markup which derives from product differentiation as

$$\frac{\varepsilon_i}{\varepsilon_i - 1} = 1 + \frac{\delta_i}{\gamma} = \mathcal{W}\left[\frac{\eta_i}{w_i(\eta_i + 1)} \exp\left\{1 + \frac{1}{\gamma N} + \overline{\ln p}\right\}\right]. \tag{7}$$

See Appendix A.2 for details. In this equation, the argument of the Lambert function $\mathcal{W}$ is a "summary measure" of the conditions that firm i faces on the labor market as well as the goods market. Since $\mathcal{W}'[Z] > 0$, a higher average log price and a lower degree of substitutability γ both lead to a higher markup. The same holds true for a smaller number of firms, whereas the markup is falling in perceived marginal cost. The second markup in (6) derives from the firm's monopsony power on the labor market, where the firm faces a finite elasticity of supply η_i. Combining markup pricing with the constraint in (4) gives rise to a best-response function $w_i = w\left[w_{i-1}, w_{i+1}, m, \overline{\ln p}, Y\right]$. Equilibrium wages then follow as the fixed point of all firms' best-response functions.

In the symmetric second-stage equilibrium, we have $p_i = p$, with $\overline{\ln p} = \ln p$, as well as $w_i = w$. Symmetry in (2) simplifies the expressions for ε and δ, allowing us to write the

[14] It is easy to verify that under the assumptions made, the second-order condition is satisfied.

profit-maximizing price (6) as

$$p[m] = \rho[m]\psi[m]\beta, \tag{8}$$

$$\text{where} \quad \rho[m] := 1 + \frac{1}{\gamma N[m]} \quad \text{and} \quad \psi[m] := \frac{\eta[m]+1}{\eta[m]}. \tag{9}$$

In (8), we have normalized the wage per efficiency unit to one.[15] Note that $\rho'[m] > 0$ as well as $\psi'[m] > 0$. Firms' market power in the either market increases as firms become larger and the number of firms falls. The elasticity of labor supply in (1) evaluated at symmetric wages, may be written as

$$\eta[m] := \left.\frac{\partial L_i^S}{\partial w_i}\frac{w_i}{L^S}\right|_{w_i=w} = -\frac{f[m]^2}{2F[m]f'[m]}, \tag{10}$$

where $F[m] := \int_0^m f[d]\mathrm{d}d$. Our assumption that $f''[m] < 0$ ensures that the labor supply elasticity is falling in m. Given a uniform distribution of the workforce around the circle, the average productivity of workers is

$$\theta[m] = \frac{1}{m}\int_0^m f[d]\mathrm{d}d. \tag{11}$$

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. Aggregate income emerges as $Y = L\theta[m]$.

This model allows for a very straightforward discussion of income inequality. In a symmetric equilibrium all inequality is within-firm inequality of wage income among workers. Given the uniform distribution of workers over the skill circle, a plausible measure of within-firm inequality is the equilibrium skill reach m. The skill reach is monotonically related to the wage income gap between the incomes of the best matched worker and the worst matched worker, $f[0] - f[m]$. Moreover, we show in Appendix A.3 that the variance of incomes across workers also increases in m, thanks to the concavity of $f[\cdot]$.

[15] We are free to do so, since our equilibrium is homogeneous of degree zero in nominal prices.

3.2 Symmetric two-stage equilibrium

We now proceed in determining the equilibrium skill distance between firms by assuming free entry and imposing a zero profit condition. Entry in the first stage of the game is guided by the zero-profit condition requiring prices as given in (8) to equal average cost:

$$p[m] = \frac{\alpha + \beta q[m]}{q[m]}. \tag{12}$$

Without loss of generality, we may choose units such that $\beta = 1$. Substituting the labor market clearing condition[16] $N[m]q[m] + \alpha N[m] = L\theta[m]$ into (12), we obtain the following representation of the zero-profit condition:

$$p[m] = g[m] := \frac{L\theta[m]}{L\theta[m] - \alpha N[m]}, \qquad \text{with} \qquad g'[m] < 0. \tag{13}$$

Intuitively, an increase in firm size coupled with a higher distance between firms leads to lower average cost. Combining the zero-profit condition (13) with the Bertrand pricing equation in (8), we finally arrive at the following condition determining m:

$$g[m] = \rho[m]\psi[m]. \tag{14}$$

Appendix A.4 characterizes the parameter space under which a symmetric equilibrium described by 14 exists and is unique. If we let H, the degree of skill differentiation, converge to zero, then this equilibrium converges to the equilibrium in the model considered by Krugman (1979) for the special case of translog preferences; see Appendix A.5.

When deriving welfare results below, we take an ex ante view, assuming that workers regard each point on the circle as being equally likely to become an ideal type for some firm. The expected income per physical unit of labor is then equal to $\theta[m]$. Hence, in a symmetric equilibrium, expected log utility of a worker is equal to

$$\ln V = \ln \theta[m] - \ln P, \tag{15}$$

where $\ln P = 1/(2\gamma N[m]) + \ln p[m]$ is the translog unit expenditure function for our symmetric equilibrium. Intuitively, this welfare measure is rising in income and the number of firms in the market while falling in the price of a typical variety of goods.

[16] Goods market equilibrium follows by Walras' law.

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 the "consumer-surplus distortion." (ii) Potential entrants ignore the positive effect on average productivity arising from a better quality of matches in the labor market. This is novel to the literature, and we call it the "productivity distortion." (iii) Potential entrants anticipate a goods price markup as well as a wage markup, but fail to see that they receive operating profits on such markups only at the expense of incumbent firms, due to the overall resource constraint. Mankiw and Whinston (1986) have called this the "business-stealing" effect. (iv) Potential entrants fail to anticipate that their entry will reduce the magnitudes of these same markups, due to enhanced competition.

Distortions (i) and (ii) constitute positive externalities, working towards insufficient entry in a laissez faire equilibrium, while distortions (iii) and (iv) work 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.6 we show that in this model the net result of distortions (i)-(iv) is excess entry. Thus, our model inherits the excess-entry result established by Salop (1979) for the circular city model. 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. The excess-entry result plays a crucial role in the determination of the gains from globalization below, since these partly unfold through a mitigation of distortions.

4 Symmetric trading equilibrium

In this section, we explore the gains from trade as well as the effect of trade on income inequality. The first subsection compares autarky with free trade, where we introduce trade simply by allowing for the number of countries to increase beyond one (which is autarky) and allowing for firms in all countries to sell on all national markets without any border frictions. Mrázová and Neary (2014) call this the extensive margin of globalization. In the second subsection we then turn to the intensive margin of globalization by holding the number of countries fixed at two, but allowing for trade to be costly, and by looking at a marginal reduction of the trading cost. We assume countries to be fully symmetric, including the extent of worker heterogeneity, so as to isolate the channels that emanate from horizontal worker heterogeneity as such.

4.1 Free trade

We assume that there are k symmetric countries and denote the total number of firms worldwide by $N^T = kN$. Absent all barriers, prices for domestic and imported goods are equal, given by

$$p[m] = \left(1 + \frac{1}{\gamma k N[m]}\right)\psi[m]. \tag{16}$$

This expression reflects the fact that firms now take into account foreign competitors, but it keeps the simplified form familiar from the autarky equilibrium; see (9). Absent all trade barriers, prices of imported and domestic varieties are fully symmetric, whence the price of any variety consumed is equal to the average price. In what follows, we define $\rho^T[m] := 1 + \frac{1}{k\gamma N[m]}$ as the goods price markup under free trade. It is obvious that $\rho^T[m] < \rho[m]$.

Equilibrium output per firm as a function of m remains unchanged, since the lower domestic demand is compensated by the larger number of countries: $q[m] = \frac{kL\theta[m]}{kN[m]p[m]} = \frac{L\theta[m]}{N[m]p[m]}$. The labor market clearing condition similarly remains unaffected. The equilibrium condition determining m then follows as

$$g[m] = \rho^T[m]\psi[m]. \tag{17}$$

The following proposition summarizes the comparison between autarky, $k = 1$, and free trade among $k > 1$ countries.

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

Proof: A formal proof is relegated to Appendix A.7.1.

The increase in variety and the pro-competitive effect on the goods market are standard results in 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 on the goods market dominates. 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. 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. The positive welfare effect involves two channels. The first runs through higher variety and lower goods prices. In addition, there is a positive first-order effect deriving from firm exit since the autarky equilibrium features excess entry. On account of $f'[m] < 0$ exit of some firms reduces the lower bound of wages. Since the upper bound of wages is fixed at $f[0] = 1$, and given a uniform distribution of workers over the skill circle, this entails an unambiguous increase in income inequality.

4.2 Costly trade and piecemeal trade liberalization

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 selling q_i units on the domestic market and q_i^* units on the export market then needs a labor input equal to $\alpha + q_i + \tau q_i^*$. We assume that markets are segmented, so that firms set market-specific quantities. The firm maximizes profits with respect to the wage, which determines its labor supply and hence total output $\bar{q}_i = q_i + \tau q_i^*$, and with respect to the quantity sold on the domestic market, observing $q_i^* = \frac{1}{\tau}(\bar{q}_i - q_i)$.

$$\max_{w_i, q_i} \{\delta_i Y + \delta_i^* Y^* - w_i(\alpha + \bar{q}_i)\} \tag{18}$$

$$\text{s.t.:} \quad \bar{q}_i = q_i + \tau q_i^* \quad \text{with} \quad q_i \geq 0,\ q_i^* \geq 0 \quad \text{and} \quad \alpha + \bar{q}_i = L^S[w_i, \boldsymbol{w}_{-i}, \boldsymbol{m}_i],$$

where the domestic expenditure share falling on domestic good i is

$$\delta_i = \frac{1}{N^T} + \gamma \left(\overline{\ln p} - \ln p_i\right) = \gamma \mathcal{W} \left[\exp\left\{ \frac{1}{\gamma N^T} + \overline{\ln p} \right\} \frac{q_i}{\gamma Y} \right]. \tag{19}$$

In this expression, $\overline{\ln p} = \frac{1}{N^T}\left(\sum_{j=1}^{N}\ln p_j + \sum_{j^*=1}^{N^*}\ln p_{j^*}\right)$ denotes the average log price. A perfectly analogous expression holds for δ_i^*, the share of foreign expenditure falling on domestic good i.[17] Due to symmetry, $N = N^*$ and the average log price is the same across markets. The first-order condition with respect to q_i requires that marginal revenue be equalized across markets, whence $p_i\left(\frac{\varepsilon-1}{\varepsilon}\right) = \frac{p_i^*}{\tau}\left(\frac{\varepsilon^*-1}{\varepsilon^*}\right)$. The first-order condition with respect to w_i requires that marginal revenue is equal to perceived marginal cost; see Appendix A.7.2 for details. A symmetric equilibrium with both wages normalized to unity then implies the following optimal pricing conditions:

$$p = \left(1 + \frac{\delta}{\gamma}\right)\psi[m] \qquad \text{and} \qquad p^* = \left(1 + \frac{\delta^*}{\gamma}\right)\psi[m]\tau. \tag{20}$$

The labor market clearing condition is

$$N[m]\left(\alpha + q[p, p^*, m] + \tau q^*[p, p^*, m]\right) = L\theta[m]. \tag{21}$$

In contrast to the autarky and the free-trade case, the pricing conditions cannot be simplified further because individual firms' prices in (19) 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 (20) and (21). This system is the analogue to the free-trade equilibrium condition (17) above.

Our preferences imply the existence of a finite prohibitive level of the trade cost. We denote this prohibitive level by $\bar{\tau}$, and it is implicitly determined by $\delta_i^* = 0$. Note that with $\delta_i^* = 0$ the price elasticity of demand for foreign goods becomes infinite; see (5). Note also that high values of γ imply low values of $\bar{\tau}$. We can now state the following proposition on piecemeal trade liberalization.

Proposition 2. *For two identical countries in a trading equilibrium, a decrease in trade costs 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) Welfare rises for sufficiently low initial levels of τ, but it falls for sufficiently high initial levels of τ. (iv) Income inequality increases.*

[17] Expenditure shares are obtained by differentiation of the log expenditure function, i.e. $\delta_i := \frac{\partial \ln P}{\partial \ln p_i}$ and $\delta_i^* := \frac{\partial \ln P}{\partial \ln p_i^*}$, and then applying the same logic as outlined in Appendix A.2 to express them in terms of q_i and q_i^*, respectively.

Proof: A formal proof is relegated to Appendix A.7.3. To obtain an intuition for this proposition consider the following decomposition of the welfare differential:

$$\mathrm{d}\ln V = \left(\frac{\partial \ln \theta}{\partial \ln m} - \frac{\partial \ln P}{\partial \ln m}\right)\mathrm{d}\ln m - N\delta\,\mathrm{d}\ln p - N\delta^{*}\mathrm{d}\ln p^{*}. \tag{22}$$

Lower trade cost means lower prices of imported goods, which also lowers the price markups for domestic goods. At the margin $\tau = \bar{\tau}$ there are no imports to start with, whence lower prices for imported goods have no first-order welfare effect. Obviously, this effect increases as the initial level of the trade cost becomes smaller. However, trade liberalization lowers domestic markups even at the margin $\tau = \bar{\tau}$. In fact, this effect is independent of the initial level of τ, working through $\overline{\ln p}$ in (18) even if $\delta^{*} = 0$.

Goods prices are also affected by the wage markup which depends on the skill distance between firms through $\psi[m]$. In addition, the skill distance determines the average productivity of workers through $\theta[m]$. Trade liberalization leads to higher firm output $\bar{q}$, which in turn implies firm exit, given the resource constraint L. A lower number of firms lowers welfare on three accounts: it lowers average productivity $\theta[m]$, it lowers the number of varieties, and it contributes to higher domestic goods prices through a higher wage markup. Crucially, this firm exit effect is magnified by a higher resource use from a higher export volume. The output equivalent of this resource use effect is equal to $(\tau - 1)\mathrm{d}q^{*}$. Thus, for any given increase in exports, the magnification of the firm exit effect is largest at the prohibitive level of $\tau = \bar{\tau}$ and it converges to zero as we move to free trade, $\tau \to 1$.

Proposition 2 tells us that at the margin $\tau = \bar{\tau}$ the higher wage markup dominates the lower goods price markup from lower import prices, which together with the magnified negative effects from firm exit implies a negative welfare effect. As the initial level of τ converges to unity (free trade), the resource use effect from higher exports converges to zero while lower import prices have a large first-order welfare effect, implying a positive welfare effect of trade liberalization, consistently with Proposition 1. Moreover, we know from Proposition 1 that free trade is better than autarky, hence there is a threshold value of $\tilde{\tau} < \bar{\tau}$ such that trade, though costly, is unambiguously better than autarky provided that $\tau < \tilde{\tau}$.[18]

[18] This U-shaped welfare curve does not hinge upon worker heterogeneity in the labor market, nor on translog preferences. Bykadorov et al. (2016) demonstrate that welfare losses for high initial trade cost obtain in any model of monopolistic competition with an additive utility function featuring a variable

The unambiguous relationship between trade and the economy-wide level of wage inequality established in Propositions 1 and 2 hinges on symmetry of firms. In a more general setting with firm-heterogeneity as in Melitz (2003), more productive (and larger) firms have a higher skill reach, which implies a higher degree of internal dispersion of wage incomes among workers than in less productive (and smaller) firms. Firm exit following a change from autarky to trade will be concentrated on the least productive firms, which implies a reallocation of workers to larger firms with a higher internal dispersion of worker incomes, which contributes to economy-wide income inequality. The inequality effect of piecemeal trade liberalization is somewhat less straightforward in the presence of firm heterogeneity. Like the move from autarky to trade, piecemeal liberalization leads to a reallocation of labor from the exiting firms with low productivity and low internal dispersion of worker incomes to more productive firms with higher internal income dispersion, contributing to aggregate income inequality. In addition, there is reallocation of labor towards the newly exporting firms, at the expense of all firms. In other words, the new exporters are the firms with the largest increase in employment. If the degree of internal income dispersion in these firms is high compared to the overall average dispersion, then the inequality effect of firm exit is reinforced, and the economy-wide income inequality is unambiguously rising in the wake of trade liberalization. This condition is met if the initial export-cut-off level of productivity is already high.

5 Migration

Consider two perfectly symmetric countries in a free-trade equilibrium of the type characterized above, symmetry also meaning that the labor force in both countries is distributed over the exact same skill circle. Interpreting average wage income as expected wage income for potential migrants, there is no incentive for international migration based on comparison of *average* incomes. However, except for an unlikely knife-edge case, in both countries some workers will find firms in the other country which provide a better match for their skill type than their present firm.[19] Hence, opening up the two labor markets will lead to a new sorting of workers between firms, domestic and foreign. We demonstrate that this leads to two-way migration which is gainful for both countries. We proceed in

elasticity of substitution and a finite prohibitive real trade cost.

[19] The knife edge case features firms in both countries positioned on identical points on the skill circle.

two steps. First, we describe how opening up labor markets to migration affects a firm's labor supply and, therefore, its wage setting, conditional the number of firms in both countries and thus the skill distance between firms. In the second step we endogenize the skill distance by invoking free entry of firms analyzing a zero-profit equilibrium with free and costless trade, but allowing for migration costs to vary continuously from prohibitive levels all the way down to zero.

5.1 Labor supply with integrated labor markets

We model the barriers to migration as reducing the productivity of a worker to a fraction $1-\lambda$ if she moves to the other country. When working for a domestic firm at distance d, a domestic worker delivers $f[d]$ efficiency units while delivering only $f[d](1-\lambda)$ efficiency units when working for a foreign firm at the same skill distance d.[20]

We start our analysis by examining the labor market environment of firms in a situation with open labor markets, conditional on some given symmetric location pattern where the distance between any two neighboring firms from the same country is equal to $2m$, and where any one firm faces two neighboring firms from the other country, with the same distance equal to m. We call this an *alternating location pattern*. Suppose that a representative domestic firm posts a wage rate w, with the neighboring foreign firm posting a wage rate w^*. In Figure 1, the domestic firm is located at s_0 on the skill circle, and we use d^n to denote the skill reach of the domestic firm for native workers to its right and its left, and d^m to denote this firm's skill reach for migrants from the other country, again symmetrically in both directions. The two skill reaches are determined by the following conditions:

$$wf[d^n] = w^* f[m-d^n](1-\lambda) \qquad \text{and} \qquad wf[d^m] = w^* f[m-d^m]\frac{1}{1-\lambda}. \tag{23}$$

As migration barriers fall, the two skill reaches converge; with $\lambda = 0$ they coincide at

[20] With positive trade cost, a migration equilibrium like the one considered here would potentially be subject to instability, since by moving one way into one of the two countries workers could avoid all trade costs, and this gain might potentially outweigh the migration cost. Assuming a zero trade cost equilibrium to start with, any positive cost of migration makes our equilibrium immune to this type of agglomeration force. In a world with zero costs of both trade and migration the equilibrium outcomes in the dispersed and the agglomeration equilibrium are the same in terms of prices and welfare. The proportionality assumption for the cost of migration is convenient but not crucial. A general characterization of the specifications generating the results derived in this section is found in Appendix A.8.6.

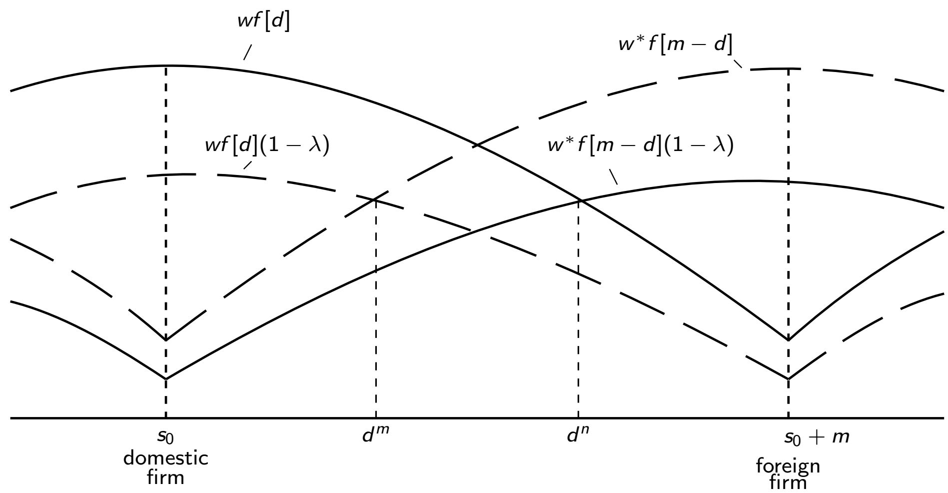

Figure 1: Sorting of workers with migration

$m/2$. The domestic firm at s_0 employs domestic workers with skill types in the interval $(s_0 - d^n, s_0 + d^n)$, and foreign workers (migrants) located in the interval $(s_0 - d^m, s_0 + d^m)$, while the foreign firm located at $s_0 + m$ employs foreign workers located in the interval $(s_0 + m - d^n, s_0 + m + d^n)$ and domestic workers (migrants) with skill types in the interval $(s_0 + m - d^m, s_0 + m + d^m)$.

The labor supply as a function of the firm's own wage now emerges as

$$L^{S,M}[w, w^*, m, \lambda] = \frac{L}{H}\left(\int\limits_0^{d^n[w,w^*,m,\lambda]} f[d]\mathrm{d}d + \int\limits_0^{d^m[w,w^*,m,\lambda]} f[d](1-\lambda)\mathrm{d}d\right), \tag{24}$$

where a superscript M indicates the case of migration, as opposed to closed labor markets. In the above equation, $d^n[w, w^*, m, \lambda]$ and $d^m[w, w^*, m, \lambda]$ are implicitly determined by (23). Symmetry across countries with $w = w^*$ implies $d^m = d^m[m, \lambda] = m - d^n[m, \lambda]$. The degree of inequality measured by the income gap between the incomes of the best-matched worker and the worst-matched worker is monotonically increasing in d^n, the skill reach for native workers. The indifference conditions (23) imply that the income of the worst-matched native at skill distance d^n equals the income of the worst-matched migrant at d^m, net of the migration cost. Hence, when determining inequality effects below, it will be enough to explore $d^n[m, \lambda]$.

By complete analogy to (10), the perceived elasticity of labor supply, evaluated at the symmetric equilibrium, can be derived as $\eta^M[m, \lambda] = \frac{\partial L^{S,M}}{\partial w}\frac{w}{L^{S,M}}\Big|_{w=w*=1}$ from (24). Details are found in Appendix A.8.1. As migration becomes less costly, each firm perceives a lower labor supply elasticity $\eta^M[m,\lambda]$, whence the labor market becomes more competitive.[21] By analogy to (9), we now use $\psi^M[m,\lambda] := \big(\eta^M[m,\lambda]+1\big)\big/\,\eta^M[m,\lambda]$ to denote the wage markup under migration. For a given level of m, the markup is unambiguously lower with migration than without. Interestingly, even if migration barriers are prohibitive, firm behavior is still influenced by *potential* migration. Let $\bar{\lambda}$ denote the prohibitive level of migration barriers, determined by setting $d^m[m,\lambda]=0$. With open labor markets but $\lambda=\bar{\lambda}$, firms do not employ any foreign workers, but setting a higher wage would now attract foreign workers in addition to domestic ones, so each firm's labor supply is more elastic than with closed labor markets. The perceived wage elasticity of labor supply evaluated at $\bar{\lambda}$ is given by

$$\eta^M[m,\bar{\lambda}] = -\frac{2f[m]^2}{f'[m]+(1-\bar{\lambda})f'[0]}\frac{1}{F[m]}. \tag{25}$$

It is obvious that our assumption of $f'[0]=0$ is sufficient to ensure that $\eta^M[m,\bar{\lambda}]$ is larger than the elasticity of supply under autarky as given in (10).

Migration also affects the average quality of skill matches between workers and firms. In a symmetric equilibrium, the average productivity of workers emerges as

$$\theta^M[m,\lambda] := \frac{1}{m}\left(\int_0^{d^n[m,\lambda]} f[d]\mathrm{d}d + \int_0^{d^m[m,\lambda]} f[d](1-\lambda)\mathrm{d}d\right). \tag{26}$$

For prohibitively high migration barriers, $\lambda=\bar{\lambda}$, the average matching quality as given in (26) is the same function of m as under autarky, given in (11): $\theta^M[m,\bar{\lambda}]=\theta[m]$. Moreover, as shown in Appendix A.8.4, θ^M is falling in λ, 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 barriers to migration, holding m constant. For

[21] Strictly speaking this requires that $f'''[d]$ is not too large (in absolute terms). 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 skill reaches 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 skill reach moves to the right.

$\lambda = 0$, we have

$$L^{S,M} = 2\frac{L}{H}\int_0^{\frac{m}{2}} f[d]\mathrm{d}d = \frac{2L}{N^M}\theta^M[m,0] = \frac{L}{N}\theta\left[m/2\right]. \tag{27}$$

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 mass of workers are doubled. However, we know from above that for $\lambda < \bar{\lambda}$ we have $\theta[m,\lambda] > \theta[m]$. Hence, firms face a larger supply of efficiency units of labor with migration than under national labor markets. While employing the same mass of workers in either case, with migration each firm finds workers with skills closer to its optimal type; the skill reach has fallen to $m/2$, compared to m in the case of closed labor markets. The productivity gains from migration, $\theta^M[m,\lambda] - \theta[m]$, are increasing in m; see Appendix A.8.2 for a proof. This implies that trade and migration are complements in the sense that firm exit brought along by trade (cp. Propositions 1 and 2) enhances the productivity gains from migration.

To summarize this first step of our analysis: Opening up labor markets in an alternating pattern of location conditional on m has two effects: First, each firm observes a higher effective labor supply due to better matches, and this efficiency gain increases as the cost of migration falls. And secondly, each firm perceives a more elastic labor supply, whence the labor market is now more competitive.

5.2 Equilibrium with open labor markets

We now turn to the determination of m in a symmetric equilibrium with free trade and open labor markets and $\lambda \in [0, \bar{\lambda}]$. Based on the assumption of free entry, the equilibrium condition requires zero profits. By complete analogy to (17), we obtain

$$g^M[m,\lambda] = \rho^T[m]\psi^M[m,\lambda]. \tag{28}$$

In this equation, $\rho^T[m] = 1 + 1/(\gamma N^M)$ denotes the free-trade price markup over perceived marginal cost, where N^M is the world-wide number of firms in this trade-cum-migration equilibrium; see (16)[22] Unlike the wage markup $\psi^M[m,\lambda]$, the price markup is affected

[22] Throughout our analysis of migration, we use a superscript T to denote functions that take the same form under migration and free trade alone, while using a superscript M to denote functions that are fundamentally different under integrated labor markets compared with free trade alone.

by migration only through the number of firms. As before, the term $g^M[m,\lambda]$ measures average cost, taking into account the labor market clearing condition, which now reads as $\alpha + q = (m/H)L\theta^M[m,\lambda]$. This measure thus reads as

$$g^M[m,\lambda] = \frac{L\theta^M[m,\lambda]}{L\theta^M[m,\lambda] - \alpha H/m}. \tag{29}$$

An equilibrium as described by (28) hinges on the condition $d^m[m,\lambda] \geq 0$. We show in Appendix A.8.5 that there is a unique value $\bar{\lambda}$ solving (28) and $d^m[m,\lambda] = 0$ and that $d^m[m,\lambda] > 0$ for $\lambda \in [0,\bar{\lambda})$. Hence, under analogous restrictions on the parameter space as discussed in Appendix A.4, there exists a unique symmetric equilibrium described by (28) featuring a positive number of firms for $\lambda \in [0,\bar{\lambda}]$.

Proposition 3. *For two symmetric countries with free trade and open labor markets, an equilibrium in the two-stage game with an alternating pattern of firm locations has the following properties: (i) With prohibitive migration barriers, $\lambda = \bar{\lambda}$, both countries observe a lower number of firms, lower prices, and a higher welfare level coupled with a higher degree of inequality than in a free-trade equilibrium with national labor markets. (ii) Starting with migration barriers $\lambda \in [0,\bar{\lambda}]$, a piecemeal integration of labor markets,* $\mathrm{d}\lambda < 0$*, has an ambiguous effect on the number of firms and the degree of inequality but leads to a lower goods price and an increase in welfare in both countries.*

Proof: The analytical details of the proof are relegated to Appendix A.8.3 for part (i) and to Appendix A.8.4 for part (ii).

The intuition for part (i) is as follows. In the equilibrium considered, the degree of labor market competition is higher than in a pricing equilibrium where the same number of domestic firms interact in autarkic labor markets, while due to $\lambda = \bar{\lambda}$ labor supply is the same in both equilibria. This follows from Section 5.1 above. Comparing these two equilibria, maximum profits are lower with open labor markets than with autarkic labor markets. Therefore, the zero-profit equilibrium with open labor markets must have a lower number of firms (higher m) than the zero-profit equilibrium with autarkic labor markets. The welfare increase owes to the excess-entry property of the equilibrium. While the productivity distortion is not affected as long as no one migrates, the wage markup is lowered because firms perceive a larger elasticity of labor supply. With a lower wage-markup distortion relative to the productivity distortion, the allocation is now closer to the social optimum. Therefore, the firm exit induced by opening up labor markets entails a first-order welfare gain, even if the cost from migration barriers is prohibitive at $\bar{\lambda}$. Despite firm exit, goods prices will fall due to lower markups. Moreover, we have $d^n[m, \bar{\lambda}] = m$. Hence, the lower number of firms also implies a higher degree of inequality.

The ambiguity regarding m in part (ii) of the proposition is best grasped by remembering the mechanisms responsible for changes in firms' average cost, $g^M[m, \lambda]$, and the double markup $\Gamma^M[m, \lambda] := \rho^T[m]\psi^M[m, \lambda]$. Using subscripts to denote partial derivatives, the skill distance m obeys

$$\mathrm{d}m = \frac{\Gamma^M_\lambda - g^M_\lambda}{g^M_m - \Gamma^M_m}\mathrm{d}\lambda. \tag{30}$$

An increase in m makes firms larger, but it also lowers the productivity of the average worker. Appendix A.8.4 shows that the size effect always dominates, leading to a lower average cost, $g^M_m < 0$. A lower number of firms (higher m) reduces both the perceived price elasticity of goods demand as well the perceived labor supply elasticity, meaning that $\Gamma^M_m > 0$. A reduction in λ improves the productivity of the average worker through savings in migration cost as well as through a resorting of workers across the border. From (29), we therefore have $g^M_\lambda > 0$. And finally, we have shown above that the perceived elasticity of labor supply increases with lower migration barriers, leading to a lower markup $\psi^M[m, \lambda]$. This implies $\Gamma^M_\lambda > 0$ as well. In view of Equation (30), all of this implies that for $\lambda < \bar{\lambda}$ the reaction of m to a reduction of λ is ambiguous. The skill distance m falls, if the pricing effect dominates the productivity effect via better matching, i.e., if $\Gamma^M_\lambda < g^M_\lambda$.

The ambiguity in m is also responsible for the ambiguity in inequality. The change in the skill reach for natives is

$$\mathrm{d}d^n = \frac{(1-\lambda)f'[m - d^n]}{f'[d^n] + (1-\lambda)f'[m - d^n]}\mathrm{d}m - \frac{f[m - d^n]}{f'[d^n] + (1-\lambda)f'[m - d^n]}\mathrm{d}\lambda. \tag{31}$$

The second term is negative for $\mathrm{d}\lambda < 0$, working towards a lower degree of inequality, However, this may be offset by an increase in m, so that the overall inequality effect is ambiguous.

Turning to the price effect in part (ii) of the proposition, we first note that $p = g^M$. Inserting Equation (30) back into $\mathrm{d}g^M = g^M_m \mathrm{d}m + g^M_\lambda \mathrm{d}\lambda$, we obtain an unambiguous fall in the goods price upon a reduction of λ. Intuitively, a lower λ lowers average cost g^M as well as markups, leading to unambiguously lower prices regardless of whether the number of firms is increasing or decreasing. The welfare effect is determined by the change in real income $\theta^M[m,\lambda]\big/ p[m]$ and the number of varieties. In view of (15), the welfare effect in Proposition 3 (ii) may therefore be described as

$$\mathrm{d}\ln V = \frac{\partial \ln\left[\theta^M/g^M\right]}{\partial\lambda}\mathrm{d}\lambda + \left[\frac{\partial \ln\left[\theta^M/g^M\right]}{\partial m} - \frac{1}{4\gamma H}\right]\mathrm{d}m, \tag{32}$$

where we again replace $p = g^M$. The first term describes the direct effect of lower migration barriers, $\mathrm{d}\lambda < 0$, on real income. This term is unambiguously positive because a lower cost of migration increases the average productivity of workers and lowers the average cost. The remaining term involving $\mathrm{d}m$ seems ambiguous, but we know that the equilibrium considered involves excess entry. This implies that $\mathrm{d}m > 0$ involves a positive the first-order effect on welfare, which means $\partial \ln\left[\theta^M/g^M\right]\big/\partial m > 1/(4\gamma H)$. It thus follows that a lower λ leads to higher welfare, if it leads to firm exit. If it leads to firm entry, then the second term in Equation (32) above is negative. However, $\mathrm{d}m$ is driven by $\mathrm{d}\lambda$, and inserting $\mathrm{d}m = (\partial m/\partial\lambda)\mathrm{d}\lambda$ one can show that the first term in (32) dominates, leading to an unambiguously positive welfare effect from $\mathrm{d}\lambda < 0$ for any initial $\lambda \in [0, \bar{\lambda}]$; see Appendix A.8.4.

Proposition 3 implies the following corollary:

Corollary 1. *A free-trade equilibrium with open labor markets and an alternating pattern of location delivers a level of welfare that is strictly larger for both countries, regardless of the value of $\lambda \in [0, \bar{\lambda}]$, than the level of welfare in the free-trade equilibrium with national labor markets.*

We close this section by briefly discussing the robustness of our findings. A first point to note is that positive welfare effects of Proposition 3 part (ii) do not hinge on the excess-entry property of our equilibrium. To see this, consider a case where entry is efficient. In this case, a change in m has no first-order effect on welfare, whence the final two terms in Equation (32) would offset each other. We are thus left with the effects of a lower λ through a better matching quality and lower markups from more competitive labor markets in the first term of Equation (32), which is unambiguously positive.

A second point relates to the symmetric location pattern of firms on the skill circle. To isolate the mechanisms through which migration affects welfare, we have restricted our analysis to a "perfectly symmetric" equilibrium with equal distances m between any two firms from different countries. This location pattern maximizes the productivity gains from migration, given any non-prohibitive level of migration barriers. However, one can imagine alternative equilibria that preserve symmetric skill distances between firms from the same country but not necessarily between domestic and foreign firms. An extreme case is a location where domestic and foreign firms are positioned on the same points on the skill circle. This pattern involves zero migration, provided that the migration cost is strictly positive. More generally, for any positive level of the migration cost, there exist quasi-symmetric equilibria where domestic firms are sufficiently close to their nearest foreign neighbor to preclude any incentive for workers to migrate. In Heiland and Kohler (2018) we model the entry game for arbitrary location patterns and discuss conditions under which such an equilibrium with quasi-symmetric locations and de-facto national labor markets will prevail over the fully symmetric alternating location equilibrium underlying Proposition 3.

6 Conclusion

We have readdressed the common narrative of variety-based gains from trade. Traditional models of monopolistic competition stress the importance of a large resource base for a high 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 for the economy at large. If anything, it increases the average productivity level through a positive selection effect among heterogeneous firms. This view neglects an important aspect of the labor market: If the labor force is heterogeneous in terms of skill types that are specific to the production of certain goods, then firms have monopsony power on the labor market and some workers will be employed in less than ideal matches. As a result, productivity gains from specializing on a coarser set of goods come at the expense of a less competitive labor market and more workers being employed in less than ideal matches.

In this environment, opening up to trade is a less benign force than portrayed in conventional models of monopolistic competition. In particular, trade-induced firm exit worsens the average quality of matches between the types of skills that workers bring to their firms and the specific skill requirements of the goods produced by these firms, and it increases the distortion between the marginal productivity of labor and the wage rate. This latter effect works against the conventional pro-competitive effect of trade on the goods markets. However, 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 worker-firm matches and from higher markups on the labor market. The gains from trade theorem survives. Looking at piecemeal trade liberalization between two symmetric countries, we find a positive aggregate welfare effect, unless the initial level of trade cost is above a certain threshold level. And we generally find that trade liberalization increases the degree of income inequality among workers.

In this environment, labor market integration generates an incentive for workers to migrate for employment in certain foreign firms, even if countries are completely symmetric. Barring prohibitive migration cost, an equilibrium with integrated labor markets thus involves two-way migration. Broadly speaking, migration tends to undo the negative labor market effects of trade: it tends to improve the quality of matches while at the same time lowering firms' monopsony power on the labor market. Unlike piecemeal trade liberalization, any scenario of piecemeal liberalization of migration is welfare enhancing. Moreover, while trade unambiguously increases income inequality, migration has an ambiguous effect on inequality. Trade and migration are complements, rather than substitutes, since trade-induced specialization increases migration incentives. Our model thus 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.

Arkolakis, C., Costinot, A., Donaldson, D., and Rodríguez-Clare, A. (2019). The Elusive Pro-Competitive Effects of Trade. The Review of Economic Studies, 86:46–80.

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

Becker, S., Egger, H., Koch, M., and Muendler, M.-A. (2019). Tasks, Occupations, and Wage Inequality in an Open Economy. mimeo.

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

Bykadorov, I., Ellero, A., Funari, S., Kokovin, S., and Molchanov, P. (2016). Painful Birth of Trade Under Classical Monopolistic Competition. HSE Working Paper.

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. Econometrica, 87(3):741–835.

Costinot, A. (2009). An Elementary Theory of Comparative Advantage. Econometrica, 77(4):1165–1192.

Costinot, A. and Vogel, J. (2010). Matching and Inequality in the World Economy. Journal of Political Economy, 118(4):pp. 747–786.

Costinot, A. and Vogel, J. (2015). Beyond Ricardo: Assignment Models in International Trade. Annual Review of Economics, 7(1):31–62.

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.

Egger, H., Kreickemeier, U., Moser, C., and Wrona, J. (2019). Exporting and Offshoring with Monopsonistic Competition. CESifo Working Paper Series 7774.

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

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. (2015). Chapter 18 - Migration, International Trade, and Capital Formation: Cause or Effect? In Chiswick, B. R. and Miller, P. W., editors, Handbook of the Economics of International Migration, volume 1, pages 913 – 1025. North-Holland.

Galor, O. (1986). Time preference and international labor migration. Journal of Economic Theory, 38(1):1 – 20.

Gathmann, C. and Schönberg, U. (2010). How General Is Human Capital? A Task-Based Approach. Journal of Labor Economics, 28(1):1–49.

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

Grogger, J. and Hanson, G. H. (2011). Income Maximization and the Selection and Sorting of International Migrants. Journal of Development Economics, 95(1):42 – 57.

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

Grossman, G. M., Helpman, E., and Kircher, P. (2017). Matching, sorting, and the distributional effects of international trade. Journal of Political Economy, 125(1):224–264.

Gulyas, A. (2018). Identifying Labor Market Sorting with Firm Dynamics. Meeting Papers 856, Society for Economic Dynamics.

Heiland, I. and Kohler, W. (2018). Heterogeneous Workers, Trade, and Migration. CESifo Working Paper Series 7355.

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.

Iranzo, S. and Peri, G. (2009). Migration and Trade: Theory with an Application to the Eastern-Western European Integration. Journal of International Economics, 79(1):1–19.

Kreickemeier, U. and Wrona, J. (2017). Two-Way Migration between Similar Countries. The World Economy, 40(1):182–206.

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. (1980). Intra-Industry Trade under Perfect Monopolistic Competition. Journal of International Economics, 10(2):151 – 175.

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

Lazear, E. P. (2009). Firm-Specific Human Capital: A Skill-Weights Approach. Journal of Political Economy, 117(5):914–940.

Lindenlaub, I. (2017). Sorting Multidimensional Types: Theory and Application. The Review of Economic Studies, 84(2):718–789.

MacKenzie, G. (2019). Trade and Market Power in Product and Labor Markets. unpublished manuscript.

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

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

Manning, A. (2003). Monopsony in Motion: Imperfect Competition in Labor Markets. Princeton University Press.

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

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

Moscarini, G. (2001). Excess Worker Reallocation. The Review of Economic Studies, 68(3):593–612.

Mrázová, M. and Neary, J. P. (2014). Together at Last: Trade Costs, Demand Structure, and Welfare. American Economic Review, 104(5):298–303.

Mrázová, M. and Neary, J. P. (2017). Not So Demanding: Demand Structure and Firm Behavior. American Economic Review, 107(12):3835–74.

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

Rosen, S. (1978). Substitution and Division of Labour. Economica, 45(179):235–250.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. Oxford Economic Papers, 3(2):pp. 135–146.

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

Vogel, J. (2008). Spatial Competition with Heterogeneous Firms. Journal of Political Economy, 116(3):pp. 423–466.

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

Appendix

A.1 Stylized facts on two-way migration

Table A.1: Correlation between bilateral stocks of immigrants and emigrants

	OECD		ALL	OECD, by AGE			OECD, by EDU			OECD, POISSON	
	levels	resid.	resid.	15-24	25-64	65+	prim.	second.	tert.	excl. zeros	incl. zeros
$\rho(Em, Im)$	.69	.59	.44	.61	.54	.46	.64	.56	.53	.43	.39
N	231	231	1332	217	231	209	221	227	224	231	269

All correlations are significantly different from zero at the 1% significance level. All correlations except column 1 based on gravity residuals controlling for *bil. distance, contiguity, common language, EU* or *RTA membership, common colonizer, current colonial relationship, common legal origin, time difference,* as well as origin and destination fixed effects. Residuals in columns 2-9 (10,11) are based on log-linear OLS estimation (Poisson estimation in levels). All columns except column 3 include only OECD origin and destination countries (26 countries, all members in 2000 except DEU,ISL,JAP,KOR, unbalanced). Column 3 includes 60 additional non-OECD countries. Source: OECD-DIOC-E database 2000 and CEPII gravity database.

A.2 Expenditure share and markup

A.2.1 Proof of Equation (3)

Starting out from (2), inserting $p_i = \frac{\delta_i Y}{q_i}$ gives

$$\delta_i = \frac{1}{N} + \gamma\overline{\ln p} - \gamma \ln \frac{\delta_i Y}{q_i}. \tag{A.1}$$

This can be rewritten as

$$\frac{\delta_i}{\gamma} + \ln \frac{\delta_i}{\gamma} = \frac{1}{\gamma N} + \overline{\ln p} - \ln \frac{Y}{q_i} - \ln \gamma. \tag{A.2}$$

Applying the Lambert function $\mathcal{W}[z]$, defined as the solution to $\ln x + x = \ln z$, we obtain $\delta_i = \delta[q_i, \overline{\ln p}, N, Y]$ as given in (3).

A.2.2 Proof of Equation (7)

Similar logic can be applied to obtain an explicit solution for the optimal price determined by the first-order condition (6). Defining perceived marginal cost as $\tilde{w}_i := [(\eta_i + 1)/\,\eta_i]w_i\beta$

and observing (2) and (5), this condition can be written as

$$\frac{p_i}{\tilde{w}_i} + \ln p_i = 1 + \frac{1}{\gamma N} + \overline{\ln p}. \tag{A.3}$$

The left-hand side is an implicit function of the profit maximizing price p_i. Rewriting (A.3) as

$$\frac{p_i}{\tilde{w}_i} + \ln p_i - \ln \tilde{w}_i = 1 + \frac{1}{\gamma N} + \overline{\ln p} - \ln \tilde{w}_i \tag{A.4}$$

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

$$p_i = \mathcal{W}\left[\tilde{w}_i^{-1} \exp\left\{1 + \frac{1}{\gamma N} + \overline{\ln p}\right\}\right] \tilde{w}_i. \tag{A.5}$$

which implies the price markup as given in (7).

A.3 Proof that the variance of incomes across individuals increases in m

Let $\sigma^2[m] = \frac{1}{m}\int_0^m f[x]^2 dx - \left(\frac{1}{m}F[m]\right)^2$ and note that $\sigma^2[m] \propto Var[m]$, the variance of income across all inviduals, by a factor $(L/H)^2$. Then,

$$\begin{aligned}
\frac{\partial \sigma^2[m]}{\partial m} &= -\frac{1}{m^2}\int_0^m f[x]^2 dx + \frac{1}{m} f[m]^2 - 2\left(\frac{1}{m}F[m]\right)\left(-\frac{1}{m^2}F[m] + \frac{1}{m}f[m]\right) \\
&= -\frac{1}{m^2}\int_0^m \left(\theta - f[x]\right)^2 dx + \frac{1}{m}\left(\theta - f[m]\right)^2
\end{aligned}$$

implies

$$\frac{\partial \sigma^2[m]}{\partial m} \geq 0 \qquad \Leftrightarrow \qquad \frac{1}{m^2}\int_0^m \left(\theta - f[x]\right)^2 dx \leq \frac{1}{m}\left(\theta - f[m]\right)^2.$$

Concavity implies

$$|\theta - f[x]| \leq |\theta - f[m]| \qquad \forall x \in [0, m].$$

Hence,

$$\frac{1}{m^2}\int_0^m \left(\theta - f[x]\right)^2 dx \leq \frac{1}{m^2}\int_0^m \left(\theta - f[m]\right)^2 dx = \frac{1}{m}\left(\theta - f[m]\right)^2.$$

A.4 Existence and uniqueness of the symmetric two-stage equilibrium

As described in Section 3.2, for symmetric distance patterns a zero-profit equilbrium in the two-stage game is given by a root of the function

$$G[m] := \rho[m]\psi[m] - g[m], \tag{A.6}$$

where $g[m] > 1$ equals average costs. 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 $\rho_m = 1/(\gamma H) > 0$. As shown in Section 3.2, $\eta_m < 0$, whence we 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 positive profits, while $G[m] < 0$ implies that firms make losses.

The following conditions are *sufficient* for a symmetric zero-profit equilibrium to *exist* and to be *unique*: a) $G[H] > 0$, b) $G[m]$ is continuous and $G'[m] > 0$ in the interval $(\tilde{m}, H]$, where $\tilde{m}$ is defined by $\frac{L}{N[\tilde{m}]}\theta[\tilde{m}] = \alpha$.

Condition a) requires that a single firm in the market makes non-negative profits, that is,

$$\frac{L\theta[H]}{L\theta[H] - \alpha} \leq \left(1 + \frac{1}{\gamma}\right)\psi[H]. \tag{A.7}$$

Observing that $\psi[m]$ increases in m, we can set $\psi[H]$ on the right-hand side to its minimum level of unity to obtain

$$\frac{\alpha}{L\beta}(1+\gamma) \leq \frac{F[H]}{H}, \tag{A.8}$$

which is a sufficient condition for (A.7). It shows, that given α, β, L and H, the degree of substitutability of goods in the utility function γ must not be too large. Relating back to (A.7) in its original form, these restrictions imply that the price markup over marginal cost that a single firm can choose exceeds its average cost.[23]

Condition b) requires that firm entry associated with a decrease in the skill reach m lowers profits in the relevant range where firms produce positive output, that is, for $m \in (\tilde{m}, H]$. Since we know from above that $\rho'[m] > 0$ as well as $\psi'[m] > 0$, condition b)

[23] This condition is well known from the standard New Trade Theory model with homogeneous workers (cp. Equation (10) in Krugman, 1980).

is satisfied if $g_m < 0$. It is straightforward to show that

$$g'[m] = \frac{\frac{L}{H} f[m]}{\frac{mL}{H}\theta[m] - \alpha} \left(1 - \frac{\frac{mL}{H}\theta[m]}{\frac{mL}{H}\theta[m] - \alpha}\right) < 0 \qquad \text{for } m \in (\tilde{m}, H]. \tag{A.9}$$

Hence, there exists a unique $N = \frac{H}{m} \geq 1$ satisfying $G[m] = 0$.

A.5 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 standard 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 decreasing $g[m, H]$. 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 (14), we obtain[24]

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

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 (14) 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

$$\frac{L}{L - \alpha\underline{N}} = 1 + \frac{1}{\gamma\underline{N}} \tag{A.10}$$

[24] To ease notation, in what follows we use subscripts to denote partial derivatives of functions with multiple arguments.

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

A.6 The constrained social optimum

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

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

The first-order condition is

$$\frac{Lf[m]}{L\theta[m] - \frac{\alpha H}{m}} = 1 + \frac{m}{2\gamma H}. \tag{A.11}$$

The second-order condition requires

$$\frac{\mathrm{d}^2 \ln V}{\mathrm{d}m^2} = -\frac{\left(L\theta'[m] + \frac{\alpha H}{m^2}\right)^2}{\left(L\theta[m] - \frac{\alpha H}{m}\right)^2} + \frac{L\theta''[m] - 2\frac{\alpha H}{m^3}}{L\theta[m] - \frac{\alpha H}{m}} < 0. \tag{A.12}$$

A sufficient condition for this to hold is

$$\theta''[m] := \frac{\partial^2\theta[m]}{\partial m^2} = \frac{1}{m}\left(f'[m] - \frac{2}{m}f[m] + \frac{2}{m}\theta[m]\right) \leq 0 \tag{A.13}$$

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 $\theta''[m] \leq 0$ 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 (14) we rewrite (A.11) as

$$g[m] = \frac{\theta[m]}{f[m]}\frac{1}{\psi[m]}\psi[m]\rho[m/2]. \tag{A.14}$$

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

$\rho'[m] > 0$,

$$\frac{\theta[m]}{f[m]}\frac{1}{\psi[m]} < 1 \tag{A.15}$$

is a sufficient condition for this to hold. Rearranging (A.15) 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.15) is fulfilled and it follows that the market equilibrium firm size is too small compared to the socially optimal allocation.

A.7 Further details of the trading equilibrium

A.7.1 Proof of Proposition 1

(i) Log-differentiating the equilibrium condition (17) and setting $k = 1$, we obtain

$$\hat{m} = A\hat{k} \qquad \text{with} \qquad A := \frac{\psi[m]\frac{1}{\gamma H}}{-g'[m] + \psi[m]\frac{1}{\gamma H} + \rho^T[m]\psi'[m]}. \tag{A.16}$$

Since $g'[m] < 0$ and $\psi'[m] > 0$ (see Appendix A.4 for details) we find that $0 < A < 1$, which implies $0 < \hat{m} = A\hat{k} < \hat{k}$. Hence, m increases and the number of firms in each country falls. However, $A < 1$ implies that the total number of available varieties $N^T = kN > N^A$ is still larger with trade than under autarky.

(ii) As the price markup depends negatively on the number of available varieties kN, it follows directly from the previous result that it must fall. Furthermore, we know from above that the wage markup increases. Log-differentiating (16) and setting $k = 1$ yields

$$\hat{p} = B\hat{k} \qquad \text{with} \qquad B = \frac{\frac{m}{\gamma H}}{\left(1 + \frac{m}{\gamma H}\right)}\frac{g'[m]}{\left(-g'[m] + \psi[m]\frac{1}{\gamma H} + \rho^T[m]\psi'[m]\right)}. \tag{A.17}$$

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

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

(iv) We know from above that goods prices are lower in the free-trade equilibrium, which contributes to higher real incomes. At the same time, exit increases m and thus average productivity $\theta[m]$, which has a negative effect on real income. The logic of A.6 implies that the free-trade equilibrium, like autarky, is characterized by excess entry. Hence, the net effect of an increase in m must be positive. With higher real income and a larger variety available for consumption as established in (i), it follows from (15) that welfare of the worker earning average income increases.

A.7.2 The first-order conditions with two symmetric countries and trade cost

Under the assumption that the constraints $q_i, q_i^* \geq 0$ never bind, we may write (18) as

$$\max_{w_i, q_i} \left\{ r_i[q_i, N, \overline{\ln p}, Y] + r_i^* \left[\frac{\bar{q}_i - q_i}{\tau}, N, \overline{\ln p}, Y \right] - w_i L_i \right\}.$$

The first-order condition with respect to w_i then obtains as

$$\frac{p^*}{\tau} \left(\frac{\partial \ln p^*}{\partial \ln \frac{\bar{q}_i - q_i}{\tau}} + 1 \right) \frac{\partial L_i}{\partial w_i} = w_i \frac{\partial L_i}{\partial w_i} + L_i \quad \Leftrightarrow \quad p^* = \frac{\varepsilon_i^*}{\varepsilon_i^* - 1} \frac{\eta_i + 1}{\eta_i} w_i \tau,$$

and the first-order condition with respect to q_i reads

$$p \left(\frac{\partial \ln p}{\partial \ln q_i} + 1 \right) \frac{\partial L_i}{\partial w_i} = \frac{p^*}{\tau} \left(\frac{\partial \ln p^*}{\partial \ln \frac{\bar{q}_i - q_i}{\tau}} + 1 \right) \frac{\partial L_i}{\partial w_i} \quad \Leftrightarrow \quad p \frac{\varepsilon_i - 1}{\varepsilon_i} = \frac{p^*}{\tau} \frac{\varepsilon_i^* - 1}{\varepsilon_i^*}.$$

Both first-order conditions together imply equations (20).

A.7.3 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 (20), we can use the same logic as in A.2 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:

$$p = \frac{\mathcal{W}[\tilde{Z}]}{2} \psi \qquad \text{with } \tilde{Z} = \frac{2}{\psi} \exp \left\{ 2 + \frac{m}{\gamma H} + \ln p^* \right\} \tag{A.18}$$

$$p^* = \frac{\mathcal{W}[\tilde{Z}^*]}{2} \psi \tau \quad \text{with } \tilde{Z}^* = \frac{2}{\psi \tau} \exp \left\{ 2 + \frac{m}{\gamma H} + \ln p \right\}. \tag{A.19}$$

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

$$p = \mathcal{W} \left[\mathcal{W}[\tilde{Z}^*] \tau \exp \left\{ 2 + \frac{m}{\gamma H} \right\} \right] \frac{\psi}{2} \tag{A.20}$$

$$p^* = \mathcal{W}\left[\frac{\mathcal{W}[\tilde{Z}]}{\tau}\exp\left\{2+\frac{m}{\gamma H}\right\}\right]\frac{\psi}{2}\tau. \tag{A.21}$$

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

$$W := W[W^*, m] = \mathcal{W}\left[W^*\tau\exp\left\{2+\frac{m}{\gamma H}\right\}\right] \tag{A.22}$$

$$W^* := W^*[W, m] = \mathcal{W}\left[\frac{W}{\tau}\exp\left\{2+\frac{m}{\gamma H}\right\}\right]. \tag{A.23}$$

Note that for zero trade costs ($\tau = 1$) 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 costs. For any $\tau > 1$, it must therefore be true that $W > W^*$.

The two-country version of (A.3) can be written as $p = \left(1 + \frac{1}{\gamma N^T} + \frac{1}{2}(\ln p^* - \ln p)\right)\tilde{w}$ and analogously for p^*. In view of (A.18) and (A.19) it follows that $\frac{W}{2} = 1 + \frac{1}{\gamma N^T} + \frac{1}{2}(\ln p^* - \ln p)$ and $\frac{W^*}{2} = 1 + \frac{1}{\gamma N^T} + \frac{1}{2}(\ln p - \ln p^*)$. The expenditure shares in (19) can therefore be written as

$$\delta = \left(\frac{W}{2} - 1\right)\gamma \qquad \text{and} \qquad \delta^* = \left(\frac{W^*}{2} - 1\right)\gamma. \tag{A.24}$$

Direct demand functions for foreign varieties in terms of W^* obtain as $q^* = \frac{\delta^* Y}{p^*}\left(1 - \frac{2}{W^*}\right)\frac{\gamma Y}{\psi}$. This implies that the prohibitive level of trade costs $\bar{\tau}$ 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 costs $W \geq W^* \geq 2$. Inserting demand and income $Y = L\theta$ into the labor market clearing condition (21), and rearranging terms gives

$$\begin{aligned}\gamma\left(2 - \frac{2}{W} - \frac{2}{W^*}\right) &= \frac{\frac{L\theta[m]}{N[m]} - \alpha}{L\theta[m]}\psi[m] \\ \gamma h[W, W^*] &= \frac{\psi[m]}{g[m]N[m]}.\end{aligned} \tag{A.25}$$

For easier reference the second line introduces $h[W, W^*] := \left(2 - \frac{2}{W} - \frac{2}{W^*}\right)$. (A.25), (A.22) and (A.23) 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 $\widehat{W} = \omega\hat{\tau}$, $\widehat{W}^* = \omega^*\hat{\tau}$ and $\widehat{m} = \mu\hat{\tau}$. 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.25), (A.22), (A.23) leads to

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

$$\begin{bmatrix} -\frac{1}{W-1-\frac{W}{W^*}} & -\frac{1}{W^*-1-\frac{W^*}{W}} & \frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N}-\alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \\ -1 & \frac{1}{W+1} & \frac{1}{\gamma N}\frac{1}{W+1} \\ \frac{1}{W^*+1} & -1 & \frac{1}{\gamma N}\frac{1}{W^*+1} \end{bmatrix} \begin{bmatrix} \widehat{W} \\ \widehat{W^*} \\ \widehat{m} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{W+1}\hat{\tau} \\ \frac{1}{W^*+1}\hat{\tau} \end{bmatrix}. \tag{A.26}$$

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

$$\omega = \frac{1}{(W+1)(W^*+1)} \left[\left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \right) W^* - \frac{1}{\gamma N h[W, W^*]} \frac{4}{W^*} \right] \frac{1}{\det[D]} \tag{A.27}$$

$$\omega^* = \frac{1}{(W+1)(W^*+1)} \left[- \left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \right) W + \frac{1}{\gamma N h[W, W^*]} \frac{4}{W} \right] \frac{1}{\det[D]} \tag{A.28}$$

$$\mu = \frac{2W^*/W - 2W/W^*}{h[W, W^*](W+1)(W^*+1)} \frac{1}{\det[D]}. \tag{A.29}$$

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

$$\det[D] = \left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \right) \frac{WW^* + W + W^*}{(W+1)(W^*+1)} - \frac{1}{\gamma N h[W, W^*]} \frac{(2+W^*)\frac{2}{W} + (2+W)\frac{2}{W^*}}{(W+1)(W^*+1)}. \tag{A.30}$$

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

$$\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} > \frac{1}{\gamma N h[W, W^*]}. \tag{A.31}$$

We know from above that $\frac{f[m]}{\theta[m]} < 1$ and $\frac{\partial \ln \psi}{\partial \ln m} > 0$. Therefore, inequality (A.31) holds if

$$\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} > \frac{1}{\gamma N h[W, W^*]}. \tag{A.32}$$

Using the equilibrium condition (A.25), we can rewrite this as $\psi[m] \geq \theta[m]/f[m]$. We have demonstrated in Appendix A.6 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.31) jointly 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 costs, converging to zero as τ approaches one. In view of A.29, the level of τ enters through W and W^*. The lower the trade cost level, 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

$$\hat{p}^* = \left(\omega^* + \frac{\partial \ln \psi}{\partial \ln m}\mu + 1\right)\hat{\tau}, \tag{A.33}$$

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.28) and (A.29) shows that $\hat{p}^*$ is positive if and only if

$$-\frac{d_{13}W - \frac{2}{W}\frac{2}{\gamma h N} + \frac{\partial \ln \psi}{\partial \ln m}\frac{1}{h[W,W^*]}\left(\frac{2W}{W^*} - \frac{2W^*}{W}\right)}{d_{13}(WW^* + W + W^*) - \frac{2}{\gamma h[W,W^*]N}\left(\frac{2+W^*}{W} + \frac{2+W}{W^*}\right)} + 1 > 0 \tag{A.34}$$

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.32), 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.34) to hold. Remembering from above that $W \geq W^* \geq 2$, it is straightforward to show that these two conditions are fulfilled.

The change in the domestic price obtains as

$$\hat{p} = \left(\omega + \frac{\partial \ln \psi}{\partial \ln m}\mu\right)\hat{\tau}. \tag{A.35}$$

We know from above that $\omega > 0$; the pro-competitive effect of lower trade costs 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 $\bar{\tau}$, the labor market effect dominates. Inserting (A.27) and (A.29) gives

$$\begin{aligned}\hat{p} = \Bigg[& W^*\left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N}-\alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m}\right) - \frac{2}{\gamma N h[W,W^*]}\frac{2}{W^*} \\ & - \frac{\partial \ln \psi}{\partial \ln m}\frac{1}{h[W,W^*]}\left(\frac{2W}{W^*} - \frac{2W^*}{W}\right)\Bigg] \times \frac{1}{(W+1)(W^*+1)}\frac{\hat{\tau}}{\det[D]}.\end{aligned} \tag{A.36}$$

Remember that prohibitive trade costs imply an infinite price elasticity and therefore a price markup of zero, whence $W^* = 2$. To see whether $\hat{p} > 0$ for $\tau = \bar{\tau}$, as stated in Proposition 2, we must therefore evaluate the bracketed term at $W^* = 2$. We obtain

$$-2\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N}-\alpha} + 2\frac{f[m]}{\theta[m]} - 2 - 2\frac{\partial \ln \psi}{\partial \ln m} + \frac{2}{\gamma N h[W,W^*]}\frac{2}{W^*} + \frac{\partial \ln \psi}{\partial \ln m}(W+2). \tag{A.37}$$

Inserting the equilibrium condition (A.25), 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

$$\psi W \frac{f[m]}{\theta[m]} < 2\frac{f[m]}{\theta[m]} + W - 2 + W\frac{\partial \ln \psi}{\partial \ln m}. \tag{A.38}$$

Inserting the explicit expressions for ψ and $\frac{\mathrm{d}\ln\psi}{\mathrm{d}\ln m}$ leads to

$$\frac{W}{\theta[m]}\frac{f[m]^2 - 2f'[m]F[m]}{f[m]} < W - 2 + \frac{2f[m]}{\theta[m]} + W\left(\frac{-2f''[m]\theta}{f[m]^2\psi} - \frac{2mf'[m]}{f[m]}\right). \tag{A.39}$$

Since $f''[m] \leq 0$, the inequality holds if

$$\frac{W}{\theta[m]}\frac{f[m]^2 - 2f'[m]F[m]}{f[m]} < W - 2 + \frac{2f[m]}{\theta[m]} - W\frac{2mf'[m]}{f[m]}. \tag{A.40}$$

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 costs is given by $\ln V = \ln\theta[m] - \ln P^T[p, p^*, m]$, where

$$\ln P^T[p, p^*, m] = \frac{1}{2\gamma N^T} + \frac{1}{N^T}\sum_{i=1}^{N^T}\ln p_i + \frac{\gamma}{2N^T}\sum_{i=1}^{N^T}\sum_{j=1}^{N^T}\ln p_i(\ln p_j - \ln p_i), \tag{A.41}$$

with $N^T = N + N^*$ and i, j indexing domestic *and* foreign varieties. Under symmetry, $N^* = N = N^T/2$ and the price index simplifies to

$$\ln P^T[p, p^*, m] = \frac{1}{4\gamma N} + \frac{1}{2}\ln p + \frac{1}{2}\ln p^* - \frac{\gamma N}{4}\left(\ln p - \ln p^*\right)^2. \tag{A.42}$$

The change in indirect utility is then

$$\widehat{V} = \left(\frac{\partial \ln\theta}{\partial \ln m} - \frac{\partial \ln P}{\partial \ln m}\right)\hat{m} - \frac{\partial \ln P}{\partial \ln p}\hat{p} - \frac{\partial \ln P}{\partial \ln p^*}\hat{p}^*, \tag{A.43}$$

with $\frac{\partial\ln\theta}{\partial\ln m} = \frac{f[m]-\theta[m]}{\theta[m]} < 0$, $\frac{\partial \ln P}{\partial \ln m} = \frac{1}{4\gamma N} + \frac{\gamma N}{4}\left(\ln p - \ln p^*\right)^2 > 0$, $\frac{\partial\ln P}{\partial \ln p} = \frac{1}{2} + \frac{\gamma N}{2}\left(\ln p - \ln p^*\right) \geq 0$ and $\frac{\partial \ln P}{\partial \ln p^*} = N\delta^* \geq 0$. Inserting yields (22).

Using the results that at the prohibitive level of trade costs $\delta^* = 0$, $\hat{p} > 0$ and $\hat{m} > 0$, it follows from (22) that $\widehat{V} < 0$ at $\tau = \bar{\tau}$. At $\tau = 1$ it holds that $\hat{m} = 0$, $\hat{p} < 0$ and $\hat{p}^* < 0$. Hence, $\widehat{V} > 0$ at $\tau = 1$.

A.8 Additional details of the trade and migration equilibrium

A.8.1 The elasticity of labor supply

The elasticity of labor supply in the symmetric alternating equilibrium is defined as $\frac{\partial L^{S,M}}{\partial w_i}\frac{w_i}{L^{S,M}}$. From (24) and (23), we obtain

$$\frac{\partial L^{S,M}}{\partial w_i} = \frac{L}{H}\frac{\partial d_i^n}{\partial w_i}f[d_i^n] + (1-\lambda)\frac{L}{H}\frac{\partial d_i^m}{\partial w_i}f[d_i^m] \qquad \text{with} \tag{A.44}$$

$$\frac{\partial d_i^n}{\partial w_i} = \frac{f[d_i^n]}{-w_i f'[d_i^n] - w^*(1-\lambda)f'[m-d_i^n]} \tag{A.45}$$

$$\frac{\partial d_i^m}{\partial w_i} = \frac{(1-\lambda)f[d_i^m]}{-w_i(1-\lambda)f'[d_i^m] - w^* f'[m-d_i^m]}. \tag{A.46}$$

Evaluating $\frac{\partial L^{S,M}}{\partial w_i}\frac{w_i}{L^{S,M}}$ 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

$$\eta^M = \frac{\partial L^{S,M}}{\partial w_i}\frac{w_i}{L^S}\bigg|_{w_i=w} = \frac{L}{H}\left(\frac{f[d^n]^2}{-f'[d^n]-(1-\lambda)f'[m-d^n]} + \frac{(1-\lambda)^2 f[d^m]^2}{-(1-\lambda)f'[d^m]-f'[m-d^m]}\right)$$
$$\times \frac{1}{\frac{L}{H}\left(\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\int_0^{d^m} f[d]\mathrm{d}d\right)}$$
$$= \frac{2f[d^n]^2}{f'[d^n]+(1-\lambda)f'[d^m]}\frac{-1}{\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\int_0^{d^m} f[d]\mathrm{d}d}. \qquad \text{(A.47)}$$

The elasticity of labor supply decreases in m:

$$\eta^M_m = \eta^M\left[\frac{2f'[d^n]}{f[d^n]}\frac{\partial d^n}{\partial m} - \frac{-f''[d^n]\frac{\partial d^n}{\partial m} - (1-\lambda)f''[d^m]\frac{\partial d^m}{\partial m}}{-f'[d^n]-(1-\lambda)f'[d^m]} - \frac{f[d^n]}{m\theta^M}\right] < 0, \qquad \text{(A.48)}$$

where $\frac{\partial d^n}{\partial m} = \frac{(1-\lambda)f'[d^m]}{f'[d^n]+(1-\lambda)f'[d^m]} > 0$ and $\frac{\partial d^m}{\partial m} = \frac{f'[d^n]}{f'[d^n]+(1-\lambda)f'[d^m]} > 0$.

Furthermore, provided that $f'''[\cdot]$ is not too positive, η^M decreases in λ:

$$\eta^M_\lambda = \eta^M\left[\underbrace{\frac{2f'[d^n]}{f[d^n]}\frac{\partial d^n}{\partial \lambda}}_{<0} + \underbrace{\frac{f''[d^n]\frac{\partial d^n}{\partial \lambda} + (1-\lambda)f''[d^m]\frac{\partial d^m}{\partial \lambda} + f'[d^m]}{-f'[d^n]-(1-\lambda)f'[d^m]}}_{<0} + \underbrace{\frac{F[d^m]}{F[d^n]+(1-\lambda)F[d^m]}}_{>0}\right] < 0 \qquad \text{(A.49)}$$

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

$$\frac{2f'[d^n]}{f[d^n]}\frac{\partial d^n}{\partial \lambda} = 2\underbrace{\frac{f[d^m]}{f[d^n]}\frac{f'[d^n]}{f'[d^n]+(1-\lambda)f'[d^m]}}_{\geq 1} \geq \underbrace{\frac{F[d^m]}{F[d^n]+(1-\lambda)F[d^m]}}_{<1}. \qquad \text{(A.50)}$$

A.8.2 The productivity gains from migration

For $\lambda \in [0, \bar{\lambda})$, the productivity gains from migration $\theta^M[m,\lambda] - \theta[m]$ increase in m, since $\frac{\partial(\theta^M[m,\lambda]-\theta[m])}{\partial m} = \frac{1}{m}\left(f[d^n] - \theta^M[m,\lambda] - f[m] + \theta[m]\right) > 0$, where the sign follows from

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

The first inequality is due to the condition $(1-\lambda)f[d^m] = f[d^n]$ and the concavity of $f[\cdot]$, which implies $(1-\lambda)f[d] - f[d^n + d] \leq (1-\lambda)f[d^m] - f[m]$ for $d \in [0, d^m]$.

A.8.3 Analytical details of the proof of Proposition 3, part (i)

In the second-stage zero-profit equilibrium with alternating firm locations and $\lambda = \bar{\lambda}$, wage markups are smaller while average cost are the same compared to a trade-only equilibrium with a similar distance $2m$ between domestic firms. That is, $\psi^M[m, \bar{\lambda}] < \psi[m]$ (implied by (9) and (25)) and $g^M[m, \bar{\lambda}] = g[m]$ (see (13) and (29)). Hence, the zero-profit equation (28) evaluated at $\lambda = \bar{\lambda}$ features a smaller number of firms. This entails a positive welfare effect since, as we prove next, the migration equilibrium inherits the excess-entry property of the autarky and trade-only equilibrium.

The number of firms is too large in the migration equilibrium. The social planner solves the same maximization problem as in Appendix A.6, additionally taking into account the integrated labor market.[25] The first-order condition of the planner then obtains as

$$\frac{Lf\left[d^n\right]}{L\theta^M - \frac{\alpha H}{m}} = 1 + \frac{m}{4\gamma H}, \tag{A.51}$$

where d^n, θ^M are shorthands for $d^n[m, \lambda], \theta^M[m, \lambda]$, respectively. A comparison with the market solution (28) 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. 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 $1 - \frac{m\theta^M(f'[d^n]+(1-\lambda)f'[d^m])}{2f[d^n]^2} > \frac{\theta^M}{f[d^n]}$. 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 show below that $f\left[\frac{d^n}{2}\right] \geq \theta^M$. Then, this inequality holds if

$$f[d^n] > f\left[\frac{d^n}{2}\right] + \frac{m}{2}\frac{f'[d^n]}{f[d^n]}\theta^M. \tag{A.52}$$

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 $f\left[\frac{d^n}{2}\right] + f'[d^n]\frac{d^n}{2} > f\left[\frac{d^n}{2}\right] + f'[d^n]\frac{m}{2}\frac{\theta^M}{f[d^n]}$ because $m \geq d^n$ and $\theta^M > f[d^n]$. Therefore, (A.52) holds a fortiori. Hence, the markup distortion exceeds the productivity distortion and

[25] 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.

consequently, the number of firms in the market equilibrium with migration is too large.[26] For the case of $\lambda = \bar{\lambda}$ we can thus state: $\psi[m] > \psi[m, \bar{\lambda}] > \frac{\theta^M[m,\bar{\lambda}]}{f[d^n]} = \frac{\theta[m]}{f[m]}$. Compared to the autarky laissez-faire equilibrium, the markup distortion is reduced but still larger than the productivity distortion. Hence, the number of firms is closer to the planner's optimal choice and welfare must be higher.

Proof that $\theta^M \leq f\left[\frac{d^n}{2}\right]$. Using the expression for θ^M in (26) 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

$$\theta^M = \frac{1}{m}\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d \leq \frac{d^n}{m} f\left[\frac{d^n}{2}\right] + (1-\lambda)\frac{d^m}{m} f\left[\frac{d^m}{2}\right]. \quad \text{(A.53)}$$

Since $d^n + d^m = m$, we have $\theta^M \leq \frac{d^n}{m} f\left[\frac{d^n}{2}\right] + (1-\lambda)\frac{d^m}{m} f\left[\frac{d^m}{2}\right]$. 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 (23) 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.8.4 Proof of Proposition 3, part (ii)

Totally differentiating (28) yields $\hat{m} = C\hat{\lambda}$ where C is given by[27]

$$C = \frac{g_\lambda^M - \rho^T \psi_\lambda^M}{-g_m^M + \rho^T \psi_m^M + \psi^M \rho_m^T} \frac{\lambda}{m} \lessgtr 0 \qquad \text{with} \quad \text{(A.54)}$$

$$g_\lambda^M = \frac{L\theta_\lambda^M}{L\theta^M - \alpha N} - \frac{L\theta^M}{\left(L\theta^M - \alpha N\right)^2} L\theta_\lambda^M \geq 0 \quad \text{and} \quad \theta_\lambda^M = -\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d \leq 0 \quad \text{(A.55)}$$

$$g_m^M = \frac{L\theta_m^M}{L\theta^M - \alpha N} - \frac{L\theta^M}{\left(L\theta^M - \alpha N\right)^2}\left(L\theta_m^M + \frac{\alpha N}{m}\right) < 0 \quad \text{and} \quad \theta_m^M = \frac{1}{m}\left(f[d^n] - \theta^M\right) < 0 \quad \text{(A.56)}$$

$$\psi_\lambda^M = -\frac{1}{(\eta^M)^2}\eta_\lambda^M > 0 \qquad \text{with } \eta_\lambda^M \text{ as in (A.49)} \quad \text{(A.57)}$$

$$\psi_m^M = -\frac{1}{(\eta^M)^2}\eta_m^M > 0 \qquad \text{with } \eta_m^M \text{ as in (A.48)} \quad \text{(A.58)}$$

---

[26] 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.51 still describes a global maximum and that the social welfare function is monotonically increasing in the relevant range. Details of the proof are available upon request.

[27] 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.

$$\rho_m^T = \frac{1}{2\gamma H} > 0. \tag{A.59}$$

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 barriers fall.

The effect on average income is ambiguous. While the partial effect of lower migration barriers 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 barriers fall, as the decrease in prices overcompensates the potential decrease in average income. We show this by log-differentiating real income $\frac{\theta^M}{p} = \frac{L\theta^M - \frac{\alpha H}{m}}{L}$ using (28):

$$\mathrm{d}\ln\left[\theta^M/p\right] = \frac{\partial \ln\left[\theta^M/p\right]\lambda}{\partial\lambda}\hat{\lambda} + \frac{\partial \ln\left[\theta^M/p\right]m}{\partial m}\hat{m} \tag{A.60}$$

with

$$\frac{\partial \ln\left[\theta^M/p\right]}{\partial\lambda} = \frac{L\theta_\lambda^M}{L\theta^M - \frac{\alpha H}{m}} < 0 \qquad \text{and} \qquad \frac{\partial \ln\left[\theta^M/p\right]}{\partial m} = \frac{L\theta_m^M + \frac{\alpha H}{m^2}}{L\theta^M - \frac{\alpha H}{m}} > 0. \tag{A.61}$$

In these equations $\theta_\lambda^M = -\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d < 0$ and $\theta_m^M = \frac{1}{m}\left(f[d^n] - \theta^M\right) < 0$. It follows from (A.51) that $\frac{\partial \ln\left[\theta^M/p\right]}{\partial m} > 0$ in the relevant range. 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.54) and (A.61) to rewrite (A.60) as

$$\begin{aligned}\mathrm{d}\ln\left[\frac{\theta^M}{p}\right] =& \frac{\lambda}{(L\theta^M - \alpha N)\left(-g_m^M + \rho\psi_m^M + \psi^M\rho_m\right)} \\ &\times \left[\left(L\theta_m^M + \frac{\alpha N}{m}\right)\left(g_\lambda^M - \rho\psi_\lambda^M\right) + \left(-g_m^M + \rho\psi_m^M + \psi^M\rho_m\right)L\theta_\lambda^M\right]\hat{\lambda}.\end{aligned} \tag{A.62}$$

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

$$\left(L\theta_m^M + \frac{\alpha N}{m}\right)g_\lambda^M = \left[\frac{L\theta_m^M + \frac{\alpha N}{m}}{L\theta^M - \alpha N} - \frac{L\theta^M\left(L\theta_m^M + \frac{\alpha N}{m}\right)}{(L\theta^M - \alpha N)^2}\right]L\theta_\lambda^M \text{ and} \tag{A.63}$$

$$L\theta_\lambda^M g_m^M = \left[\frac{L\theta_m^M}{L\theta^M - \alpha N} - \frac{L\theta^M\left(L\theta_m^M + \frac{\alpha N}{m}\right)}{(L\theta^M - \alpha N)^2}\right]L\theta_\lambda^M, \tag{A.64}$$

we can reduce the expression in squared brackets on the right-hand side of (A.62) to

$L\theta_\lambda^M \left(\frac{\frac{\alpha N}{m}}{L\theta^M - \alpha N} + \psi_m^M \rho + \rho_m \psi^M \right) - \left(L\theta_m^M + \frac{\alpha N}{m} \right) \rho \psi_\lambda^M$. This is negative since $\theta_\lambda^M < 0$ and $\psi_\lambda^M > 0$. Hence, a decrease in λ raises real income also if it leads to exit of firms. This completes the proof.

A.8.5 Uniqueness of $\bar{\lambda}$

The prohibitive level of migration cost $\bar{\lambda}$ satisfies $d^m \left[m[\bar{\lambda}], \bar{\lambda} \right] = 0$ with $m[\bar{\lambda}]$ denoting the solution to $g^M[m, \bar{\lambda}] = \rho^T[m]\psi^M[m, \bar{\lambda}]$. We prove uniqueness of $\bar{\lambda}$ by showing that $d^m \left[m[\lambda], \lambda \right]$ is differentiable in λ and strictly decreasing in λ at the solution point $d^m[m[\bar{\lambda}], \bar{\lambda}] = 0$, implying that the solution point is unique. Differentiability follows from

$$\frac{\mathrm{d}d^m}{\mathrm{d}\lambda} = \frac{f[d^m]}{f'[d^n] + (1-\lambda)f'[d^m]} + \frac{f'[d^n]}{f'[d^n] + (1-\lambda)f'[d^m]} \frac{m}{\lambda} C,$$

with C given in (A.54). The strictly negative slope at the solution point follows from the fact that $C < 0$ at $\lambda = \bar{\lambda}$ because $\theta_\lambda^M = 0$ at $\lambda = \bar{\lambda}$; see (A.55). This completes the proof.

A.8.6 Robustness with respect to the specification of migration barriers

The proof of Proposition 3 reveals that our results are valid for more general specifications of migration barriers. The positive welfare effect of the potential of migration established in Proposition 3 part (i) stems from a first-order welfare gain due the reduction of the markup distortion. Hence, the validity of Proposition 3 part (i) is maintained, provided that the excess-entry property of the autarky equilibrium is preserved and $\eta_\lambda^M < 0$. The proof of Proposition 3 part (ii) shows that positive welfare gains from lower migration barriers occur for every specification of the cost of migration for which, in addition to excess entry and $\eta_\lambda^M < 0$, it holds that $\theta_\lambda^M < 0$.