Heterogeneous Workers, Trade, and Migration*

Inga Heiland[a] Wilhelm Kohler[b]

University of Oslo
SSB, CESifo, CEPR

University of Tuebingen
CESifo, GEP and IAW

February 10, 2022

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 markdowns 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 markdowns and lower goods prices and thus to higher welfare in both countries. Our model 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 University of Michigan International Economics Seminar, 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). We are grateful for very helpful comments from the editor as well as from an associate editor and two referees.

1 Introduction

Economic well-being crucially depends on how well workers' characteristics are matched with firms' production 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. Trade increases firms' monopsony power on the labor market while worsening the average quality of worker-firm matches, yet the standard result of gains from trade 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.

The extant literature on worker-firm matching has mostly focused on how workers with a higher or lower level of skills are matched with firms with a higher or lower productivity level.[1] Arguably, however, there is a lot of worker heterogeneity in characteristics that cannot be ranked according a criterion like the level of skills, which prompts us to analyze heterogeneity in skill *types*, rather than skill *levels*, or simply *horizontal* heterogeneity. In a similar vein, there is much heterogeneity among firms that defies a clear ranking, such as the type of workers required for production. Clearly, matching under horizontal heterogeneity must follow a logic different from that under vertical heterogeneity. What does this this logic imply for earnings inequality? What are the implications of horizontal heterogeneity for trade and migration? Conversely, how does openness to trade and migration affect the quality of worker-firm matching based on horizontal heterogeneity?

In this paper, we use a stylized general equilibrium model to answer these questions. Our model relies on the notion of *distance* between skill types. A skill type is best thought of as a certain combination of abilities for a potentially vast array of different tasks. Workers with the same level of skills may represent different skill types, and any two workers' skill types may be more distant or less distant from each other, depending on the exact combination of abilities. To give an example, consider an architect and a fashion designer who might be regarded as having the same level of skills but embody different types of skills. The same can be said when comparing a bridge building engineer with

[1] See Section 2 for a detailed review of the literature.

an architect or with a fashion designer. Moreover, the bridge building engineer arguably represents a skill type more similar (or at a smaller distance) to that of an architect than to that of a fashion designer.

Allocation of skill types to firms may seem trivial: architects should work in architecture, bridge building engineers should build bridges, and so on. However, closer inspection reveals that different people calling themselves architects differ in their skill types, some being marginally closer to a bridge building engineer and others being marginally closer to a fashion designer. Indeed, looking at workers in this way we find that occupations are mere labels for certain ranges of skill types describing a much finer underlying variety of worker characteristics. We take this to the extreme in treating the skill distance as a continuous variable. This proves convenient as it makes the analysis of horizontal skill heterogeneity amenable to tools familiar from the theory of spatial competition. More specifically, we follow Amiti and Pissarides (2005) in adapting the well-known circular model of spatial competition due to Salop (1979). Instead of firms choosing a location in geographic space, we have firms choosing a position in skill space (a particular way to produce a particular product), called a firm's *ideal* skill type.

The income a worker earns in a specific firm depends on the distance between the worker's skill type and the firm's ideal type; a smaller distance makes for higher earnings. With a given and finite number of firms, workers maximize their earnings by choosing the firm that offers the highest wage for their particular type of skill. This has three important consequences. First, in this labor market each firm has wage setting power. Posting a higher wage means the firm will attract workers with skill types farther away from its ideal type, which were previously employed in "neighboring" firms. Secondly, even if all workers have the same skill level, they receive different earnings. Thus, there is residual earnings inequality. And finally, the number of firms becomes crucial for both, the degree of inequality and the average quality of worker-firm matches. Other things equal, fewer firms means larger firms and firm expansion always means using less suitable skill types who earn less.

Adding a standard element of product differentiation ("love of variety") on goods markets, we use our model to "re-tell" the story of gains from trade as well as the story of trade and wages. We also use it to tell a new story of gainful two-way migration, i.e., both emigration and immigration of workers between completely symmetric countries. Key ingredients of both stories are the effects that firm entry and exit, guided by a zero-profit condition, have on the degree of monopsony power in the labor market and on the quality of worker-firm matching. Intuitively, firm exit increases monopsony power and it worsens the average quality of matching, because it increases the skill distance between firms. We distinguish between the extensive and the intensive margin of liberalization. The *extensive margin* of liberalization of trade or migration refers to the number of countries practicing free trade or free migration between them. The *intensive margin* of liberalization involves a piecemeal reduction in the real trade cost or the cost of migration. To provide a clean and distinct description of the mechanisms involved with trade liberalization as opposed to migration, we first analyze liberalization of trade assuming closed labor markets. When dealing with liberalization of migration at either margin, we always assume an arbitrary number of countries practicing free trade, including the case of autarky.

In a nutshell, our results are as follows. Trade liberalization at either margin entails firm exit, as expected from existing models of trade under monopolistic competition, but this now has hitherto neglected detrimental effects: less competition on the labor market (and thus higher wage markdowns) and a poorer average matching quality as well as an increase in earnings inequality, measured by the gap between the highest and the lowest earning generated by worker-firm matching. However, these effects notwithstanding, extensive margin trade liberalization still unambiguously lowers firms' average cost and goods prices due to scale economies and increases all participating countries' welfare.

In contrast, intensive margin liberalization is not unambiguously welfare increasing. It is gainful only if the initial level of trade costs is not too high. High real trade costs imply low levels of imports to start with whence lower import prices from trade liberalization carry a low welfare weight. At the same time, if trade costs are high to start with, any increase in the volume of exports implies a large additional resource use for trade costs. This, together with larger wage markdowns and a poorer average matching quality caused by firm exit, is responsible for a negative welfare effect of piecemeal trade liberalization, if the initial level of trade costs is sufficiently high.

In our world of horizontal worker heterogeneity, there are incentives for two-way migration among all countries practicing open labor markets, even if these countries are perfectly symmetric, including identical worker heterogeneity. This is remarkable, since two-way migration looms large in the data (as demonstrated in the next section) but is difficult to explain. In our setup, the explanation is simple: Workers and firms in all countries find better matches by searching and hiring across borders. Migration effectively increases the pool of workers of any skill type with the effect that firms find more matches closer to their optimal type. This increases the average quality of matches and reduces the dispersion. Moreover, opening up labor markets to migration enhances competition; wage markdowns fall. As a result, liberalising migration on the extensive margin leads to an increase in welfare as well as lower earnings inequality, even if the number of firms within each country falls (as it may). The existence of a strong incentive for migration even if countries already practice free trade means that trade and migration are complements, which - by and large - is what we find in the data (see again the next section).

Opening labor markets has pro-competitive and welfare effects even if the cost of migration are so large that migration is negligibly small. The reason is that the potential of drawing on the foreign pool of workers through an increase in the wage offer makes each domestic firm's labor supply more elastic than it would be with closed labor markets. Consequently, wage markdowns as well as goods prices are lower than with closed labor markets, and zero profits require a lower number of firms, which implies a poorer average matching quality and higher earnings inequality. However, lower prices still lead to higher welfare than with closed labor markets, even though migration is negligibly small. Liberalization along the intensive margin through a reduction of the migration cost from this prohibitive level further increases welfare but has an ambiguous effect on the number of firms and on earnings inequality within each country.

The remainder of the paper is organized as follows. In Section 2, we discuss the related literature, including available empirical evidence supporting key features of our model. In Section 3, we describe the general model framework and characterize the autarky equilibrium. In Section 4, we then discuss the welfare and inequality effects of trade liberalization at the extensive and the intensive margin, respectively, at that point assuming closed labor markets. In Section 5, we introduce labor mobility and analyze the liberalization of migration. Again, we do so for both the extensive and the intensive of liberalization, in each case assuming a given (but arbitrary) level of trade liberalization at the extensive margin. Section 6 concludes.

2 Related literature and empirical evidence

2.1 Worker heterogeneity, trade and wages

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 with vertical skill heterogeneity among workers. In their models, workers with different ability levels are assigned to industries based on submodularity or supermodularity of production techniques. Log-supermodularity means that the productivity of a more able worker, *relative* to that of a less able worker, is larger in a more productive firm than in a less productive firm. In other words, a more able worker has a comparative advantage in working for a more productive firm. Log-supermodularity typically leads to positive assortative matching: more able workers are matched with more productive firms. This matching translates worker heterogeneity into wage inequality. Costinot (2009) and Costinot and Vogel (2015) develop a generalized theory of comparative advantage for linear technologies where factors are assigned to industries and sectors are assigned to countries, based on log-supermodularity of technologies. Costinot and Vogel (2010) develop a model of this kind incorporating workers distributed along a continuous variable measuring the skill level. Workers are assigned to a continuum of tasks (equivalently: goods or industries), again based on log-supermodularity of technology.[2] The focus in Costinot and Vogel (2010) lies on how within-country wage inequality is affected by trade between countries differing in their skill endowments and the skill bias of their technologies. Wage inequality is measured by the return to skill (or the steepness of the wage function), i.e., the inequality explained by workers' skill levels.

Grossman et al. (2017) analyze wage inequality across skill levels, occupations (workers, managers), and industries in a model featuring matching of heterogeneous workers to heterogeneous managers. As in Costinot and Vogel (2010), worker heterogeneity is described by a density over a continuous ability variable. The same holds true for managers, and log-supermodularity generates positive assortative matching of workers to managers. In addition, worker-manager teams sort themselves into one of two sectors differing in their intensity of using workers. As a result, the model is able to address within-industry-and-occupation inequality, again explained by vertical skill heterogeneity. Sorting, and thus the distribution of earnings, is driven by the prices of the two goods produced in the two sectors. Hence, the effect of trade liberalization on inequality along the different margins is a-priori ambiguous. Sampson (2014) introduces skill-level heterogeneity of workers into a Melitz (2003) model. In this setting, trade liberalization increases the productivity level

[2] There is a long tradition of assignment models, unrelated to trade and starting with Roy (1951), that investigates assignment of skill types, defined as incorporating certain bundles of different skills, to jobs with multi-dimensional skill requirements; see Mandelbrot (1962), Rosen (1978), Sattinger (1993), Teulings (1995), Moscarini (2001), and Lazear (2009). An important recent contribution is Lindenlaub (2017) who analyzes assignment of workers characterized by multidimensional heterogeneity.

of the least productive of all surviving firms, leaving workers with a more productive range of firms to match with. With log-supermodularity and positive assortative matching, this implies a rise in wage inequality, again of the explained sort.

This literature leaves a gap in that horizontal heterogeneity of workers is not amenable to the logic of supermodularity for the simple reason that skill types, unlike skill levels, defy any notion of ranking or "having more or less".[3] Our paper attempts to fill this gap by analyzing the mechanism of worker-firm matching for horizontal heterogeneity. To model worker-firm matching we use a formal approach, borrowed from Amiti and Pissarides (2005), that employs logic from the circular model of spatial competition due to Salop (1979). A cornerstone of that logic is entry of firms at certain locations. In our case location simply means a certain technology featuring a certain ideal skill type. This is analogous to firms choosing idiosyncratic combination of product characteristics when facing Lancasterian consumers whose ideal product varieties are represented by points on a circle, as in Economides (1989) and Vogel (2008).

Allowing for endogenous firm entry is crucial because under plausible conditions, entry is likely to be inefficient; see Mankiw and Whinston (1986). Indeed, our analysis shows that horizontal worker heterogeneity generates entry distortions in and of its own and that the question of efficient firm entry is also a question of efficient worker-firm matching. This is in sharp contrast to the aforementioned literature where assortative matching as such is always efficient.[4] We discuss the inefficiency of firm entry in the context of a model incorporating a love-of-variety-type goods market (in addition to a labor market with skill-type variety), thus contributing to the new trade literature going back to Krugman (1979) where inefficient firm entry has recently received renewed attention; see Baldwin (2005); Dhingra and Morrow (2019); Bilbiie et al. (2019); Caliendo et al. (2015, 2021); Jung and Kohler (2021). We readdress welfare for love-of-variety economies under horizontal worker heterogeneity under different regimes of trade openness, thus contributing to the modern gains from trade literature (see, e.g., Arkolakis et al. (2012, 2019); Melitz and Redding (2015).

[3] As pointed out in the introduction, a skill type must be conceptualized as a combination of a worker's abilities. This is akin to the definition of a good as a combination of certain characteristics in Lancaster's theory of consumption; see Lancaster (1966). However, we do not explicitly model these characteristics but simply define a firm by a certain ideal skill type, much like a consumer is characterized by an ideal product variety in Lancaster's model of intra-industry trade; see Lancaster (1980).

[4] 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.

Among the distortions generated by horizontal worker heterogeneity is monopsony in the labor market, which is a pervasive empirical phenomenon; see Manning (2003). Yet, this phenomenon takes somewhat of a backseat in the trade literature. In MacKenzie (2019), monopsony power derives from worker-firm-specific productivity draws in a random utility model. As in our model, this leads to matching of a continuum of workers with a finite set of heterogeneous firms. The set of firms, as well as their set of export markets are exogenously given, but firms are suboptimally small. Importantly, the size distortion is particularly strong for more productive firms. 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. (2021) use a similar microfoundation for monopsonistic labor markets that relies on discrete choice based on random utility. As in MacKenzie (2019), firms are heterogeneous, and each worker draws an idiosyncratic random utility component capturing the worker's preference for the amenities offered a specific firm. Upward-sloping firm-specific labor demand generates a novel incentive for two-way 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. A similar microfoundation of monopsony on the labor market is also employed by Jha and Rodriguez-Lopez (2021). In addition to wage inequality driven by Melitz-type firm heterogeneity, they highlight a novel mechanism which they call workers' "love of firm variety": the larger the number of firms, the more likely will workers find firms they would ideally want to work for.[5] Obviously, this resembles our matching-quality effect. In contrast to these random utility models, however, our explanation of monopsony power and matching quality squarely lies on the supply side of the economy, generated by horizontal skill differentiation among workers. In addition, we investigate the consequences of such heterogeneity for both migration and trade.

[5] Egger et al. (2021) call this effect a "market thickness" effect but deliberately shut this channel down by a suitable normalization of preferences.

2.2 Trade and residual wage inequality

Assortative worker-firm matching generates a wage function describing how worker heterogeneity translates into wage inequality.[6] For this reason, assortative matching models have acquired workhorse status in the literature addressing the relationship between trade and wage inequality; see Helpman (2017) for a survey. However, empirical evidence tells us that there is a lot of *residual* wage inequality, i.e., inequality that cannot be explained by observable worker characteristics of the type assumed by assortative matching models.

There is a strand of literature that explains such residual wage inequality through heterogeneity of the firms with which ex-ante homogeneous workers are matched. For instance, Egger and Kreickemeier (2009) demonstrate that fair-wage preferences generate wage inequality among perfectly homogeneous workers, if - plausibly - these preferences are sensitive to the profit of the firm a worker works for. Of course, in this modeling environment with full homogeneity of the work force, there is no assortative matching. But there is *endogenous* residual wage inequality since the rent sharing generated by the fair-wage constraint implies that more productive and larger firms pay higher wages, which is also what we observe in the data. A similar outcome is generated if the fair-wage paradigm is replaced by the efficiency-wage paradigm, as in Davis and Harrigan (2011).[7]

Egger and Kreickemeier (2012) propose a model with individuals who have different levels of ability only when working as managers, but are perfectly homogeneous as workers when matched with firms run by managers. More able individuals make larger profits when running a firm, and this generates occupational sorting into managing firms. Wages are again subject to a fair-wage constraint as in Egger and Kreickemeier (2009), leading to within-group inequality among both managers and workers. Whereas the former is explained by underlying ability levels, the latter is again endogenous residual inequality.[8]

Wage inequality that essentially mirrors heterogeneity of firms where workers happen to be employed may also be generated when job formation is subject to costly search. For instance, Helpman et al. (2010) assume firms facing a work force which is perfectly homogeneous ex ante, i.e., prior to matching, but once workers are matched with firms they draw a stochastic ability level which is specific to the firm. All firms have a screening

[6] Note that log-supermodularity skews the wage distribution to the right, because assortative matching magnifies the productivity difference between any two workers with different skill levels; see Hao (2017).
[7] A model like this is brought to Indonesian data in Amiti and Davis (2011).
[8] In Kohl (2020), a similar setup is used to explore redistributive tax policies, but assuming away any fair wage constraint whence all (homogeneous) workers receive the same wage.

technology allowing them to control the minimum ability level among their hires. Since more productive firms screen more ambitiously, they also end up with a work force with a larger average ability level. Multilateral bargaining between workers and the firm leads to wages that depend on the firm's productivity. From an ex-ante perspective, earnings inequality is endogenous residual inequality, although ex post the wage rate a worker receives is positively related to the average ability of all workers employed by the same firm. As in other models featuring Melitz (2003)-type firm heterogeneity, trade impacts wage inequality through affecting the productivity distribution of firms and earnings inequality through a reallocation of labor between firms with different productivity levels.[9]

The common features of these models of residual wage inequality are that all workers employed by a firm receive the same wage and that all workers employed by any one firm have the same (ex-ante) ability level. Moreover, vertical firm heterogeneity leads to endogenous residual wage inequality, which may be called *within-group*, but *between-firm* inequality. Empirical evidence for this type of between-firm inequality is relatively well established; see Song et al. (2018). In contrast, our model delivers *within-firm* inequality among workers who differ in their skill types rather than their skill levels. To motivate our analysis, we next look at empirical evidence of within-firm (or within-plant) inequality.

Using Swedish and Brazilian worker-level data, respectively, Akerman et al. (2013) and Helpman et al. (2017) estimate a Mincer equation that includes a firm-fixed effect in order to decompose the wage inequality observed within different sector-occupation cells into a within-firm component and a between-firms component, controlling for (vertical) worker characteristics. For Brazil in 1994, the within-firm component and the between-firms component contribute equally (37 and 39 percent) to the within-sector-occupation inequality observed in 1994. For Sweden in 2001, the within-firm component explains by far the largest part (65 percent) followed by 19 percent for the between-firm component. For Sweden the within-component has increased by 75 percent from 2001 to 2007, whereas for Brazil it has fallen by 12 percent between 1986 and 1995. Using similar methods, Becker et al. (2019) use German linked plant–worker data (LIAB) data for 1996-2014 to

[9] Autor et al. (2020) provide evidence suggesting that between-firm reallocation of activity from less productive to highly productive ("superstar") firms is responsible for the decline in the labor share in the US observed over recent decades. The underlying mechanism, they argue, is that under reasonably general conditions high-productivity firms also charge higher price-cost markups, which implies that they have lower firm-specific labor shares than low-productivity firms. In this way, the aggregate wage share is driven by a composition effect. However, residual wage inequality of the type described above runs counter to this mechanism in that "superstar" firms pay higher wages than low-productivity firms although both use the same type of labor.

estimate the within-plant component at 71 percent of residual wage inequality.[10]

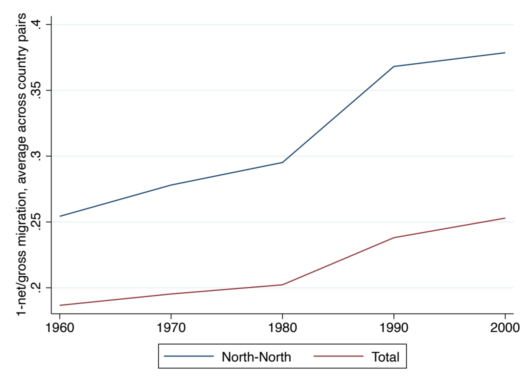
Horizontal worker heterogeneity has important consequences for firm-level employment and earnings dynamics. For instance, our model of within-firm inequality implies that firms wanting to increase employment have to reach out to workers with skill types farther away from their ideal types, and closer to other firms' ideal types. There is empirical evidence also for this aspect of our model. Becker et al. (2019) show that residual wage inequality is larger in larger plants. Using German employer-employee data, Gulyas (2018) finds that the worker-firm matching quality is below average for newly hired workers in growing firms. Evidence on earnings dynamics is presented in Autor et al. (2014). They estimate the effect of US workers' exposure to increased imports from China on earnings per year employed during the time span 1992-2007, separately for workers who have changed the firm but remained in the same industry and those who remained with the initial firm. They find negative effects for both types of workers, but a stronger effect for workers who have left the firm. A possible explanation consistent with horizontal heterogeneity is that displaced workers are forced to find employment in firms for which their skills are less valuable.[11] Gathmann and Schönberg (2010) present evidence relating the portability of skills across jobs to a measure of distance between the skill requirements of the old and the new employer. The firm-level dynamics of earnings and employment that follow from shocks involving firm exits are, of course, determined also by costs of switching between firms. Generally, such costs will reduce the amount of switches taking place while at the same time reducing aggregate welfare, relative to a case without such costs. For simplicity, our baseline model abstracts from switching cost, but as we shall see when dealing with migration scenarios, the cost of migration in our model acts like a switching cost in the short run.[12]

[10] In their explanation of wage inequality, Becker et al. (2019) employ a circular representation of skill types similar to ours. Within-firm inequality results from firms' internal organization of production using workers with heterogeneous core abilities (our skill types). Firms with different productivity levels differ in their organization of production and, therefore, in their degree of wage inequality. Importantly, their model does not feature a monopsonistic labor market resulting from firms' locations on the skill circle.

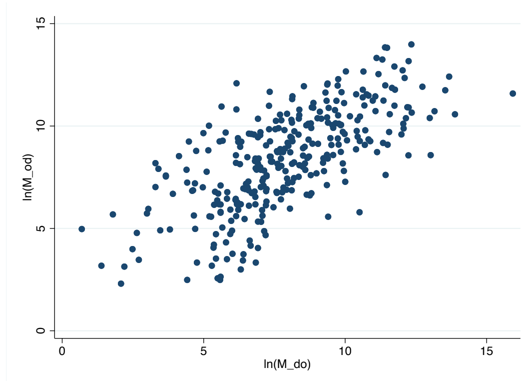
[11] Our model is designed to address long-run adjustments to globalization shocks, involving entry, exit and relocation of firms on the skill circle. However, an implication of our model in the short run is that workers displaced by shrinking or exiting firms must accept employment in existing firms for which they are worse matches, which is consistent with the evidence provided by Gulyas (2018).

[12] See Artuç et al. (2010) for an in depth analysis of the cost of adjustment to trade liberalization in a dynamic Ricardo-Viner model.

Figure 1: Two-way migration



a) Two-way migration index



b) Bilateral stocks (OECD countries, 2000)

Note: Panel a) is based on the World Bank's Global Bilateral Migration Database. North-to-North refers to migration between countries that were OECD member in 2000. Panel b) is based on the OECD DIOC database.

2.3 Two-way migration

Our paper provides novel explanations for two salient patterns in global migration data. The first is two-way migration between countries of similar standards of living, the second is complementarity between trade and migration.

Figure 1 shows how two-way migration relative to total migration has evolved over time, globally and for north-to-north (here, intra-OECD) migration. Panel a) plots the development of an index for two-way migration that resembles the Grubel-Lloyd index for inter-industry trade and is defined as $1 - \frac{|M_{odt} - M_{dot}|}{M_{odt} + M_{dot}}$, where M_{odt} denotes the stock of migrants from origin o residing in destination d at time t and vice versa for M_{dot}. The index relates net migration $|M_{odt} - M_{dot}|$ to gross migration $M_{odt} + M_{dot}$ at the country-pair level. If all migration is two-way, that is, $M_{odt} = M_{dot}$, then the index is equal to one. Conversely, it equals zero if net migration equals gross migration, that is, if no two-way migration occurs at all. Panel a) plots the average index value across country pairs. It shows that two-way migration has become more prevalent between 1960 and 2000, and is more relevant for migration among rich countries, reaching a level of close to 40 percent of total migration in 2000.

Panel b) plots the log of M_{od} against the log of M_{do} for OECD countries in 2000. The correlation between the two stocks is positive and strong (.69), supporting the relevancy of two-way migration. Admittedly, this correlation may be due to country size and symmetry in bilateral migration barriers. Yet, using residuals obtained from a regression equation explaining M_{od} and M_{do} by origin and destination fixed effects as well as the usual gravity-type covariates, the correlation coefficient remains high (.59) and statistically significant.[13] We find correlation coefficients in the same vicinity also within age groups and within education levels; see Table A.1 in the appendix.

Models featuring two-way migration typically rest on assumptions regarding differences in technologies or endowments, as in Iranzo and Peri (2009).[14] Two-way migration among similar countries proves harder to explain. Recent migration literature on sorting across destinations has invoked location-specific preferences to rationalize these flows; see, e.g., 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). We add a novel explanation of two-way migration between symmetric countries that relies on firm-specificity of skills and firms' endogenous location choices in the skill space. An important feature of migration as explained by our model is that it is the outcome of specific worker-firm matches in open labor markets, rather than anonymous movements from one country's labor force into some other country's labor force.

The second salient fact of international migration explained by our model relates to the relationship between trade and 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). Empirically, trade and migration flows often reveal patterns of complementarity. A sizeable body of empirical literature (surveyed in Felbermayr et al., 2015) attributes

[13] Specifically, we regress both M_{od} and M_{do} on bilateral distance, time difference and dummies indicating contiguity, a common language, joint EU or RTA membership, a common colonizer or current colonial relationship, or a common legal origin and then compute the correlation between the residuals from the two regressions.

[14] 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 the relative price of capital might lead to two-way migration when individuals are heterogeneous with respect to the degree of altruism.

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.

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.[15] The circumference 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 that would ideally draw on a specific skill type on the skill circle. In line with the vast majority of the related literature, we consider only symmetric location equilibria where all firms are located at equal distances on the skill circle.[16] 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 of firms determined in stage one. When deciding on entry in stage one, firms anticipate the equilibrium of stage two (subgame perfection). We assume an infinite number of potential entrepreneurs with zero outside options as well as free entry, which implies zero equilibrium profits.

[15] Our model differs from Amiti and Pissarides (2005) by featuring a more general functional form of the relationship between skill distance and the productivity of a worker. 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.

[16] 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. Symmetric location patterns are analyzed in the classical circular models of Vickrey et al. (1999) and Salop (1979). In a trade context, the symmetric location pattern is, among others, analyzed by Helpman (1981), Grossman and Helpman (2005), Amiti and Pissarides (2005), and Eckel (2009a,b). For examples of circular labor market models with symmetric firm locations, see Bhaskar and To (1999, 2003) and Hamilton et al. (2000).

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 skill type is at (arc) distance d from the firm's ideal type.[17] We assume that $f'[d] < 0$, $f'[0] = 0$, $f''[d] < 0$, and $f[d] = f[-d]$. Without loss of generality, we set $f[0] = 1$. The concavity assumption $f''[d] < 0$ will prove crucial in the subsequent analysis. It means that the "marginal penalty" from using a less suitable skill match is increasing in the distance from the firm's ideal skill type and it is justified by communication or coordination cost associated with replacing a task performed by one worker with a (more) optimal skill type by two workers with less than optimal skills. It implies that it becomes increasingly costly for the firm to employ an ever more diverse lot of workers.

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 firms positioned in her neighborhood on the skill circle as well as the relevant firm-specific worker productivity schedules $f[d]$. In contrast, we assume that firms know the distribution of skills in the neighborhood of their optimal type but are unable to identify skill differences between individual workers ex ante. The assumed information advantage on the part of workers is reasonable, given that workers may infer the demand for particular skill types from observing the product produced by the firm as well as from the job postings.[18] In effect, this assumption negates wage discrimination among workers with regard to the rate paid per efficiency unit. Workers choose employers so as to maximize individual earnings, 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.

When firms set wages and prices, the number of firms and their ideal skill type on the circle is given. We use m to denote the *half* distance between any one firm's position and the nearest neighbor to its left and to its right and henceforth refer to it as the "skill distance between firms". It is inversely related to the number of firms through $N = \frac{H}{m}$. If firm i posts a wage rate w_i per efficiency unit of labor it attracts workers up to a

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

[18] Going back to the example given in the introduction, the informational asymmetry should not be taken as meaning that firms cannot tell an architect from a fashion designer. What it means is that firms cannot identify marginal differences between worker's skills at the hiring margin.

skill distance d_r to its right, with d_r determined 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 $i+1$. The solution to this equation is denoted by $d_r[w_i, w_{i+1}, m_i]$ and we call it firm i's "skill reach" to the right. Analogously, we use d_l to denote firm i's skill reach to the left. Both skill reaches are increasing in w_i. Given a uniform distribution of workers along the skill circle, we obtain firm i's labor supply schedule as

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

This labor supply function is increasing in the firm's own wage.[19] In what follows, we shall use η_i to denote the wage elasticity of firm i's labor supply.

We describe demand by a symmetric translog expenditure function, which implies homothetic preferences and a variable demand elasticity. Both properties are desirable for our study of the welfare effects of globalization on heterogenous workers.[20] Aggregate demand for variety i is

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

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*.[21] 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

[19] 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 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).

[20] 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.

[21] The Lambert function $\mathcal{W}[z]$ is defined as the implicit solution to $xe^x = z$ for $z > 0$. It satisfies $\mathcal{W}'[z] = \frac{\mathcal{W}[z]}{(\mathcal{W}[z]+1)z} > 0$, $\mathcal{W}''[z] < 0$, $\mathcal{W}[0] = 0$ and $\mathcal{W}[e] = 1$.

A.2 has the details. Given consumers' love of 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 labor requirement β and a fixed labor requirement α, measured 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 $L_i = L^S\left[w_i, w_{i-1}, w_{i+1}, m\right]$. The restriction ensures that the 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.3. 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 take the average log price $\overline{\ln p}$ and aggregate income Y as given. The perceived price elasticity of demand for variety i then 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 markup of the price and a markdown of the wage:[22]

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

It is easy to verify that under the assumptions made, the second-order condition is satisfied. From (5) and (2), we may write the markup deriving from product differentiation as

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

[22] Firms set perceived marginal revenue equal to perceived marginal cost, implying a wage below the perceived marginal revenue product and a price above the perceived marginal cost. Hence, we speak of a "wage markdown" and a "price markup".

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 $\mathcal{W}'[\cdot] > 0$, 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 wage markdown 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 $w_i = w$. With equal wages, the skill reach of a firm on either side is equal to the half distance between two firms, that is, $d_r = d_l = m$. Symmetry in (2) simplifies the expressions for ε and δ, allowing us to write the 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. We are free to do so, since our equilibrium is homogeneous of degree zero in nominal prices. Note that $\rho'[m] > 0$ as well as $\psi'[m] > 0$. Firms' market power in 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

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

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 work force around the circle, the average matching quality 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 $Y = L\theta[m]$.

3.2 Symmetric two-stage equilibrium

We now proceed in determining the equilibrium number of firms and their distance on the skill circle by invoking free entry and imposing the zero-profit condition which reads as

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

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

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

It seems intuitive that an increase in firm size implied by a higher m leads to lower average cost $g[m]$. However, it should be noted that a higher m also lowers the average matching quality. In Appendix A.3 we demonstrate that the size effect dominates. 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.3 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.4.

We take a "veil of ignorance" view of welfare, 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 varieties while falling in the price of a typical variety of goods.

This model allows for a straightforward discussion also of income inequality. In a symmetric equilibrium all inequality is within-firm inequality of wage income among workers. Following Helpman (2017), we measure inequality by the real income gap between the best-matched worker and the worst-matched worker, $f[0] - f[m]$.[23] This measure is monotonically increasing in the skill reach, which in the equilibrium with closed labor markets is equal to m. In fact, given the assumptions of our model, the skill reach also serves as proxy for other measures of income inequality. More specifically, in Appendix A.5 we show that, given concavity of $f[\cdot]$, the variance of incomes across workers as well as the Gini coefficient are similarly increasing in m.[24]

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 the average matching quality arising from a better quality of matches in the labor market. We call this the "matching distortion." (iii) Potential entrants anticipate a certain goods price markup as well as a wage markdown, but fail to see that they receive operating profits only at the expense of incumbent firms, due to the overall resource constraint. Mankiw and Whinston (1986) have called this the "business-stealing" effect, but it is perhaps better described as "market crowding." (iv) Potential entrants fail to anticipate that their entry will reduce the magnitudes of these profit margins, 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 distortions (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-

[23] Note that all workers face the same prices and enjoy the same degree of variety in (15) above.

[24] With earnings inequality at the heart of our analysis and with our welfare function relying on the "veil of ignorance", it seems natural to also consider the Rawlsian social welfare function. In the eyes of the Rawlsian social planner welfare is given by the utility of the least fortunate worker. It is straightforward that the qualitative effect of globalization on the utility of the poorest individual is the same as the effect on average income in scenarios that increase average utility while at the same time lowering the degree of inequality (proxied by the skill reach). In general, however, the two measures of welfare may move in opposite directions. We will return below to a discussion of more general conditions under which the welfare results stated in terms of average income carry over to the welfare of the poorest individual.

stealing effect dominates the consumer-surplus effect, giving rise to excess entry. The excess-entry result will prove important for signing the welfare gains from globalization below, since these partly unfold through a mitigation of distortions.[25]

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 trade by allowing for trade between an arbitrary number of countries to be costly and looking at a marginal reduction of the trading cost. We call this the intensive margin of globalization. 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 The extensive margin of trade

We assume that there are k symmetric countries. The number of varieties available worldwide is 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

[25] Returning to our discussion of the Rawlsian welfare criterion, we might now state that a globalization scenario that leads to higher welfare for the average worker but also to higher inequality due to firm exit will improve the welfare of the poorest individual provided that a Rawlsian social planner would prefer a smaller number of firms compared to the laissez-faire equilibrium. In our setting, excess-entry from a Rawlsian planner's point of view obtains only under additional restrictions on the parameter space and/or the shape of $f[\cdot]$. More details are available upon request.

$\rho^T[m/k] := 1 + \frac{1}{k\gamma N[m]} = 1 + \frac{m}{k\gamma H}$ as the goods price markup under free trade. It is obvious that $\rho^T[m/k] < \rho[m]$ for $k > 1$.

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

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

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 markdown, coupled with a lower price markup, but goods prices are unambiguously lower. (iii) Each country suffers a fall in the average matching quality, implying lower average income. (iv) On average, individuals in each country enjoy a higher welfare. (v) Income inequality increases.*

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

The increase in variety, the pro-competitive effect on goods markets and the decline in prices due to larger firms producing at lower average cost are standard results in trade models with monopolistic competition and endogenous markups; see Krugman (1979). 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 markdown. 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 average matching quality. However, the variety effect, the gains from scale and the pro-competitive effect on the goods market more than compensate for the adverse effects on the labor market, making the economy better off under free trade than under autarky.[26] The positive welfare effect involves two channels. The first runs through higher variety and lower goods prices. In addition, there

[26] Note that a worsening of the average matching quality doesn't make firms less productive since the scale effect dominates the matching quality effect, as pointed out above. The productivity of the firm is measured by average cost $g[m]$, which unambiguously falls in this scenario.

is a positive first-order effect deriving from firm exit since the autarky equilibrium features excess entry. Exit of firms reduces the lower bound of earnings. Since the upper bound of earnings is fixed at $f[0] = 1$ this entails an unambiguous increase in income inequality.

4.2 The intensive margin: piecemeal trade liberalization

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

$$\max_{w_i, q_i} \{\delta_i Y + (k-1)\delta_i^* Y - w_i(\alpha + \bar{q}_i)\} \tag{18}$$
$$\text{s.t.:} \quad \bar{q}_i = q_i + (k-1)\tau q_i^* \quad \text{with} \quad q_i \geq 0,\ q_i^* \geq 0 \quad \text{and} \quad \alpha + \bar{q}_i = L^S[w_i, \boldsymbol{w}_{-i}, \boldsymbol{m}_i],$$

where the domestic expenditure share falling on domestic good i is

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

In this expression, $\overline{\ln p} = \frac{1}{kN}\left(\sum_{j=1}^{N} \ln p_j + \sum_{j^*=1}^{(k-1)N} \ln p_{j^*}\right)$ denotes the average log price and j (j^*) indexes domestic (foreign) firms. A perfectly analogous expression holds for δ_i^*, the share of foreign expenditure falling on domestic good i.[27] Due to symmetry, 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 wages normalized to unity then implies the following optimal pricing conditions:

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

[27] 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.

The labor market clearing condition is

$$N[m]\left(\alpha+q[p,p^*,m]+(k-1)\tau q^*[p,p^*,m]\right)=L\theta[m]. \tag{21}$$

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 the average price in the economy. The equilibrium skill distance between firms and the corresponding skill reach, 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 $k\geq 2$ symmetric 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) The average welfare level across individuals rises for sufficiently low initial levels of τ, but it falls for sufficiently high initial levels of τ. (iv) Income inequality increases.*

Proof: A formal proof is relegated to Appendix A.7.3.

To obtain an intuition for this proposition consider the welfare differential

$$\mathrm{d}\ln V=\left(\frac{\partial\ln\theta}{\partial\ln m}-\frac{\partial\ln P}{\partial\ln m}\right)\mathrm{d}\ln m-N\delta\,\mathrm{d}\ln p-(k-1)N\delta^*\mathrm{d}\ln p^*. \tag{22}$$

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 markdown which depends on m, the skill distance between firms, through $\psi[m]$. In addition, m determines the average matching quality 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 the average matching quality $\theta[m]$, it lowers the number of varieties, and it contributes to higher domestic goods prices through a higher wage markdown. Crucially, this firm exit effect is magnified by a higher resource use from a higher export volume via the the resource constraint (21), since $\tau - 1$ additional efficiency units of labor are dissipated by trade cost per unit exported. 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 markdown 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 $\tau < \tilde{\tau}$.[28]

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 view of the positive relationship between firm size and within-firm wage inequality implied by horizontal worker heterogeneity, it appears likely that the impact of trade liberalization on the economy-wide level of wage inequality depends on the firm-size distribution.[29] With firm heterogeneity as in Melitz (2003), firm exit following a change from autarky to trade will be concentrated on the least productive (and smallest) firms, which implies a reallocation of workers to larger firms with a higher internal dispersion of worker incomes, contributing to economy-wide income inequality. The inequality effect of piecemeal trade

[28] 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 elasticity of substitution and a finite prohibitive real trade cost.

[29] The higher skill reach of more productive (larger) firms is formally analogous to the higher skill reach of a firm posting a higher wage; see above. A larger within-firm wage inequality for larger firms is documented empirically in Becker et al. (2019).

liberalization is 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.[30] 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 International migration

Whatever the degree of commodity market integration between countries, there is an economic rationale for international migration, because workers in low-quality matches with domestic firms are likely to find foreign firms that provide a better skill match. Similarly, with open labor markets firms will want to make sure they hire good skill matches with foreign workers before reaching out to less suitable domestic ones. These migration incentives obtain, simultaneously for all countries, even if the type of horizontal skill differentiation is literally the same everywhere, so that all countries' labor markets offer the exact same skill types with the same uniform distribution, and if firms in all countries have the same technology. This is, indeed, the case we want to analyze in this section. Importantly, this migration is targeted in nature in that a migrant worker moves to a specific foreign firm that offers the best match for her skill type.

In the following analysis, we continue using k to denote the number of countries practicing free trade. In perfect analogy to our analysis of trade, we analyze migration at both the extensive and the intensive margin. The extensive margin relates to the number of countries practicing free migration, denoted by ℓ, and the intensive margin refers to a

[30] An aggravation of wage inequality as a result of such a trade scenario is also found in Sampson (2014), but it should be noted that the mechanism is different. In Sampson (2014) higher between-firm inequality results from workers with different skill levels being matched to a set of more productive firms (lowest productivity firms truncated away by exit) under conditions of log-supermodularity (see Section 2.1 above). Here, the selection effect as such is the same but higher inequality follows from reallocation to more productive firms featuring higher within-firm inequality.

reduction in the cost of migration between any two countries. When investigating these migration margins below, we allow for a flexible regime of trade integration by assuming that each of the ℓ countries practicing open labor markets also practices free trade with k countries. Note that k may be smaller or larger than ℓ, and k may well be equal to one (autarky on goods markets). However, since we focus on symmetric equilibria, we require k to be the same for each of the ℓ countries practicing free migration.

A key effect of migration in our model is to enlarge the supply of labor. Opening up to migration thus resembles an increase in the population of an autarkic economy. Studying goods markets based on a circular product space, Salop (1979) and Hummels and Lugovsky (2009) have shown that a larger population provides for an increase in variety through firm entry and for lower prices thanks to enhanced competition. As we demonstrate below, both effects play out in our labor market. However, since we study an increase in labor supply without a simultaneous increase in demand, the general equilibrium effect on product variety is decoupled from the number of firms competing in the labor market and thus ambiguous a priori.[31]

5.1 The extensive margin of migration

First, consider the extensive margin of migration. A symmetric equilibrium for any number ℓ of symmetric countries sharing a common labor market with costless migration is the same as the trading equilibrium described above but with a larger labor force equal to ℓL. We continue to denote with $N = H/m$ the number of *domestic* firms in each country. With free migration, the total number of firms drawing on a common labor market (with the same skill circle) is $\ell N = H\ell/m$.[32] At the same time, the labor force over any range of skill types along the circle is multiplied by ℓ, relative to the case of closed labor markets.

Integrating labor markets affects a firm's environment in two ways: through a higher supply of labor efficiency units and through more intense competition on the labor market. Specifically, the labor supply per firm is equal to

[31] With free migration and free firm entry there also is the theoretical possibility of agglomeration equilibria where some countries are abandoned altogether. We rule out such equilibria as unlikely outcomes but note in passing that the equilibrium where all individuals and firms from ℓ countries agglomerate in a single country is formally identical to the case of ℓ countries practicing both free migration and free trade with each other and free trade also with some number k' of outside countries.

[32] Remember that the circumference of the circle is $2H$ and, hence, m/ℓ denotes the *half* distance between any two neighboring firms on the skill circle.

$$L^{S,L} = \frac{\ell L}{\ell N}\theta[m/\ell] = \frac{mL}{H}\theta[m/\ell]. \tag{23}$$

For $\ell = 1$, $L^{S,L}$ collapses to the autarkic labor market of the previous sections; see (1) and (11). Note that each firm's skill reach is now reduced to m/ℓ. Notionally holding m (and thus the number of firms per country) constant, labor supply *per firm* is larger in the integrated labor market with $\ell > 1$ countries. This is due to the fact that the average matching quality $\theta[m/\ell]$ improves as firms exchange poorly matched domestic workers with high-quality matches from abroad whilst the mass of workers per firm is unchanged. Accordingly, our measure of the average cost becomes

$$g^L[m,\ell] = \frac{L\theta[m/\ell]}{L\theta[m/\ell] - \alpha H/m} \quad \text{with} \quad g_\ell < 0, \tag{24}$$

which is falling in ℓ. Intuitively, a larger supply of efficiency units with an unchanged fixed cost per firm allows all firms to become larger and thus to move down their average cost curves. Again, for $\ell = 1$ this collapses to $g[m]$ as defined in (13).

The second channel through which open labor markets affect the firm's decision problem is increased competition in the labor market. This is reflected in a higher perceived elasticity of labor supply. Invoking (10), this elasticity may be written as

$$\eta^L = \frac{f[m/\ell]}{-2f'[m/\ell]F[m/\ell]}, \tag{25}$$

which increases in ℓ. We write $\psi^L[m/\ell] := (\eta^L + 1)/\eta^L$ for the wedge between marginal cost over the common wage of all countries with open labor markets. This markdown is falling in ℓ, which we refer to as the pro-competitive effect of integrated labor markets.

To summarize, if all countries had the same number of firms as with closed labor markets, then pooling labor markets of ℓ countries through free migration would imply lower average cost as well as lower wage markdowns than under closed labor markets. However, the number of firms (and thus the distance between them) will not be the same, because these changes potentially disturb the zero-profit condition that governs firm entry. We know from above that the markup on goods prices depends on the number of countries practicing free trade and on the number of firms (varieties) per country; see (16). The zero-profit logic of (17) therefore requires that the number of firms adjusts according to

$$g^L[m,\ell] = \rho^T[m/k]\psi^L[m/\ell]. \tag{26}$$

The following proposition summarizes the comparison between closed labor markets, $\ell = 1$, and complete integration of labor markets among $\ell > 1$ countries, assuming that an arbitrary number k of countries practice free trade.

Proposition 3. *A complete integration of $\ell > 1$ symmetric countries' labor markets has the following effects on these countries, relative to closed labor markets ($\ell = 1$): (i) There is a higher average matching quality. (ii) Goods prices are lower. (iii) There is higher real income as well as greater welfare for individuals on average, coupled with lower income inequality. (iv) The change in the number of firms per country is ambiguous. These effects obtain independently of whether goods markets are autarkic or characterized by free trade with an arbitrary number of k countries.*

Proof. The proof of this proposition is found in Appendix A.8.1.

The mechanisms responsible for gains from migration are new to the literature: a reduction in the degree of firms' monopsony power on the integrated labor market and a higher average quality of matching obtained by each firm. Both are not immediately obvious since the number of firms in each country may well fall (part iv of the proposition), which contributes to an increase in the skill distance between firms. However, as we demonstrate in the proof, if an increase in the number of countries with integrated labor markets does lead to firm exit, then the associated increase in m is lower, in relative terms, than the increase in ℓ. Consequently, the skill reach m/ℓ falls even in this case, which in turn implies a higher average matching quality and thus higher expected earnings for workers. Of course, welfare additionally depends on goods prices and the variety of goods available. Goods prices move in line with average cost. As we have seen above, average cost is falling in ℓ, holding m constant. But the aforementioned result that m/ℓ always falls is sufficient for average cost to fall also in general equilibrium with an endogenous adjustment of m. The variety effect is positive if there is firm entry, and it is negative if there is firm exit. However, in this latter case we know from the excess-entry property of the equilibrium demonstrated above that the positive price effect on welfare dominates the negative variety effect. As the skill reach falls, earnings inequality measured by the income gap between the best and the worst match declines.

5.2 The intensive margin of migration

By the intensive margin of migration, we mean a change in the magnitude of migration that is caused by a change in the migration cost. To analyze the intensive margin of migration, we proceed in two steps. First, we describe how a firm's labor supply in an open labor market is affected by the cost of migration, conditional on the number of (domestic and foreign) firms and thus conditional on the skill distance between firms. We show that the partial effects of lower migration cost on wage markdowns and on the quality of matches are qualitatively similar to the effects of labor market integration along the extensive margin. In the second step we then endogenize the number of firms by imposing the zero-profit condition. When doing so, we let the migration cost vary continuously from the prohibitive level all the way down to zero. Again, we allow for a flexible degree of commodity market integration, assuming that every country engages in free trade with $k \geq 1$ other countries. However, for tractability, we restrict the number of countries with mutually open labor markets to two.

Thus, consider two perfectly symmetric countries opening their labor markets to migration in either direction, and assume a migration cost equal to λ, with $0 \leq \lambda \leq 1$, so that moving to the other country reduces a worker's productivity down to a fraction $1-\lambda$ of her domestic earnings, assuming the same skill distance to the domestic and the foreign firm. We study a symmetric location pattern where any firm's two neighbors are located in the other country. Figure 2 depicts such a symmetric location pattern where a domestic firm is located at s_0, at distance m from a foreign firm on either side and at distance $2m$ from the nearest domestic competitor on either side.[33]

For now, we treat m as exogenous, but we will return to the equilibrium value of m below. The domestic firm's labor supply is composed of domestic and foreign workers, and it is governed by an efficiency disadvantage of foreign (migrant) workers relative to domestic (native) workers at the same skill distance. In the figure, this disadvantage is measured by the vertical distance between the solid lines (for domestic workers) and the dashed lines (for foreign workers), respectively. Hence, for any offered wage, the domestic firm will attract more native workers than migrants. We use d^n to denote the domestic firm's skill reach for native workers on either side of its position, and d^m to denote this firm's skill reach for migrants, again symmetrically in both directions. The two skill reaches are determined by the following indifference conditions:

[33] Alternative symmetric location patterns are discussed at the end of this section.

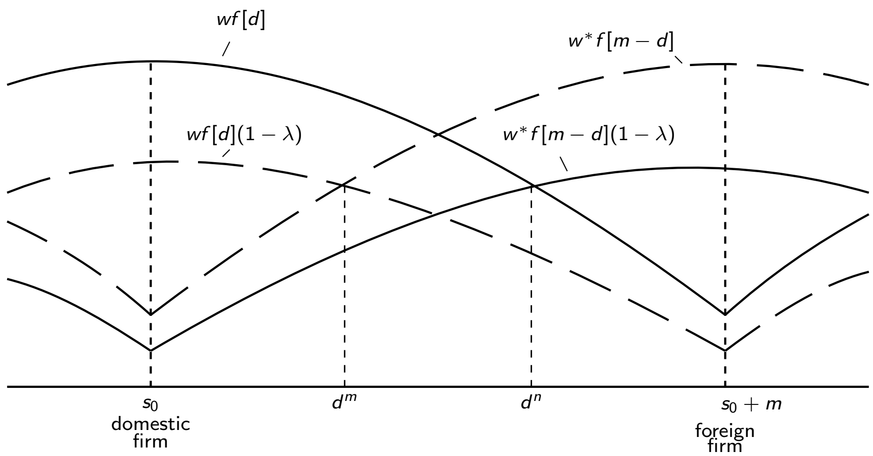


Figure 2: Worker-firm matching with migration

$$wf[d^n] = w^* f[m - d^n](1 - \lambda) \qquad \text{and} \qquad wf[d^m] = w^* f[m - d^m]\frac{1}{1-\lambda}. \tag{27}$$

For equal wages, the two skill reaches converge as migration barriers fall; with $\lambda = 0$ they coincide at $m/2$. The indifference condition implies that income of the worst-matched native is equal to the income of the worst-matched migrant. Hence, for any level of λ we may use d^n to measure the gap between the highest and lowest earnings.[34]

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

$$L^{S,M}[w, w^*, m, \lambda] = \frac{L}{H}\left(\int_0^{d^n[w,w^*,m,\lambda]} f[d]\mathrm{d}d + \int_0^{d^m[w,w^*,m,\lambda]} f[d](1-\lambda)\mathrm{d}d\right), \tag{28}$$

where a superscript M indicates the case of costly migration, as opposed to L^S for closed labor markets or $L^{S,L}$ for the intensive margin of migration. In the above equation, $d^n[w, w^*, m, \lambda]$ and $d^m[w, w^*, m, \lambda]$ are implicitly determined by the two equations in (27). Symmetry across countries with $w = w^*$ implies $d^m = d^m[m, \lambda] = m - d^n[m, \lambda]$. The migration cost becomes prohibitive if $f[m] = f[0](1 - \lambda)$.

Firm behavior on this labor market is governed by the perceived elasticity of labor

[34] We would also like to note that with positive and non-prohibitive migration cost, concavity of $f[\cdot]$ is no longer enough to establish a monotone relationship between the skill reach and the other measures of inequality analyzed in Appendix A.5 (variance of earnings and Gini coefficient). The reason is that the distribution of earnings is now given by a weighted average of the distribution among migrants and native workers. A change in λ affects both the distribution within the two groups and the relative weights of the groups. Hence, the effect on the overall distribution of earnings depends among other parameters on the change in the curvature of $f[\cdot]$ and the initial level of λ.

supply, evaluated at the symmetric equilibrium: $\eta^M[m,\lambda] = \frac{\partial L^{S,M}}{\partial w} \frac{w}{L^{S,M}}\Big|_{w=w*=1}$. We use $\psi^M[m,\lambda] := \left(\eta^M[m,\lambda]+1\right) \big/ \eta^M[m,\lambda] > 1$ to denote the wage markdown. Details on $\eta^M[m,\lambda]$ are found in Appendix A.8.2, where we also prove that a reduction in the migration cost λ increases the labor supply elasticity $\eta^M[m,\lambda]$, which implies a lower wage markdown.[35] Moreover, for a given level of m, the markdown is unambiguously lower with migration than without. Interestingly, even if migration barriers approach the prohibitive level and migration becomes negligibly small, 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 that each firm's labor supply is more elastic than with closed labor markets. As λ converges to its prohibitive level $\bar{\lambda}$ from below, the perceived wage elasticity of labor supply converges to

$$\eta^M[m,\bar{\lambda}] = -\frac{2f[m]^2}{f'[m]+(1-\bar{\lambda})f'[0]} \frac{1}{F[m]}. \tag{29}$$

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).

In analogy to the extensive margin, the intensive margin of migration also affects the average quality of skill matches between workers and firms. In a symmetric equilibrium, the average matching quality emerges as

$$\theta^M[m,\lambda] := \frac{1}{m}\left(\int_0^{d^n[m,\lambda]} f[d]\mathrm{d}d + \int_0^{d^m[m,\lambda]} f[d](1-\lambda)\mathrm{d}d\right). \tag{30}$$

As shown in Appendix A.8.5, θ^M is falling in λ, reaching $\theta^M[m,0] = \theta[m/2]$ for frictionless migration where $\lambda = 0$. For prohibitively high migration barriers, $\lambda = \bar{\lambda}$, the average matching quality as given in (30) is the same function of m as under autarky, given in (11): $\theta^M[m,\bar{\lambda}] = \theta[m]$.

As with the productivity gains at the extensive margin above, effective labor supply

[35] Strictly speaking this requires that $f'''[d]$ is not too large (in absolute terms). The reasoning behind this condition is as follows: A lower λ leads firms to increase the share of migrants employed by shifting d^m outwards and d^n inwards. If the curvature of $f[d]$ falls (in absolute terms) as the skill reaches move to the right, a decrease in λ helps firms to avoid competition by employing more migrant workers in the range where the curvature of $f[d]$ is strong and fewer natives in the range where the curvature of $f[d]$ is weak. We rule this out by assuming that the curvature does not decrease too much (in absolute terms) as the skill reaches moves to the right.

per firm, $L^{S,M} = \frac{L}{N}\theta^M[m,\lambda]$, is larger than under autarkic labor markets since for $\lambda < \bar{\lambda}$ we have $\theta^M[m,\lambda] > \theta[m]$. Accordingly, average cost are smaller than under autarky conditional on the same number of firms per country. In Appendix A.8.3 we prove that the matching gains from migration, $\theta^M[m,\lambda] - \theta[m]$, are increasing in m and thus decreasing in N. 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 matching gains from migration.[36]

Thus, the first step of our analysis reveals that, conditional on m, opening up labor markets via a piecemeal reduction in the cost of migration affects firms' environment in a way that is qualitatively similar to the effects of liberalization on the extensive margin. 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, and this pro-competitive effect is increasing with lower migration cost.

With open labor markets and migration cost $\lambda \in [0, \bar{\lambda})$, the equilibrium condition determining the number of firms and the skill distance between them reads as

$$g^M[m,\lambda] = \rho^T[m/k]\psi^M[m,\lambda]. \tag{31}$$

As before, the term $g^M[m,\lambda]$ measures average cost, taking into account the labor market clearing condition, which now reads as $\alpha + kq = (m/H)L\theta^M[m,\lambda]$. Average cost read as

$$g^M[m,\lambda] = \frac{L\theta^M[m,\lambda]}{L\theta^M[m,\lambda] - \alpha H/m}. \tag{32}$$

An equilibrium as described by (31) hinges on the condition $d^m[m,\lambda] > 0$. We show in Appendix A.8.6 that there is a unique value $\bar{\lambda}$ solving (31) 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.3, there exists a unique symmetric equilibrium described by (31) featuring a positive number of firms for $\lambda \in [0, \bar{\lambda})$. Note the formal analogy to (24) which looks at the extensive margin of migration: ℓ, the number of countries connected by costless migration, is now replaced by λ, the cost of migration. For $\lambda = 0$ (free migration) we have that (31) is identical to (26) for $\ell = 2$ countries.

[36] The same relationship between the matching gains from migration and m holds with regard to the extensive margin of migration, where we have that $\theta^L[m/\ell] - \theta[m]$ increases in m; see Appendix A.8.3.

Proposition 4. *An equilibrium with costly migration between two countries has the following properties: (i) With the migration cost so high that migration is negligibly small, both countries observe a lower number of firms, lower prices, and a higher average welfare level, coupled with a higher degree of inequality, than in an equilibrium with autarkic 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 the average welfare level in both countries.*

Proof: The analytical details of the proof are relegated to Appendix A.8.4 for part (i) and to Appendix A.8.5 for part (ii).

The intuition for part (i) is as follows. Lower monopsony power on open labor markets implies that firms would make losses, if the number of firms were the same as in the reference equilibrium with autarkic labor markets. A zero-profit equilibrium therefore requires firm exit and thus a rise in m. In addition, goods prices will fall due to lower wage markdowns. The welfare increase follows from the excess-entry property of the equilibrium. Moreover, in the equilibrium considered, the skill reach approaches $d^n[m, \bar{\lambda}] = m$. Hence, with a higher m than under autarky we also observe a higher earnings gap between the best and worst matches.

The ambiguity regarding the number of firms in part (ii) of the proposition is familiar from Proposition 3 above. A lower migration cost lowers both, the wage markdown (more competition) as well as the average cost (better matching). If the pro-competitive effect dominates, then zero profits require firm exit and thus a rise in m, and vice versa if the matching effect dominates. In contrast to Proposition 3, however, ambiguity in m now also leads to an ambiguity regarding the degree of inequality. The reason is that the skill reach is no longer equal to m/ℓ but equal to d^n. We have two effects here. A lower migration cost lowers the skill reach while a lower number of firms increases it. If the lower migration cost leads to firm entry, then both effects operate towards a lower skill reach and thus less inequality. But if it leads to firm exit, then we have two opposing effects on the skill reach. And in contrast to the extensive margin analysis, the firm exit effect may dominate the migration cost effect.

The price effect in part (ii) of the proposition is intuitive since a lower λ lowers average cost as well as wage markdowns, leading to 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 4 (ii) may be described as

$$\mathrm{d}\ln V = \frac{\partial \ln \left[\theta^M/g^M\right]}{\partial \lambda}\mathrm{d}\lambda + \left[\frac{\partial \ln \left[\theta^M/g^M\right]}{\partial m} - \frac{1}{4\gamma H}\right]\mathrm{d}m, \tag{33}$$

where we 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 matching quality 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 (33) 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 (33) dominates, leading to an unambiguously positive welfare effect from $\mathrm{d}\lambda < 0$ for any initial $\lambda \in [0, \bar{\lambda})$; see Appendix A.8.5.

Proposition 4 implies the following corollary:

Corollary 1. *A symmetric equilibrium with open labor markets delivers an average level of welfare that is larger than with autarkic labor markets, irrespective of the value of* $\lambda \in [0, \bar{\lambda})$.

This follows from a positive welfare effect of opening up labor markets at a migration cost that is below but arbitrarily close to the prohibitive level (part i), coupled with positive welfare effects for successive marginal reductions in the cost of migration over the entire interval $[0, \bar{\lambda})$ (part ii).

We close this section by briefly discussing the robustness of our findings. A first point to note is that excess entry is not a necessary condition for the welfare effects established in Proposition 4 and the corollary. 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 (33) would offset each other. We are thus left with the effects of a lower λ through a better matching quality and lower markdowns from more competitive labor markets in the first term of (33), which is unambiguously positive.

The second point relates to the conditions under which the results of Proposition 4 part (ii) generalize to more than two countries integrating their labor markets. The results hinge on two properties of the labor supply function. Namely, that, for a given size of the work force, lower migration cost make it easier to attract more workers for a given wage ($\eta_\lambda^M < 0$) and increase the average matching quality ($\theta_\lambda^M < 0$). Both properties are highly plausible features of a more general sorting scheme but not straightforward to establish in the present setting.[37] A similar argument also applies to other specifications of the migration cost.

Finally, how about alternative firm location patterns? The "perfectly symmetric" equilibrium analyzed above, with equal distances m between any two firms from different countries, maximizes labor supply per firm since it equalizes the productivity of all marginal workers at positions d^n and d^m around the circle. One can imagine alternative patterns of location 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 at the same (symmetric) points on the skill circle. In this case, the incentive for migration disappears; nobody perceives any matching advantage from migration. More generally, with positive migration cost, if domestic firms locate sufficiently close to foreign firms so that nobody sees a matching advantage from migration large enough to outweigh the migration cost, then we have a quasi-symmetric location pattern with zero migration. Such a pattern seems a possible strategy for firms to avoid fiercer competition from open labor markets. In Heiland and Kohler (2018) we model the entry game for arbitrary location patterns and discuss conditions under which the fully symmetric location equilibrium underlying Proposition 4 prevails and indeed is the unique equilibrium of the entry game.

6 Conclusion

We have readdressed the common narrative of variety-based gains from trade. Traditional models of monopolistic competition in the spirit of Krugman (1991) stress the importance

[37] To be more specific, with more than two countries the labor supply function will have kinks, so that the partial derivative (and thus the labor supply elasticity) has no unique value. Such kinks arise at wage levels where domestic workers with a certain skill type are indifferent between migrating to two different foreign countries. Increasing the wage makes the domestic firm competing with a firm from a country positioned farther away on the skill circle, whereas lowering the wage makes it competing with a firm from the country positioned less far away. The specific wage levels generating such kinks depend on the magnitude of the migration cost. Obviously, such kinks cannot arise if there are only two countries with mutually open labor markets. A complete description of the labor supply curve for four countries is available upon request.

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. We argue that this view neglects an important aspect of the labor market: If the labor force is heterogeneous in terms of skill types (as opposed to skill levels), then firms are likely to 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.

We have developed a model that combines standard features of monopolistic competition on goods markets with skill-type (or horizontal) heterogeneity of workers, represented by a circular skill space. In this modeling 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 wage markdowns. 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, scale, and pro-competitive effects on goods markets dominate the adverse effects from a lower average quality of worker-firm matches and from higher markdowns on the labor market. Therefore, the standard result of gains from trade survives and, lower matching quality notwithstanding, we observe firms with higher productivity. However, looking at piecemeal trade liberalization between two symmetric countries, we find a positive aggregate welfare effect only if the initial level of trade cost is below a certain threshold level. And we generally find that trade liberalization increases the degree of within-firm income inequality among workers.

We also find that in this modeling environment labor market integration generates an incentive for workers to migrate for employment in specific 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, however, 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

Akerman, A., Helpman, E., Itskhoki, O., Muendler, M.-A., and Redding, S. (2013). Sources of wage inequality. American Economic Review, 103(3):214–19.

Amiti, M. and Davis, D. R. (2011). Trade, Firms, and Wages: Theory and Evidence. The Review of Economic Studies, 79(1):1–36.

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

Artuç, E., Chaudhuri, S., and McLaren, J. (2010). Trade Shocks and Labor Adjustment: A Structural Empirical Approach. American Economic Review, 100(3):1008–45.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms*. The Quarterly Journal of Economics, 135(2):645–709.

Autor, D. H., Dorn, D., Hanson, G. H., and Song, J. (2014). Trade Adjustment: Worker-Level Evidence. The Quarterly Journal of Economics, 129(4):1799–1860.

Baldwin, R. (2005). Heterogeneous Firms and Trade: Testable and Untestable Properties of the Melitz Model. Working Paper 11471, National Bureau of Economic Research.

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

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

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

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

Bilbiie, F. O., Ghironi, F., and Melitz, M. J. (2019). Monopoly power and endogenous product variety: Distortions and remedies. American Economic Journal: Macroeconomics, 11(4):140–74.

Bykadorov, I., Ellero, A., Funari, S., Kokovin, S., and Molchanov, P. (2016). Painful Birth of Trade Under Classical Monopolistic Competition. HSE Working Paper, (WP BRP 132/EC/2016).

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.

Caliendo, L., Feenstra, R. C., Romalis, J., and Taylor, A. M. (2015). Tariff Reductions, Entry, and Welfare: Theory and Evidence for the Last Two Decades. Working Paper 21768, NBER.

Caliendo, L., Feenstra, R. C., Romalis, J., and Taylor, A. M. (2021). A Second-best Argument for Low Optimal Tariffs. Working Paper 28380, NBER.

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.

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

Davis, D. R. and Harrigan, J. (2011). Good Jobs, Bad Jobs, and Trade Liberalization. Journal of International Economics, 84(1):26–36.

Dhingra, S. and Morrow, J. (2019). Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity. Journal of Political Economy, 127(1):196–232.

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

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

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

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

Egger, H. and Kreickemeier, U. (2009). Firm Heterogeneity and the Labor Market Effects of Trade Liberalization. International Economic Review, 50(1):187–216.

Egger, H. and Kreickemeier, U. (2012). Fairness, Trade, and Inequality. Journal of International Economics, 86(2):184 – 196.

Egger, H., Kreickemeier, U., Moser, C., and Wrona, J. (2021). Exporting and offshoring with monopsonistic competition. The Economic Journal, https://doi.org/10.1093/ej/ueab078.

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.

Grossman, G. M. and Maggi, G. (2000). Diversity and trade. American Economic Review, 90(5):1255–1275.

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

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

Hao, L. (2017). Assortative Matching, pages 1–5. Palgrave Macmillan UK, London.

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.

Helpman, E. (2017). Globalisation and Wage Inequality. Journal of the British Academy, 5:125–162.

Helpman, E., Itskhoki, O., Muendler, M.-A., and Redding, S. J. (2017). Trade and Inequality: From Theory to Estimation. The Review of Economic Studies, 84(1):357–405.

Helpman, E., Itskhoki, O., and Redding, S. (2010). Inequality and Unemployment in a Global Economy. Econometrica, 78(4):1239–1283.

Hummels, D. and Lugovsky, V. (2009). International Pricing in a Generalized Model of Ideal Variety. Journal of Money, Credit and Banking, 41(s1):3–33.

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.

Jha, P. and Rodriguez-Lopez, A. (2021). Monopsonistic Labor Markets and International Trade. European Economic Review, forthcoming.

Jung, B. and Kohler, W. (2021). Input-Output Linkages and Monopolistic Competition: Input Distortion and Optimal Policies. Hohenheim Discussion Papers in Business, Economics and Social Sciences 06-2021, University of Hohenheim.

Kohl, M. (2020). Redistribution, Selection, and Trade. Journal of International Economics, 122:103255.

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. (1991). Increasing Returns and Economic Geography. Journal of Political Economy, 99(3):483–99.

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. and Redding, S. J. (2015). New Trade Models, New Welfare Implications. American Economic Review, 105(3):1105–46.

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.

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.

Sampson, T. (2014). Selection into Trade and Wage Inequality. American Economic Journal: Microeconomics, 6(3):157–202.

Sattinger, M. (1993). Assignment Models of the Distribution of Earnings. Journal of Economic Literature, 31(2):831–880.

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

Song, J., Price, D. J., Guvenen, F., Bloom, N., and von Wachter, T. (2018). Firming Up Inequality. The Quarterly Journal of Economics, 134(1):1–50.

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

Teulings, C. N. (1995). The Wage Distribution in a Model of the Assignment of Skills to Jobs. Journal of Political Economy, 103(2):280–315.

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

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

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	.36
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 Existence and uniqueness of the symmetric two-stage equilibrium

As described in Section 3.2, for symmetric distance patterns a zero-profit equilibrium 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 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 $\psi[m] := \frac{\eta[m]+1}{\eta[m]}$ denote the markup on the goods market and the markdown on 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 the price markup and the wage markdown 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.[38]

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) 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.4 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[39]

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

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

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

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

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

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

A.5 Alternative measures of inequality

Proof that the variance of incomes 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 individuals, 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,

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

Proof that the GINI coefficient increases in m. The GINI coefficient is defined as

$$GINI = 1 - 2B, \tag{A.11}$$

where $B = \int_0^1 L[y]\mathrm{d}y$ is the area below the Lorenz curve. The Lorenz curve $L[y]$ gives the share of income earned by the lowest y percent of the income distribution. In our setting, the Lorenz curve is

$$L[x,m] = \frac{F[m] - F\left[(1-x)m\right]}{F[m]} \qquad \text{for } 0 \leq x \leq 1 \qquad \text{and with } F[z] = \int_0^z f[d]\mathrm{d}d. \tag{A.12}$$

Hence, we have

$$B = \int_0^1 L[x,m]dx = \int_0^1 \left(1 - \frac{F\left[(1-x)m\right]}{F[m]}\right)\mathrm{d}x \quad \text{with} \quad \frac{\partial B}{\partial m} = \int_0^1 \frac{\partial L[x,m]}{\partial m}\mathrm{d}x. \tag{A.13}$$

It holds that $\frac{\partial B}{\partial m} < 0$ if

$$\frac{\partial L[x,m]}{\partial m} = -\left(\frac{f\left[(1-x)m\right](1-x)}{F[m]} - \frac{f[m]F\left[(1-x)m\right]}{F[m]^2}\right) < 0 \qquad \forall\, x. \tag{A.14}$$

Condition (A.14) can be written as $\frac{f[(1-x)m](1-x)m}{F[(1-x)m]} > \frac{f[m]m}{F[m]}$ and, defining $a[y] := \frac{f[y]y}{F[y]}$, it holds if

$$\frac{\partial a[y]}{\partial y} = \frac{yf'[y]}{F[y]} + \frac{f[y]}{F[y]} - \frac{yf[y]^2}{F[y]^2} < 0. \tag{A.15}$$

Note that a measures the ratio of the (rectangular) area spanned by $f[y]$, y and the area below the curve $f[y]$ up to y. Using $F[y] = \theta[y]y$, the condition can be written as

$$\frac{f'[y]}{\theta[y]} + \frac{f[y]}{y\theta[y]} - \frac{f[y]^2}{y\theta[y]^2} < 0 \qquad \Leftrightarrow \qquad \theta[y] < f[y] - yf'[y]\frac{\theta[y]}{f[y]} \tag{A.16}$$

$\theta[y] < 1$, $\frac{\theta[y]}{f[y]} > 1$ and concavity ($f[y] - yf'[y] > 1$) imply that condition (A.16) is true. Hence, $\frac{\partial a[y]}{\partial y} < 0$, which implies $\frac{\partial B}{\partial m} < 0$. (A.11) then implies that the GINI coefficient increases in m (implying higher inequality), provided $f[\cdot]$ is concave.

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

$$\frac{Lf[m]}{L\theta[m] - \frac{\alpha H}{m}} = 1 + \frac{m}{2\gamma H}. \tag{A.17}$$

The second-order condition requires

$$\frac{\mathrm{d}^2 \ln V}{\mathrm{d}m^2} = -\frac{\left(L\theta'[m] + \frac{\alpha H}{m^2}\right)^2}{\left(L\theta[m] - \frac{\alpha H}{m}\right)^2} + \frac{L\theta''[m] - 2\frac{\alpha H}{m^3}}{L\theta[m] - \frac{\alpha H}{m}} < 0. \tag{A.18}$$

A sufficient condition for this to hold is

$$\theta''[m] := \frac{\partial^2 \theta[m]}{\partial m^2} = \frac{1}{m}\left(f'[m] - \frac{2}{m} f[m] + \frac{2}{m}\theta[m]\right) \leq 0 \tag{A.19}$$

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.17) as

$$g[m] = \frac{\theta[m]}{f[m]} \frac{1}{\psi[m]} \psi[m] \rho[m/2]. \tag{A.20}$$

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 $\rho'[m] > 0$,

$$\frac{\theta[m]}{f[m]} \frac{1}{\psi[m]} < 1 \tag{A.21}$$

is a sufficient condition for this to hold. Rearranging (A.21) 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.21) 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

We prove this proposition by total differentiation of (17), demonstrating that the directions of all changes involved are independent of k. This ensures monotonicity as we move from $k = 1$ to any positive number of countries with integrated labor markets.

A.7.1 Proof of Proposition 1

(i) Log-differentiating the equilibrium condition (17), we obtain

$$\hat{m} = A\hat{k} \qquad \text{with} \qquad A := \frac{\psi[m]\frac{1}{\gamma H k}}{-g'[m] + \psi[m]\frac{1}{\gamma H k} + \rho^T[m/k]\psi'[m]}. \tag{A.22}$$

Since $g'[m] < 0$ and $\psi'[m] > 0$ (see Appendix A.3 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 kN increases nevertheless. Since (A.22) holds for any initial level of k, it follows that the number of firms (number of varieties) in an equilibrium with free trade among k countries is smaller (larger) 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 markdown increases. Log-differentiating (16) yields

$$\hat{p} = B\hat{k} \qquad \text{with} \qquad B = \frac{\frac{m}{\gamma H k}}{\left(1 + \frac{m}{\gamma H k}\right)} \frac{g'[m]}{\left(-g'[m] + \psi[m]\frac{1}{\gamma H k} + \rho^T[m/k]\psi'[m]\right)}. \tag{A.23}$$

Since $-1 < B < 0$, it follows that $\hat{p} < 0$. Since (A.23) holds for any initial level of k, prices are lower with free trade than under autarky.

(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

$$\max_{w_i, q_i} \left\{ r_i[q_i, N, \overline{\ln p}, Y] + (k-1) r_i^* \left[\frac{\bar{q}_i - q_i}{(k-1)\tau}, N, \overline{\ln p}, Y \right] - w_i L_i \right\}.$$

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

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

and the first-order condition with respect to q_i reads

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

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 markets is the same and given by $\overline{\ln p} = \overline{\ln p}^* = 1/k \ln p + (k-1)/k \ln p^*$. Inserting $\overline{\ln p}$ and $\overline{\ln p}^*$ into (19), we can use the same logic as in A.2 to obtain explicit solutions for p and p^* using (20), where the price markups no longer depend on the own price, but only on the respective other price and the number of firms:

$$p = \mathcal{W}[\tilde{Z}] \frac{\psi(k-1)}{k} \qquad \text{with } \tilde{Z} = \frac{k}{(k-1)\psi} \exp\left\{ \frac{k}{k-1} + \frac{1}{k-1} \frac{m}{\gamma H} + \ln p^* \right\} \tag{A.24}$$

$$p^* = \mathcal{W}[\tilde{Z}^*] \frac{\psi\tau}{k} \quad \text{with } \tilde{Z}^* = \frac{k}{\psi\tau} \exp\left\{ k + \frac{m}{\gamma H} + \ln p \right\}. \tag{A.25}$$

Inserting $p = \mathcal{W}[\tilde{Z}] \frac{\psi(k-1)}{k}$ and $p^* = \mathcal{W}[\tilde{Z}^*] \frac{\psi\tau}{k}$ into the $\tilde{Z}$-terms, we obtain

$$p = \mathcal{W}\left[\frac{\mathcal{W}[\tilde{Z}^*]\tau}{k-1} \exp\left\{ \frac{k}{k-1} + \frac{m}{(k-1)\gamma H} \right\} \right] \frac{\psi(k-1)}{k} \tag{A.26}$$

$$p^* = \mathcal{W}\left[\frac{\mathcal{W}[\tilde{Z}](k-1)}{\tau} \exp\left\{ k + \frac{m}{\gamma H} \right\} \right] \frac{\psi\tau}{k} \tag{A.27}$$

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 emerges as

$$W := W[W^*, m] = \mathcal{W}\left[\frac{W^*\tau}{k-1}\exp\left\{\frac{k}{k-1} + \frac{m}{(k-1)\gamma H}\right\}\right] \tag{A.28}$$

$$W^* := W^*[W, m] = \mathcal{W}\left[\frac{W(k-1)}{\tau}\exp\left\{k + \frac{m}{\gamma H}\right\}\right]. \tag{A.29}$$

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 $p = \left(1 + \frac{1}{\gamma k N} + \frac{k-1}{k}(\ln p^* - \ln p)\right)\tilde{w}$ and analogously for p^*. In view of (A.24) and (A.25) it follows that $W\frac{k-1}{k} = 1 + \frac{1}{\gamma k N} + \frac{k-1}{k}(\ln p^* - \ln p)$ and $W^*\frac{1}{k} = 1 + \frac{1}{\gamma k N} + \frac{1}{k}(\ln p - \ln p^*)$. The expenditure shares in (19) can therefore be written as

$$\delta = \left(W\frac{k-1}{k} - 1\right)\gamma \qquad \text{and} \qquad \delta^* = \left(\frac{W^*}{k} - 1\right)\gamma. \tag{A.30}$$

Direct demand functions for foreign varieties in terms of W^* obtain as $q^* = \frac{\delta^* Y}{p^*} = \left(1 - \frac{k}{W^*}\right)\frac{\gamma Y}{\psi\tau}$. This implies that the prohibitive level of trade costs $\bar{\tau}$ for which $q^* = 0$ satisfies $\mathcal{W}\left[\frac{W(k-1)}{\bar{\tau}}\exp\left\{k + \frac{m}{\gamma H}\right\}\right] \equiv k$. It follows that for non-prohibitive trade costs $W \geq W^* > k$. Inserting demand and income $Y = L\theta$ into the labor market clearing condition (21), and rearranging terms gives

$$\begin{aligned}
\gamma\left(k - \frac{k}{(k-1)W} - \frac{k(k-1)}{W^*}\right) &= \frac{\frac{L\theta[m]}{N[m]} - \alpha}{L\theta[m]}\psi[m] \\
\gamma h[W, W^*] &= \frac{\psi[m]}{g[m]N[m]}.
\end{aligned} \tag{A.31}$$

For easier reference the second line introduces $h[W, W^*] := \left(k - \frac{k}{(k-1)W} - \frac{k(k-1)}{W^*}\right)$. (A.31), (A.28) and (A.29) 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.31), (A.28), (A.29) leads to

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

$$\begin{bmatrix} \frac{-1}{(k-1)W-1-\frac{(k-1)^2W}{W^*}} & \frac{-1}{\frac{W^*}{k-1}-1-\frac{W^*}{(k-1)^2W}} & \frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N}-\alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \\ -1 & \frac{1}{W+1} & \frac{1}{(k-1)\gamma N}\frac{1}{W+1} \\ \frac{1}{W^*+1} & -1 & \frac{1}{\gamma N}\frac{1}{W^*+1} \end{bmatrix} \begin{bmatrix} \widehat{W} \\ \widehat{W}^* \\ \widehat{m} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{W+1}\hat{\tau} \\ \frac{1}{W^*+1}\hat{\tau} \end{bmatrix}. \tag{A.32}$$

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

$$\omega = \frac{1}{(W+1)(W^*+1)} \left[d_{13}W^* - \frac{1}{\gamma N h[W,W^*]} \frac{k^2}{W^*} \right] \frac{1}{\det[D]} \tag{A.33}$$

$$\omega^* = \frac{1}{(W+1)(W^*+1)} \left[-d_{13}W + \frac{1}{\gamma N h[W,W^*]} \frac{k^2}{(k-1)^2 W} \right] \frac{1}{\det[D]} \tag{A.34}$$

$$\mu = \frac{\frac{k}{k-1}\frac{W^*}{W} - k(k-1)\frac{W}{W^*}}{h[W,W^*](W+1)(W^*+1)} \frac{1}{\det[D]} \tag{A.35}$$

where d_{13} is the element in row 1 and column 3 of D. The signs of the elasticities hinge upon the sign of the determinant which is given by

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

Since $W > \frac{k}{k-1}$ and $W^* \geq k$, we have $WW^* + W + W^* > (k+W^*)\frac{k}{(k-1)^2W} + (k+kW - W)\frac{k}{W^*}$. This implies that $\det[D] > 0$ if

$$\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N}-\alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} > \frac{1}{\gamma N h[W,W^*]}. \tag{A.37}$$

We know from above that $\frac{f[m]}{\theta[m]} < 1$ and $\frac{\partial \ln \psi}{\partial \ln m} > 0$. Therefore, inequality (A.37) holds if

$$\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N}-\alpha} > \frac{1}{\gamma N h[W,W^*]}. \tag{A.38}$$

Using the equilibrium condition (A.31), 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 elasticities ω, ω^*, we note that $W^* \geq \frac{k^2}{W^*}$, $W \geq \frac{k^2}{(k-1)^2W}$, $\det[D] > 0$ and (A.37) jointly imply $\omega > 0$ and $\omega^* < 0$. And finally, $W \geq W^*$ and $k \geq 2$ 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.35), 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 $\hat{m} = 0$. 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 the price markup and the wage markdown

$$\hat{p}^* = \left(\omega^* + \frac{\partial \ln \psi}{\partial \ln m}\mu + 1\right)\hat{\tau}, \tag{A.39}$$

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.34) and (A.35) shows that $\hat{p}^*$ is negative for $\hat{\tau} < 0$ if and only if

$$-\frac{d_{13}W - \frac{1}{\gamma N h}\frac{k^2}{(k-1)^2 W} + \frac{\partial \ln \psi}{\partial \ln m}\frac{1}{h}\left(\frac{k}{k-1}\frac{W^*}{W} - \frac{k(k-1)W}{W^*}\right)}{d_{13}(WW^* + W + W^*) - \frac{1}{\gamma h N}\left((k+W^*)\frac{k}{(k-1)^2 W} + (k + kW - W)\frac{k}{W^*}\right)} + 1 > 0. \tag{A.40}$$

Canceling identical terms in the denominator and the numerator shows that this holds if

$$\frac{\frac{\partial \ln \psi}{\partial \ln m}\frac{1}{h}\left(\frac{k}{k-1}\frac{W^*}{W} - \frac{k(k-1)W}{W^*}\right)}{d_{13}(WW^* + W^*) - \frac{1}{\gamma h N}\left(W^*\frac{k}{(k-1)^2 W} + (k + kW - W)\frac{k}{W^*}\right)} < 1.$$

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.38), it follows that $WW^* + W^* \geq W^*\frac{k}{(k-1)^2 W} + (k + kW - W)\frac{k}{W^*}$ and $WW^* + W^* \geq \frac{1}{h}\left(\frac{k}{k-1}\frac{W^*}{W} - \frac{k(k-1)W}{W^*}\right)$ is sufficient for the inequality in (A.40) to hold. Remembering from above that $W \geq W^* \geq k \geq 2$, it is straightforward to show that these two conditions are fulfilled.

The change in the domestic price obtains as

$$\hat{p} = \left(\omega + \frac{\partial \ln \psi}{\partial \ln m}\mu\right)\hat{\tau}. \tag{A.41}$$

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 markdown. 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.33) and (A.35) gives

$$\hat{p} = \left[W^* \left(\frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - \frac{f[m]}{\theta[m]} + 1 + \frac{\partial \ln \psi}{\partial \ln m} \right) - \frac{1}{\gamma N h[W, W^*]} \frac{k^2}{W^*} \right.$$
$$\left. + \frac{\partial \ln \psi}{\partial \ln m} \frac{1}{h[W, W^*]} \left(\frac{k}{k-1} \frac{W^*}{W} - k(k-1) \frac{W}{W^*} \right) \right] \frac{1}{(W+1)(W^*+1)} \frac{\hat{\tau}}{\det[D]}. \quad \text{(A.42)}$$

Remember that prohibitive trade costs imply an infinite price elasticity and therefore a price markup of zero, whence $W^* = k$. To see whether $\omega + \frac{\partial \ln \psi}{\partial \ln m} \mu < 0$ for $\tau = \bar{\tau}$, as stated in Proposition 2, we must therefore evaluate the bracketed term at $W^* = k$. We obtain[40]

$$k \frac{\frac{Lf[m]}{N}}{\frac{L\theta[m]}{N} - \alpha} - k \frac{f[m]}{\theta[m]} + k + k \frac{\partial \ln \psi}{\partial \ln m} - (k-1)W - \frac{\partial \ln \psi}{\partial \ln m} \left(k + (k-1)W \right). \quad \text{(A.45)}$$

Inserting the equilibrium condition (A.31), which reduces to $\gamma h[W, W^*] = \frac{L\theta[m]/N - \alpha}{L\theta[m]/N} \psi = \frac{k}{(k-1)W}$ at $\tau = \bar{\tau}$, shows that the expression is negative, if

$$\psi W \frac{f[m]}{\theta[m]} (k-1) < k \frac{f[m]}{\theta[m]} + (k-1)W - k + \frac{\partial \ln \psi}{\partial \ln m} (k-1)W. \quad \text{(A.46)}$$

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

$$\frac{(k-1)W}{\theta[m]} \frac{f[m]^2 - 2f'[m]F[m]}{f[m]} < (k-1)W - k + k \frac{f[m]}{\theta[m]}$$
$$+ W(k-1) \left(\frac{-2f''[m]\theta}{f[m]^2 \psi} - \frac{2mf'[m]}{f[m]} \right). \quad \text{(A.47)}$$

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

$$\frac{(k-1)W}{\theta[m]} \frac{f[m]^2 - 2f'[m]F[m]}{f[m]} < (k-1)W - k + k \frac{f[m]}{\theta[m]} + W(k-1) \left(-\frac{2mf'[m]}{f[m]} \right). \quad \text{(A.48)}$$

Rearranging terms and using $m\theta = F[m]$ shows that this inequality holds since $W(k-$

[40] The derivation uses the fact that, at $W^* = k$,

$$\frac{1}{h[W, W^*]} \left(\frac{k}{k-1} \frac{W^*}{W} - k(k-1) \frac{W}{W^*} \right) = -\frac{k^2 - (k-1)^2 W^2}{k - (k-1)W} = -(k + (k-1)W) \quad \text{(A.43)}$$

and

$$\frac{1}{\gamma N h[W, W^*]} \frac{k^2}{W^*} = (k-1)W \quad \text{(A.44)}$$

since $h[W, k] = 1 - \frac{k}{(k-1)W}$ and $N\gamma h[W, k] = \frac{N\delta k}{(k-1)W} = \frac{k}{(k-1)W}$. The last step follows from $\delta = \frac{1}{N}$.

$1)/k > 1$ and $f[m] < \theta$. 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

$$\ln P^T[p, p^*, m] = \frac{1}{2\gamma kN} + \frac{1}{kN}\sum_{i=1}^{kN} \ln p_i + \frac{\gamma}{2kN}\sum_{i=1}^{kN}\sum_{j=1}^{kN} \ln p_i (\ln p_j - \ln p_i), \tag{A.49}$$

with i, j indexing domestic *and* foreign varieties. Under symmetry, the price index simplifies to

$$\ln P^T[p, p^*, m] = \frac{1}{2\gamma kN} + \frac{1}{k}\ln p + \frac{k-1}{k}\ln p^* - \frac{\gamma(k-1)N}{2k}\left(\ln p - \ln p^*\right)^2. \tag{A.50}$$

The change in indirect utility is then

$$\widehat{V} = \left(\frac{\partial \ln \theta}{\partial \ln m} - \frac{\partial \ln P}{\partial \ln m}\right)\widehat{m} - \frac{\partial \ln P}{\partial \ln p}\hat{p} - \frac{\partial \ln P}{\partial \ln p^*}\hat{p}^*, \tag{A.51}$$

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

A.8.1 Analytical details of the proof of Proposition 3

We prove this proposition by total differentiation of (26), demonstrating that the *directions* of all changes involved are independent on k, as well as independent on ℓ. The latter ensures monotonicity as we move from $\ell = 1$ to any positive number of countries with integrated labor markets. Thus, differentiating (26) and using subscripts to indicate partial derivatives, we obtain

$$\mathrm{d}m = \frac{-\rho^T\psi_\ell^L + g_\ell^L}{\rho^T\psi_m^L - g_m^L + \psi^L\rho_m^T}\mathrm{d}\ell, \qquad \text{with} \tag{A.52}$$

$$g_\ell^L = -\frac{\alpha H/mL\theta_\ell}{(L\theta[m/\ell] - \alpha H/m)^2} < 0 \quad \text{and} \quad g_m^L = \frac{-\alpha H/mL\theta_m - L\theta\alpha N/m}{(L\theta[m/\ell] - \alpha H/m)^2} < 0, \tag{A.53}$$

$\psi_\ell^L < 0$, $\psi_m^L > 0$, and $\rho_m^T > 0$. The denominator of the fraction on the right of (A.52) is unambiguously positive while the numerator is ambiguous, which proves part (iv) of the proposition. If the pro-competitive effect of labor market integration dominates the average cost effect from better matches, so that $-\rho^T \psi_\ell^L + g_\ell^L > 0$, then zero profits imply firm exit and thus d$m > 0$ – and vice versa. To proceed further we write

$$\hat{m} = A\hat{\ell}, \quad \text{where } A := \frac{-\rho^T \psi_\ell^L + g_\ell^L}{(\rho^T \psi_m^L - g_m^L + \psi^L \rho_m^T)\, m/\ell}. \tag{A.54}$$

The core of the proof is to demonstrate that $A < 1$, which implies that $\hat{m} - \hat{\ell} = (A-1)\hat{\ell} < 0$ if $\hat{\ell} > 0$. If this is true, then labor market integration leads to a reduction in the skill reach per firm m/ℓ, even if the skill distance m between domestic firms increases. We know from above that the wage markdown ψ and the average worker productivity θ both depend only on m/ℓ, whence $\psi_\ell = -\psi_m m/\ell$ and $\theta_\ell = -\theta_m m/\ell$. This latter equality implies

$$g_m^L = -g_\ell^L \ell/m - \frac{L\theta\alpha N/m}{(L\theta[m/\ell] - \alpha H/m)^2}. \tag{A.55}$$

Substituting in A, we obtain

$$A = \frac{-\rho^T \psi_\ell^L + g_\ell^L}{-\rho^T \psi_\ell^L + g_\ell^L + \left(\frac{L\theta\alpha N/m}{(L\theta[m/\ell] - \alpha H/m)^2} + \psi^L \rho_m^T\right)(m/\ell)}. \tag{A.56}$$

We have

$$A < 1 \quad \Longleftrightarrow \quad 0 < \left(\frac{L\theta\alpha N/m}{(L\theta[m/\ell] - \alpha H/m)^2} + \psi^L \rho_m^T\right)(m/\ell). \tag{A.57}$$

Part (i) of the proposition requires that θ increases with an increase in ℓ. Given $A < 1$, this follows from

$$\hat{\theta} = \frac{\partial \ln \theta}{\partial \ln m}\hat{m} + \frac{\partial \ln \theta}{\partial \ln \ell}\hat{\ell} = \frac{\partial \ln \theta}{\partial \ln m}(A-1)\,\hat{\ell} > 0 \tag{A.58}$$

since $\frac{\partial \ln \theta}{\partial \ln \ell} = -\frac{\partial \ln \theta}{\partial \ln m}$.

Part (ii) of the proposition holds that $\hat{p} < 0$ with $\hat{\ell} > 0$. From zero profits, $p = g^L$, we have $\hat{p} = \hat{g}^L = g_m^L(m/g)\hat{m} + g_\ell^L(\ell/g)\hat{\ell}$. Using Equation (A.55) we obtain

$$\begin{aligned} \hat{p} &= g_m^L(m/g)\hat{m} - g_m^L(m/g)\hat{\ell} - \frac{L\theta}{(L\theta[m/\ell] - \alpha H/m)^2}\frac{\alpha N}{g}\hat{\ell} \\ &= g_m^L(m/g)(A-1)\hat{\ell} - \frac{\alpha N}{L\theta[m/\ell] - \alpha H/m)}\hat{\ell} < 0. \end{aligned} \tag{A.59}$$

Finally, turning to part (iii), we first note that with $\hat{\ell} > 0$ we have $\hat{\theta} - \hat{p} > 0$, a rise in real income. In view of (15) we have

$$\hat{V} = \hat{\theta} - \hat{p} - \frac{m}{2k\gamma H}\hat{m}. \tag{A.60}$$

Since $\hat{m} = A\hat{\ell}$, welfare clearly increases if $A < 0$. In this case we have firm entry, and a positive variety effect adds to the real income increase, leading to higher welfare. If $0 < A < 1$ we have firm exit and thus a loss in variety which countervails the real income increase. However, due to the excess entry property of our equilibrium, we know that the marginal utility due to an increase in m (from lower average cost) exceeds the marginal loss from less variety.

The increase in the lowest earning is

$$\mathrm{d}\ln f[m/\ell] = \frac{f'[m/\ell]}{f[m/\ell]}\frac{m}{\ell}(1-A) > 0, \tag{A.61}$$

which implies lower inequality.

A.8.2 The elasticity of labor supply

The elasticity of labor supply in the symmetric equilibrium is defined as $\frac{\partial L^{S,M}}{\partial w_i}\frac{w_i}{L^{S,M}}$. From (28) and (27), we obtain

$$\frac{\partial L^{S,M}}{\partial w_i} = \frac{L}{H}\frac{\partial d_i^n}{\partial w_i}f[d_i^n] + (1-\lambda)\frac{L}{H}\frac{\partial d_i^m}{\partial w_i}f[d_i^m] \qquad \text{with} \tag{A.62}$$

$$\frac{\partial d_i^n}{\partial w_i} = \frac{f[d_i^n]}{-w_i f'[d_i^n] - w^*(1-\lambda)f'[m-d_i^n]} \tag{A.63}$$

$$\frac{\partial d_i^m}{\partial w_i} = \frac{(1-\lambda)f[d_i^m]}{-w_i(1-\lambda)f'[d_i^m] - w^* f'[m-d_i^m]}. \tag{A.64}$$

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

$$\begin{aligned}\eta^M = \frac{\partial L^{S,M}}{\partial w_i}\frac{w_i}{L^S}\bigg|_{w_i=w} =& \frac{L}{H}\left(\frac{f[d^n]^2}{-f'[d^n]-(1-\lambda)f'[m-d^n]} + \frac{(1-\lambda)^2 f[d^m]^2}{-(1-\lambda)f'[d^m]-f'[m-d^m]}\right)\\ &\times \frac{1}{\frac{L}{H}\left(\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\int_0^{d^m} f[d]\mathrm{d}d\right)}\\ =& \frac{2f[d^n]^2}{f'[d^n]+(1-\lambda)f'[d^m]}\frac{-1}{\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\int_0^{d^m} f[d]\mathrm{d}d}.\end{aligned} \tag{A.65}$$

The elasticity of labor supply decreases in m:

$$\eta_m^M = \eta^M \left[\frac{2f'[d^n]}{f[d^n]} \frac{\partial d^n}{\partial m} - \frac{-f''[d^n]\frac{\partial d^n}{\partial m} - (1-\lambda)f''[d^m]\frac{\partial d^m}{\partial m}}{-f'[d^n] - (1-\lambda)f'[d^m]} - \frac{f[d^n]}{m\theta^M} \right] < 0, \tag{A.66}$$

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:

$$\eta_\lambda^M = \eta^M \left[\underbrace{\frac{2f'[d^n]}{f[d^n]} \frac{\partial d^n}{\partial \lambda}}_{<0} + \underbrace{\frac{f''[d^n]\frac{\partial d^n}{\partial \lambda} + (1-\lambda)f''[d^m]\frac{\partial d^m}{\partial \lambda} + f'[d^m]}{-f'[d^n] - (1-\lambda)f'[d^m]}}_{<0} + \underbrace{\frac{F[d^m]}{F[d^n] + (1-\lambda)F[d^m]}}_{>0} \right] < 0 \tag{A.67}$$

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

$$\frac{2f'[d^n]}{f[d^n]} \frac{\partial d^n}{\partial \lambda} = 2 \underbrace{\frac{f[d^m]}{f[d^n]} \frac{f'[d^n]}{f'[d^n] + (1-\lambda)f'[d^m]}}_{\geq 1} \geq \underbrace{\frac{F[d^m]}{F[d^n] + (1-\lambda)F[d^m]}}_{<1}. \tag{A.68}$$

A.8.3 The productivity gains from migration

Intensive margin of 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

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

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 $d \in [0, d^m]$.

Extensive margin of migration. A similar same result holds for the extensive margin of migration, where we have $\frac{\partial(\theta^L[m,\ell]-\theta[m])}{\partial m} = \frac{1}{m}\left(f\left[\frac{m}{\ell}\right] - f(m)\right) + \frac{1}{m}\left(\theta[m] - \theta\left[\frac{m}{\ell}\right]\right) > 0$ for $\ell > 1$. The sign follows from the fact that

$$f\left[\frac{m}{\ell}\right] - \frac{\ell}{m}F\left[\frac{m}{\ell}\right] > f[m] - \frac{1}{m}F[m] \tag{A.69}$$

which can be shown by dividing the left-hand side by $f\left[\frac{m}{\ell}\right]$ and the right-hand-side by $f[m]$ (noting that $f\left[\frac{m}{\ell}\right] > f[m]$, which yields

$$1-\frac{F\left[\frac{m}{\ell}\right]}{\frac{m}{\ell}f\left[\frac{m}{\ell}\right]} > 1-\frac{F[m]}{mf[m]} \tag{A.70}$$

In Appendix A.5 we have shown that $\frac{yf[y]}{F[y]}$ decreases in y for a concave function $f[\cdot]$, which implies that condition (A.70) is true.

A.8.4 Analytical details of the proof of Proposition 4, part (i)

As λ approaches $\bar{\lambda}$ and migration ceases, average cost converge to the level obtaining under closed labor markets, since $g^M[m,\bar{\lambda}] = g[m]$ (see (13) and (32)). In contrast, wage markdowns in the open labor market converge to a lower value, since $\psi^M[m,\bar{\lambda}] < \psi[m]$ (implied by (9) and (29)). Hence, the zero-profit equilibrium (31) evaluated at values of λ below but arbitrarily close to $\bar{\lambda}$ 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.[41] The first-order condition of the planner reads as

$$\frac{Lf\left[d^n\right]}{L\theta^M-\frac{\alpha H}{m}} = 1+\frac{m}{2k\gamma H}, \tag{A.71}$$

where d^n, θ^M are shorthands for $d^n[m,\lambda], \theta^M[m,\lambda]$, respectively. A comparison with the market solution (31) shows that, as before, the number of firms in the market equilibrium is too large, if the markdown distortion is larger than the matching distortion. This is the case in the migration equilibrium with $\lambda < \bar{\lambda}$. 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 yields $f[d^n] > \theta^M + \frac{m}{2}\frac{f'[d^n]}{f[d^n]}\theta^M$. We show below that $f\left[\frac{d^n}{2}\right] \geq \theta^M$. Then, this inequality holds if

$$f[d^n] > f\left[\frac{d^n}{2}\right]+\frac{m}{2}\frac{f'[d^n]}{f[d^n]}\theta^M. \tag{A.72}$$

[41] 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.

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.72) holds a fortiori. Hence, the markdown distortion exceeds the matching distortion and consequently, the number of firms in the market equilibrium with migration is too large.[42]

Proof that $\theta^M \leq f\left[\frac{d^n}{2}\right]$. Using the expression for θ^M in (30) 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

$$\theta^M = \frac{1}{m}\int_0^{d^n} f[d]\mathrm{d}d + (1-\lambda)\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d \leq \frac{d^n}{m} f\left[\frac{d^n}{2}\right] + (1-\lambda)\frac{d^m}{m} f\left[\frac{d^m}{2}\right]. \quad \text{(A.73)}$$

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 (27) it follows that a symmetric equilibrium is characterized by $(1-\lambda) = f[d^n]/f[d^m]$, so the condition becomes $f\left[\frac{d^m}{2}\right]/f\left[\frac{d^n}{2}\right] \leq \frac{f[d^m]}{f[d^n]}$, which is implied by $d^m \leq d^n$ and $f''[\cdot] \leq 0$. This completes the proof.

A.8.5 Proof of Proposition 4, part (ii)

Totally differentiating (31) yields $\hat{m} = C\hat{\lambda}$ where C is given by[43]

$$C = \frac{g_\lambda^M - \rho^T \psi_\lambda^M}{-g_m^M + \rho^T \psi_m^M + \psi^M \rho_m^T}\frac{\lambda}{m} \lesseqgtr 0 \qquad \text{with} \quad \text{(A.74)}$$

$$g_\lambda^M = \frac{L\theta_\lambda^M}{L\theta^M - \alpha N} - \frac{L\theta^M}{\left(L\theta^M - \alpha N\right)^2} L\theta_\lambda^M \geq 0 \quad \text{and} \quad \theta_\lambda^M = -\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d \leq 0 \quad \text{(A.75)}$$

$$g_m^M = \frac{L\theta_m^M}{L\theta^M - \alpha N} - \frac{L\theta^M}{\left(L\theta^M - \alpha N\right)^2}\left(L\theta_m^M + \frac{\alpha N}{m}\right) < 0 \quad \text{and} \quad \theta_m^M = \frac{1}{m}\left(f[d^n] - \theta^M\right) < 0 \quad \text{(A.76)}$$

$$\psi_\lambda^M = -\frac{1}{(\eta^M)^2}\eta_\lambda^M > 0 \qquad \text{with } \eta_\lambda^M \text{ as in (A.67)} \quad \text{(A.77)}$$

$$\psi_m^M = -\frac{1}{(\eta^M)^2}\eta_m^M > 0 \qquad \text{with } \eta_m^M \text{ as in (A.66)} \quad \text{(A.78)}$$

$$\rho_m^T = \frac{1}{k\gamma H} > 0. \quad \text{(A.79)}$$

[42] 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 order condition in A.71 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.

[43] 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.

While the denominator of C is always positive (a larger firm size m decreases average cost and increases both the price markup and the wage markdown), 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 markdown (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 (31):

$$\mathrm{d}\ln\left[\theta^M/p\right] = \frac{\partial \ln\left[\theta^M/p\right]\lambda}{\partial\lambda}\hat{\lambda} + \frac{\partial \ln\left[\theta^M/p\right]m}{\partial m}\hat{m} \tag{A.80}$$

with

$$\frac{\partial \ln\left[\theta^M/p\right]}{\partial\lambda} = \frac{L\theta^M_\lambda}{L\theta^M - \frac{\alpha H}{m}} < 0 \qquad \text{and} \qquad \frac{\partial \ln\left[\theta^M/p\right]}{\partial m} = \frac{L\theta^M_m + \frac{\alpha H}{m^2}}{L\theta^M - \frac{\alpha H}{m}} > 0. \tag{A.81}$$

In these equations $\theta^M_\lambda = -\frac{1}{m}\int_0^{d^m} f[d]\mathrm{d}d < 0$ and $\theta^M_m = \frac{1}{m}\left(f[d^n] - \theta^M\right) < 0$. It follows from (A.71) 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.74) and (A.81) to rewrite (A.80) as

$$\begin{aligned}\mathrm{d}\ln\left[\frac{\theta^M}{p}\right] =& \frac{\lambda}{\left(L\theta^M - \alpha N\right)\left(-g^M_m + \rho\psi^M_m + \psi^M\rho_m\right)} \\ & \times \left[\left(L\theta^M_m + \frac{\alpha N}{m}\right)\left(g^M_\lambda - \rho\psi^M_\lambda\right) + \left(-g^M_m + \rho\psi^M_m + \psi^M\rho_m\right)L\theta^M_\lambda\right]\hat{\lambda}.\end{aligned} \tag{A.82}$$

We know that the first fraction on the right-hand side above is positive, hence we must show that the square-bracketed term is negative. Using

$$\left(L\theta^M_m + \frac{\alpha N}{m}\right)g^M_\lambda = \left[\frac{L\theta^M_m + \frac{\alpha N}{m}}{L\theta^M - \alpha N} - \frac{L\theta^M\left(L\theta^M_m + \frac{\alpha N}{m}\right)}{(L\theta^M - \alpha N)^2}\right]L\theta^M_\lambda \text{ and} \tag{A.83}$$

$$L\theta^M_\lambda g^M_m = \left[\frac{L\theta^M_m}{L\theta^M - \alpha N} - \frac{L\theta^M\left(L\theta^M_m + \frac{\alpha N}{m}\right)}{(L\theta^M - \alpha N)^2}\right]L\theta^M_\lambda, \tag{A.84}$$

we can reduce the expression in squared brackets on the right-hand side of (A.82) to $L\theta^M_\lambda\left(\frac{\frac{\alpha N}{m}}{L\theta^M - \alpha N} + \psi^M_m\rho + \rho_m\psi^M\right) - \left(L\theta^M_m + \frac{\alpha N}{m}\right)\rho\psi^M_\lambda$. This is negative since $\theta^M_\lambda < 0$ and

$\psi_\lambda^M > 0$. Hence, a decrease in λ raises real income also if it leads to exit of firms. This completes the proof.

A.8.6 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 $g^M[m, \bar{\lambda}] = \rho^T[m/k]\psi^M[m, \bar{\lambda}]$. 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.74). 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.75). This completes the proof.