

Globalization and Wage Inequality¹

Elhanan Helpman

Harvard University and CIFAR

December 2, 2016

Abstract

Globalization has been blamed for rising inequality in rich and poor countries. Yet the views of many protagonists in this debate are not based on evidence. To help form an evidence-based opinion, I review in this paper the theoretical and empirical literature on the relationship between globalization and wage inequality. While the initial analysis that started in the early 1990s focused on a particular mechanism that links trade to wages, subsequent studies have considered several other channels, and the quantitative assessment of the size of these influences has been carried out in multiple studies. Building on this research, I conclude that trade played an appreciable role in increasing wage inequality, but that its cumulative effect has been modest, and that globalization does not explain the preponderance of the rise in wage inequality within countries.

Keywords: international trade, college wage premium, inequality, residual inequality

¹ This is the background paper for my Keynes Lecture in Economics delivered to the British Academy on September 28, 2016. I thank Gene Grossman and Stephen Redding for comments.

I. Background

While measures of globalization—which assess the importance of international transactions in the world’s economic activity—may include international trade in goods and services, international trade in financial and nonfinancial assets, and international migration, this article centers on the role of merchandise trade in the globalization process. The role of international commerce in the evolution of the world’s economy has a long history, dating back to biblical times more than three thousand years ago and extending to the Roman empire, the Dark Ages, the Middle Ages, and the post-Industrial Revolution era (see McCormick 2001, Findlay and O’Rourke 2007, and Helpman 2011, chap. 1). Despite this long history, imports plus exports as a share of the value of output were very small until the beginning of the 19th century. Starting around 1820, when the value of world imports plus exports amounted to 2% of the value of the world’s output, the value of trade relative to income started to climb, and it kept climbing until the outbreak of World War I, exceeding 22% in 1913. Judged by the evolution of the share of international trade in income, there were two waves of globalization since the beginning of the 19th century; the first started in the early part of the 19th century and ended with World War I, while the second started after World War II and continues until this very day. The ratio of trade to output declined between the wars.²

Likewise, the growth rate of income per capita in the world economy was negligible until the 19th century (see Maddison, 2001). Starting from 1820 it accelerated and remained high until World War I. The growth rate was lower during the years between the two world wars, and then reached an unprecedented peak between World War II and the oil crisis of 1973, a period known as the Golden Age of economic growth. Moreover, even after the oil crisis the rate of growth of income per capita remained high by historical standards (see Helpman 2004, chap. 1). Evidently, trade and growth followed similar trajectories during these historical episodes.

Bourguignon and Morrisson (2002) constructed a data set from which they estimated the evolution of inequality of personal income in the world beginning with the first wave of globalization and ending in 1992. Using two common measures of inequality, the Gini coefficient

²The rise in the trade–income ratio (imports plus exports relative to GDP) was temporarily interrupted after the oil crisis of 1973 and the recession of 2008; see <http://data.worldbank.org/indicator/TG.VAL.TOTL.GD.ZS?end=2015&start=1960>, accessed on July 31, 2016. According to these data, the trade–income ratio exceeded 50% in 2008. I thank Alan Taylor for providing the data cited in the text.

and the Theil index, they showed that inequality increased dramatically during that time span. In their data the Gini coefficient was 0.5 in 1820 and close to 0.66 in 1992, while the Theil index was slightly above 0.52 in 1820 and exceeded 0.85 in 1992. Importantly, they decomposed the inequality of the world's distribution of personal income into a within- versus a between-country component, where the latter is calculated under the counterfactual supposition that the income level of every individual within a country equals the country's per capita income. The Theil index is particularly suitable for this type of decomposition because the within- and between-country Theil indexes add up nicely to equal the index of the world's distribution of income. According to updated data reported in van Zanden, Baten, Foldvari and van Leeuwen (2014), which extends to the year 2000, overall inequality increased steadily until the middle of the 20th century and reached a peak in 1975. Importantly, however, the rise in inequality was primarily driven by the rise in inequality between countries; that is, over time, income per capita increased faster in rich countries than in poor countries, thereby widening the gap of income per capita between them.

Between the early 1980s and 2000 the Theil index of within-country inequality and the Theil index of between-country inequality changed little. Nevertheless, due to economic growth, extreme poverty declined dramatically in the world economy. Extreme poverty was originally defined as income of less than one dollar a day, and this threshold has been updated to less than 1.25 dollars a day adjusted for purchasing power parity (PPP) in 2005. In 1981 close to two billion people, a bit more than half the population, lived in extreme poverty, while by 2008 this number had declined to less than 1.3 billion, or a little over 22% (see Anand and Segal, 2015, Table 11.8). Noting that the world's population increased by about 50% between 1981 and 2008, this reduction in poverty is impressive indeed.

These long-term trends in the world economy suggest that economic growth, globalization in the form of expanding foreign trade, and income inequality are intertwined. Recent concerns about these issues are more narrowly focused, however, because in as much as the contribution of within-country inequality to total inequality of the world distribution of personal income was stable between 1981 and 2000, big changes occurred in individual countries afterwards. While inequality of income declined in a number of emerging economies, particularly in Latin America, it increased significantly in most OECD countries.³ But earnings gaps between skilled and unskilled workers,

³Between the mid-1980s and 2013, income inequality declined slightly in Greece and Turkey, and changed very little in Belgium, the Netherlands and France. In all other OECD countries it increased

which played a large role in rising income inequality from the late 1970s to the early 1990s, also increased in many less-developed countries that managed to significantly close the gap in their income per capita with the rich countries. The sources of this rise in inequality have been hotly debated and a great deal of research has attempted to unearth them.

II. The Rise of the College Wage Premium

In 1913 the top 1% of U.S. income earners received 18% of U.S. income. Their share declined to 8% in the mid-1970s, after which it started to climb, reaching 13% in 1990 and 18% in 2008.⁴ This U-shaped form of the share of top earners in a country's income is not unique to the United States, as shown by Atkinson, Piketty and Saez (2011). In other English-speaking countries, such as Canada and the United Kingdom, inequality evolved along similar lines. And moreover, this pattern is not restricted to top incomes; other measures of inequality, such as the Gini coefficient, show similar trends. For our purposes the more important observation is that inequality has been rising in many countries since the late 1970s, and this includes not only the English-speaking countries, the United States, United Kingdom, Canada, Ireland, Australia and New Zealand, but also the Nordic countries Sweden, Finland and Norway, as well as poorer countries such as India and China. An important source of this rise in inequality has been rising inequality of labor income, which attracted much attention in the early 1990s (see Katz and Autor, 1999).

Katz and Murphy (1992) pointed out that despite the continuous rise in the number of college graduates relative to high school graduates in the U.S. economy, the college wage premium rose sharply in the 1980s. This trend continued in later years. Autor (2014) showed that the relative supply of college graduates, measured by the share of their hours in the aggregate number of hours worked by the adult population, increased continuously from 1963 to 2012. At the same time, the college wage premium—which expresses in percentage terms how much more a college graduate earns than a high school graduate—followed a humped shape between 1963 and 1979 and sharply increased thereafter. According to these updated data, the college wage premium was 48% in 1979

substantially, and especially so in Finland, Sweden, New Zealand, the United States and Mexico (see OECD, 2015, Figure 1.3).

⁴Source: The World Wealth and Income Database (WID), <http://www.wid.world>, accessed on September 12, 2015.

and had doubled to 96% by 2012.⁵ In 1987, the last year in the Katz and Murphy sample, the college wage premium was 63%.⁶

From the vantage point of 1992, Katz and Murphy asked whether the rise in the college wage premium during the 1980s was driven by supply or demand factors, and they concluded that the dominant cause was an increase in the relative demand for skilled workers, and especially for those with a college degree. Following an analysis of alternative explanations, including the impact of foreign trade on wage inequality, they concluded: “Although much of this shift in relative demand can be accounted for by observed shifts in the industrial and occupational composition of employment toward relatively skill-intensive sectors, the majority reflects shifts in relative labor demand occurring within detailed sectors. These within-sector shifts are likely to reflect skill-biased technological change” (Katz and Murphy, 1992, p.37). The conclusion that skill-biased technological change was the principal cause behind the widening gap in the wages of skilled and unskilled workers was subsequently echoed in additional studies. Other factors, such as the decline of unionization, the decline of the minimum wage, and deregulation of labor and product markets, did not appear to play a large role in the rise of the college wage premium.⁷ Attempts to disentangle the contribution of trade from the contribution of technology came out in favor of technology.

III. What Does Basic Theory Teach Us?

A celebrated result from the factor proportions trade theory, known as the Stolper-Samuelson Theorem, was used to interpret the rising college wage premium. According to this theorem, in a country that trades low-skill-intensive and high-skill-intensive products on international markets

⁵I am grateful to David Autor for providing these data.

⁶Measured in constant 2012 dollars, the college wage premium increased from 17.4 thousand dollars in 1979 to about 35 thousand in 2012; see Autor (2014).

⁷See Bourguignon (2015, chap. 3) for a review of evidence concerning these factors in a variety of countries. In the United States, for example, the debate concerning the impact of the minimum wage on inequality led to a nuanced conclusion. Card and DiNardo (2002) argued that the decline of the real value of the minimum wage during the 1980s played a dominant role in the rise of wage inequality. On the other side, Autor, Katz and Kearney (2008) showed that the minimum wage had an impact on inequality at the lower end of the wage distribution, but not at the upper end where inequality increased most. In an updated recent analysis, Autor, Manning and Smith (2016) found that the decline in the real minimum wage explains 30 to 40 percent of the rise in wage inequality at the lower tail (the 50/10 percentile ratio) in the 1980s. But they also point out that they cannot reject the hypothesis that the spillovers of wages from the minimum to higher percentiles is spurious due to measurement errors.

an increase in the price of low-skill-intensive products raises the real wage of the country's workers with low skills and reduces the real wages of its workers with high skills.⁸ And, alternatively, if the price of low-skill-intensive products falls, then the real wage of low-skilled workers declines while the real wage of high-skilled workers rises. In the former case the gap between the wages of low-skilled and high-skilled workers narrows while in the latter case it widens. These wage outcomes do not depend on the sources of price movements; they can result from a country's change in trade policy, such as magnified import protection or trade liberalization, or from changes that occur in other countries that trade on international markets.⁹

To interpret the rising college wage premium in light of this theorem, consider the following scenario. A group of less-developed countries expands its participation in foreign trade by joining the World Trade Organization or by reducing barriers to trade. Since compared to rich countries they specialize in low-skill-intensive products, their expanded role in world trade reduces the relative price of low-skill-intensive products. Under the circumstances the Stolper-Samuelson Theorem predicts that the wages of low-skilled workers should decline in rich countries relative to the wages of high-skilled workers. Regarding high-skilled workers as college graduates then implies that the college wage premium should rise in the rich countries.

There is, however, a flip side to this argument. For less-developed countries to sell more low-skill-intensive products on world markets the relative price of these goods has to *rise* in their home markets. The logic of the Stolper-Samuelson Theorem then implies that the college wage

⁸In Stolper and Samuelson (1941), where the original theorem is stated, the assumptions restrict the economies to include two constant-returns-to-scale sectors and two factors of production, and the factors of production are labeled labor and capital. Moreover, one sector is capital intensive and the other is labor intensive, in the sense that the former uses more capital per worker for a given wage rate and rental rate on capital. Under the circumstances an increase in the price of labor-intensive products raises the nominal and real wage and reduces the nominal and real reward to capital. An increase in the price of capital-intensive products has the opposite effects. The same logic applies, of course, when instead of labor and capital the economy uses skilled and unskilled workers. An extension of this result is provided in Jones and Scheinkman (1977); they show that with many types of inputs and many types of sectors, all producing under constant returns to scale, an increase in the price of a product raises the nominal and real reward of some inputs and reduces the nominal and real reward of some other inputs.

⁹Indeed, an analysis of the impact of trade protection on wages was the original motivation for Stolper and Samuelson (1941), who assumed that one sector is labor intensive and the other is capital intensive. A tariff raises the domestic price of import-competing products, as a result of which labor gains when the import-competing sector is labor intensive and labor loses when the import-competing sector is capital intensive.

premium should *decline* in these countries. In other words, the relative price of low-skill-intensive products and the skill premium should change in opposite directions in rich and poor countries.¹⁰

Additional implications of this theory concern relative factor use. A higher college wage premium in rich countries induces manufacturers to economize on skilled workers. By the same logic a lower college wage premium in poor countries induces manufacturers to economize on low-skilled workers. For that reason the ratio of high- to low-skilled employees should decline in rich countries and rise in poor countries. Evidently, once we subscribe to this mechanism we also buy into certain subsidiary implications. The empirical validity of these subsidiary implications provides a test of the extent to which the Stolper-Samuelson mechanism is suitable for explaining the rise in the college wage premium.¹¹

IV. Evidence: The First Pass

The first attempts at empirically evaluating the role of foreign trade in raising the college wage premium heavily relied on the factor proportions trade theory. Katz and Murphy (1992) used factor content analysis to compute shifts in labor demand induced by U.S. imports and exports. Factor content analysis consists of computing the services of various factors of production embodied in a country's exports and imports. By adding the net amounts of these services (from exports minus imports) to the country's factor endowment one obtains a notional country with the same characteristics as the original country except for its factor endowment. The autarky equilibrium of the notional country is then the same as the trade equilibrium of the original country in terms of prices, factor rewards and consumption levels, which means that the output levels of the notional country's exportables equal the original country's output levels minus exports, and the output levels of the notional country's importables equal the original country's output levels plus imports. In the absence of trade the country would be in autarky with the original factor endowment. Therefore the gap between the notional country's factor endowment and the original factor endowment, which equals the factor content of trade, represents the implicit *addition* of factor availability made possible by foreign trade. This modification of factor supplies affects factor

¹⁰These opposite shifts require the relative prices of low-skill-intensive products to be higher in rich countries prior to the trade expansion by less-developed countries, which is congruent with the presence of import protection.

¹¹Rodrik (2015) provides an excellent discussion of this type of use of models in economics.

rewards in contrast to autarky. An increase in supply depresses a factor's reward, while a reduction in supply raises its reward.¹² A similar analysis applies to changes in trade that result from changes in world prices by comparing the factor contents of trade before and after the price changes.

Katz and Murphy found that changes in U.S. trade flows embodied flows of factor content that increased the demand for skilled relative to unskilled workers, thereby contributing to the rise in the wage gap between them. Yet, "Although trade-induced changes in relative demand move in the correct direction to help explain rising education differentials in the 1980s, they are quite small relative to the increase in the relative supplies of more-educated workers over the same period" (p. 65). They therefore concluded that foreign trade did not play a big role in the rise of the college wage premium. In a more detailed study of the factor content of trade flows with less-developed countries between 1980 and 1995, Borjas, Freeman and Katz (1997) concluded that trade accounted for 20 percent of the rise in the U.S. college wage premium. Immigration, which consisted primarily of workers with less than a college degree, also raised the relative supply of low-skilled workers. But the impact of these workers on the college wage premium was limited, not exceeding the impact of trade expansion with less-developed countries.¹³

The Stolper-Samuelson mechanism was also employed by Krugman (1995) for evaluating the influence of trade with less-developed countries on wages. Although, he argued, in theory this mechanism can explain the empirical pattern, reasonable parametrization of the theoretical model leads to the conclusion that it cannot explain the *magnitude* of the rise in the college wage premium.

A quantitative assessment of the impact of trade with less-developed countries on the U.S. (or some other country's) college wage premium has to use an estimate of changes in the relative price of high-skill-intensive products, or prices of exports relative to imports (the terms of trade), and it has to use an estimate of the impact of such price changes on changes in the relative wage of skilled workers. The Stolper-Samuelson Theorem implies that an increase in the relative price of high-skill-intensive products, be it due to the rise in the price of these products or a decline in the price of low-skill-intensive products, has a positive impact on the relative wage of skilled

¹²See Krugman (2008) for a clear exposition. This argument rests on some assumptions that are common in the factor proportions trade theory, and the last sentence strictly applies to a two-factor environment only.

¹³U.S. imports from less-developed countries were small in the 1980s, about 2% of GDP, which was about half the import volume from developed countries (see Krugman, 2008, Figure 2).

workers. It follows that the extent to which this mechanism explains the rise in the college wage premium depends on whether relative prices of high-skill-intensive products increased in the 1980s and how large that increase was, and on the size of the coefficient that translates relative price changes into changes in relative wages.¹⁴ In other words, the impact of trade on the college wage premium depends on how large this combined effect is in practice.

Lawrence and Slaughter (1993) estimated the relationship between sectoral skill intensity, measured as the employment of nonproduction relative to production workers, and price changes. They found no evidence that during the 1980s prices of high-skill-intensive products increased in the U.S. more than prices of low-skill-intensive products.¹⁵ Leamer (1998) tracked sectoral prices relative to the overall producer price index during three decades: the 1960s, 1970s and 1980s. Textile and apparel were the low-skill-intensive sectors in his sample, and he found that their relative prices declined markedly, by 30%, only in the 1970s. In the 1980s, when the college wage premium soared, the relative prices of these products increased only slightly. One could of course argue that the price changes in the 1970s had the biggest impact on wages only in the 1980s, because the transmission of price shocks into wages was slow. But the credibility of this argument—which is not grounded in evidence—is questionable, even if one believes that the adjustment of wages to prices is not contemporaneous. Leamer (1998) also estimated “mandated” factor price changes; that is, factor price changes mandated by the zero profit condition in competitive markets. In this approach prices equal marginal costs and therefore changes in factor prices result from either product price changes or changes in productivity. Comparing the resulting estimates of mandated wage changes with wage data provides a test of the soundness of these

¹⁴The size of the coefficient that translates relative price changes into changes in relative wages depends on a host of characteristics of the model economy deployed for the analysis. For the Stolper and Samuelson (1941) economic structure it depends only on the factor intensities of the two sectors as measured by factor shares in production costs (see Jones, 1965). If instead one envisions a sector-specific structure in which there are two industries and three factors, one of the factors being specific to one industry and another factor being specific to the other industry and the third factor being mobile between them, then this coefficient depends on the factor intensity of each sector, the relative size of each sector, and the elasticity of substitution between the two inputs in each one of them (see Jones, 1971). By regressing relative factor rewards on relative prices one obtains a reduced-form estimate that depends on these structural features. This estimate can then be interpreted as a “sufficient statistic” for the impact of relative prices on relative factor rewards. See Chetty (2009) for a review of the “sufficient statistic” approach in economics.

¹⁵There was a controversy surrounding these estimates. For example, Sachs and Shatz (1994) found a positive effect by isolating the computer industry. They justified this formulation with the argument that computer prices were grossly mismeasured.

estimates. It turns out, however, that these estimates are very sensitive to how much of the productivity growth is assumed to feed into prices, which limits the trustworthiness of the inferences drawn from this analysis (see Feenstra, 2015, pp. 87–91). Be this as it may, Leamer’s main conclusion was that Stolper-Samuelson effects were strong in the 1970s but not in the 1980s. Apparently, other mechanisms are needed to account for the rise of the college wage premium in the 1980s.¹⁶

Another challenge to the Stolper-Samuelson mechanism was presented by evidence from less-developed countries. A number of these countries implemented unilateral trade reforms in the 1980s and early 1990s, including Argentina, Brazil, Colombia, India and Mexico. Tariff reductions were far-reaching in these episodes, as they were in Chile’s trade liberalization in the 1970s (see Goldberg and Pavcnik, 2007). According to the theory these policies should have reduced the reward to skilled relative to unskilled workers (see Section 3), yet relative wages responded in the opposite direction. And similarly to what happened in the rich countries, the use of skilled workers increased within sectors despite the rise in their relative cost (see Goldberg and Pavcnik, 2007).

V. Trade vs. Technology

As pointed out in Section 2, Katz and Murphy (1992) concluded that the rise of the U.S. college wage premium during the 1980s was most likely caused by skill-biased technological change, and the majority of the shift in relative labor demand occurred within rather than between sectors. What skill-biased technological change means in this context is that the efficiency of skilled labor increased faster than the efficiency of unskilled labor.¹⁷ If, alternatively, a decline in the relative price of low-skill-intensive products were the foremost cause of the rise in the college wage premium, we would have observed a reallocation of factors of production from low-skill- to high-skill-intensive sectors on the one hand and a reduction in the employment of high-skilled relative to low-skilled workers within all manufacturing industries on the other. Since the supply of college graduates increased markedly during that time period relative to the supply of workers with lower

¹⁶See Slaughter (2000) for a more detailed discussion of these issues and a review of additional evidence.

¹⁷A simple way to understand this definition is to imagine that output depends on the amounts of “effective” units of labor and that technological change raises the effective units for a given number of workers (or hours). Skill-biased technical change then implies that, given a fixed number of low-skilled and high-skilled workers, the number of effective units rises faster among the skilled.

education levels, the allocation of college graduates to high-skill-intensive sectors should have been massive. No such shifts took place, however. According to the evidence in Berman, Bound and Griliches (1994), the relative employment of skilled workers *increased* in all manufacturing industries, and these within-sector changes account for the majority of the rise in the aggregate employment of skilled relative to unskilled workers in manufacturing. The same type of employment shifts took place in other rich countries during the 1980s. While in the United States more than 70 percent of the increase in the employment of skilled (nonproduction) workers occurred within manufacturing sectors, much larger shares (more than 90 percent) were recorded in Australia, Belgium and the U.K.

Skill-biased technological change can explain the above-described employment shifts together with the surge in the college wage premium, at least in theory. Furthermore, considerable evidence shows that advances in technology shifted factor demand toward highly skilled workers; sectors with faster increases in the demand for such workers were more innovative, more intensive in R&D, and more intensive in computer use.¹⁸ On the other side, Machin and van Reenen (1998) found that in seven OECD countries the share of imports originating from less-developed countries played a minor role in explaining the rise in the employment of skilled workers within industries. This type of evidence was interpreted to imply that the role of foreign trade in the rise of the college wage premium was modest at best.

Leamer (2000) objected to this interpretation, arguing that for given world prices the data on shifts in wages and employment are not consistent with pure skill-augmenting technological change: that is, a productivity improvement of every skilled worker. If this were the case, wages of low-skilled workers would not change and wages of high-skilled workers would rise in proportion to the rate of technological improvement, so that wages *per effective* unit would not change. Under these circumstances the growth of effective units of skilled labor should not change factor proportions within industries, measured in effective units, but should shift resources from low-skill-intensive to high-skill-intensive sectors.

While theoretically correct, Leamer's argument relies on the assumption that world prices of goods do not change, which would be appropriate if the skill-biased technical change were a

¹⁸See Berman, Bound and Griliches (1994) and Autor, Katz and Krueger (1998) for the U.S. evidence and Machin and Van Reenen (1998) for comparable evidence from the U.S., the U.K., France, Germany, Denmark, Sweden and Japan.

localized phenomenon in some small country. As pointed out by Krugman (2000), however, this assumption is quite inappropriate when the improvement in technology is widespread. In the latter case there is a direct effect on wages, captured by Leamer's analysis, and there is an indirect effect through changes in final product prices set in motion by the resulting supply shifts. Working out a complete model embodying these considerations, Krugman showed that the final outcomes are theoretically consistent with the patterns in the data.

Was skill-biased technical change ubiquitous? The evidence points to an affirmative answer. Berman and Machin (2000) showed that in the 1980s the within-industry contribution to increases in nonproduction workers' wage-bill shares was large in all 12 rich countries in their sample with the exception of Sweden, and that changes in sectoral nonproduction workers' wage-bill shares were positively correlated across these countries. For 9 of them, the upswings in their wage-bill shares were positively correlated with the U.S. upswings, and only Austria and Belgium had a few negative correlations with other countries (see their Table 2). Berman and Machin also showed that during the same time period the within-industry contribution to the increase in nonproduction workers' wage-bill shares was large in all of the 18 middle-income countries in their sample with the exception of Korea, as well as in the poor countries in their sample with the exception of Bangladesh. Furthermore, sectoral skill upgradings in the poor and middle-income countries were positively correlated with skill upgradings in the U.S. sectors (see their Table 4), and sectoral skill upgradings in all these countries—rich, middle-income and poor—were positively correlated with U.S. computer usage and OECD R&D intensity (see their Table 5). Evidently, changes in technology were widespread and exhibited similar patterns in countries at different levels of development.

Do similar patterns of wage changes in rich and poor countries necessarily contradict a foreign trade-based explanation of the rising college wage premium? Undeniably, this evidence is at odds with the implications of the theoretical analysis in Section 3. Despite that, Feenstra and Hanson (1996, 1997) managed to develop a sensible modification of the theoretical model in order to qualitatively square the theory with this evidence. To this end they proposed to view the production process as a collection of many activities or intermediate inputs that differ in factor intensity. In these circumstances rich countries, with a high relative wage of skilled workers, find it profitable to source low-skill-intensive intermediate inputs from poor countries. In other words, the rich countries specialize in high-skill-intensive production while poor countries specialize in

low-skill-intensive production. When sourcing from foreign countries becomes cheaper, be it due to a decline in transport costs or improvements in technology, a rich country stops producing some of its least-skill-intensive intermediates and instead sources them from a poorer country. This change in the sourcing pattern can take place within firm boundaries by multinational corporations, as suggested by Feenstra and Hanson, or at arm's length, as suggested by Zhu and Trefler (2005). In both cases the reallocated activities are least-skill-intensive in the rich country and most-skill-intensive in the poor country. As a result, the demand for high-skilled labor rises in both countries relative to low-skilled labor, bidding up the relative wage of skilled workers in both. Evidently, this modified model has a built-in mechanism through which trade can raise the college wage premium in rich and poor countries alike. And moreover, if all these activities and intermediate inputs are concentrated in the same industry, it also predicts a rise in the relative use of skilled workers within industries.

Feenstra and Hanson (1997) used data on U.S. multinational corporations that operated assembly plants in Mexico during the 1980s to examine these implications. Those assembly plants, erected along the U.S. border, imported intermediate inputs from the U.S. and shipped back assembled products. This business strategy was profitable because the assembly could be done primarily by unskilled Mexican workers. Feenstra and Hanson found that U.S. foreign direct investment in these plants, known as maquiladoras, was positively correlated with the rise of the share of skilled labor in Mexico's wage bill; in regions with larger foreign investment, the wage share of skilled labor increased more.

While this evidence confirms that offshoring of intermediate inputs affects relative wages in the destination country, it does not tell us how important this mechanism was in shaping U.S. wages. The latter was addressed in Feenstra and Hanson (1999). Studying the 1970s and 1980s, they sought to evaluate the importance of offshoring versus technology in shaping U.S. wages. Their estimates proved to be sensitive to the measure of sectoral use of high-tech equipment, which they employed as a proxy for a sector's technology level. Using the share of high-tech equipment in the capital stock as a measure of technology led them to conclude that about a quarter of the rise in the relative wage of nonproduction workers during the 1980s (1979–1990) was due to offshoring and around 30% was due to technology. But once more weight was given to more recent equipment, which presumably was more advanced, the contribution of offshoring halved and the contribution of technology more than tripled. In their review of the literature, Feenstra and Hanson

(2003) re-estimated these relationships, giving the trade explanation as good a chance as possible, but the results did not change much.

My conclusion from the literature discussed so far is that while several mechanisms that link globalization to the relative wages of skilled workers have affected U.S. wages during the 1980s, their impact on wage inequality was modest. Furthermore, although technological change has most likely played a bigger role, there is a paucity of quantitative evidence concerning its effects. In too many cases technological change is used as a default explanation when other alternatives are not compelling.¹⁹

VI. Broadening the Canvas

A vibrant literature has recently re-examined the relationship between foreign trade and wages, motivated by new theoretical developments on the one hand and new facts on the other. Three expansions of the standard model are at the core of this enterprise: firm heterogeneity within industries, worker heterogeneity beyond the classification into two groups of low-skilled and high-skilled individuals, and labor market frictions such as unemployment, wage bargaining and costly mobility. Each of these features adds a distinct facet to the theory, helping to address issues that were beyond the reach of previous scholarly work.

1. Firm Heterogeneity

Firm heterogeneity was introduced into trade theory in response to the discovery during the 1990s of new patterns in previously unavailable data sets. In these data firms exhibited substantial heterogeneity within industries in terms of productivity and size, and only a fraction of firms exported. Furthermore, exporters differed systematically from nonexporters, with exporters being larger and more productive.²⁰ Melitz (2003) provided the canonical model that is consistent with

¹⁹In fact, technological change plays a significant role in the Feenstra and Hanson (1996, 1997, 1999) story line concerning the role of foreign trade, because it is used to motivate the rise in offshoring.

²⁰Bernard and Jensen (1995, 1999) discovered these patterns in U.S. data, while subsequent studies confirmed them in many other countries, including Canada, Colombia, France, Mexico, Morocco, Spain and Taiwan (see Helpman, 2011, chap. 5).

these patterns, and various elaborations of his approach were applied to the study of trade and wages.²¹

Melitz assumed that labor is homogeneous and entrepreneurs pay an upfront “entry” cost to acquire a manufacturing technology. The entry cost may consist of R&D or the cost of forming a business enterprise. Importantly, however, the productivity of the manufacturing technology becomes known only after the entry cost is sunk, and only the *distribution* of productivity is known when the entry decision is made. A company’s business strategy is formed after entry, when the productivity of its technology becomes known. At that stage staying in business entails bearing a fixed operating cost in every period. For this reason only firms with a high enough productivity level are profitable, while low-productivity firms are not. The latter cut their losses by closing shop. Companies that stay in business may also export, except that exporting entails a fixed cost of establishing a beachhead in every destination country. For this reason only firms with high enough productivity levels can profitably export. In sum, not all entrants into an industry stay, and among those who do the more-productive enterprises export while the less productive serve only the domestic market. This structure replicates the main patterns in the data.²²

In Melitz (2003) all workers are paid the same wage, independently of whether they are employed by high- or low-productivity firms, by exporters or by nonexporters. For this reason international trade impacts the wage level but not wage inequality. Nevertheless, as we shall see below, by adding more features firm heterogeneity can generate a nondegenerate wage distribution that responds to foreign trade.

2. Assortative Matching

Matching has a long tradition in economics, be it for the assignment of firms to locations, of individuals to houses, or workers to firms. Becker (1973) applied it to marriages, deriving a condition under which there is Positive Assortative Matching (PAM). For illustrative purposes, suppose that there are a fixed number of “men” and a fixed number of “women” and the number of men equals the number of women. Moreover, men can be ranked by a single characteristic from low to high and so can women. A marriage consists of pairing a man and a woman, and every pair

²¹An alternative, less-influential model, was developed by Bernard, Eaton, Jensen and Kortum (2003). See, however, the discussion of Burstein and Vogel (2016) below for an interesting application.

²²See Melitz and Redding (2014) for a review of the original contribution and its many extensions.

produces a value based on the characteristic of the man and the characteristic of the woman, and this value is higher the higher the characteristic of either the man or the woman.

What types of matches maximize aggregate value? Becker showed that if the value of a match exhibits complementarity, then positive assortative matching maximizes the aggregate value of marriages; that is, the man with the largest value of the masculine characteristic is matched with the woman with the largest feminine characteristic, the man with the second largest masculine characteristic is matched with the woman with the second largest feminine characteristic, and so on, until the man with the lowest masculine characteristic is matched with the woman with the lowest feminine characteristic. Complementarity of the values of matches means that the increase in the value of a match when the masculine characteristic is replaced with a larger one is larger the larger the feminine characteristic in the match. In other words, the marginal gain from a masculine characteristic is increasing in the attractiveness of the woman in the match. And symmetrically, the marginal gain from a feminine characteristic is increasing in the attractiveness of the man in the match. This property is also known as supermodularity. Becker then showed that in a “competitive” marriage market the resulting marriages satisfy PAM.

The same logic can be applied to matching workers with managers or firms. All we need is to identify a worker characteristic, say ability, and a characteristic of managers, say managerial ability, or a characteristic of firms, say technological sophistication, in order to study the matching of workers to managers or firms. An important difference between these types of matches and those in the above-described marriage market is that while one man is typically matched with one woman in a marriage, many workers are matched with a single manager or a single firm. These models use a stronger notion of complementarity in order to obtain clear portrayals of inequality; they assume that the *natural logarithm* of the value of a match exhibits complementarity. This property is also known as log supermodularity. This implies that a marginal increase in the characteristic of one party raises the marginal value of the other party’s characteristic *proportionately more* than the value of the match.²³

This last feature has the following implication. Suppose, for concreteness, that workers with ability levels between a and b , measured in appropriate units (e.g., years of schooling) are

²³For this condition to be necessary and sufficient for the inequality results derived in these studies, characteristics and quantities of the two parties have to interact in specific ways; see Eeckhout and Kircher (2016).

matched with firms whose technological sophistication lies between c and d , also measured in appropriate units. Then workers with higher ability are matched with more-sophisticated firms. In particular, workers with ability a are matched with firms whose technological sophistication is c while workers with ability b are matched with firms whose technological sophistication is d . In this event more-able workers are paid higher wages, and the rate at which wages rise with ability depends on how strong the complementarity between worker ability and firm sophistication is in the productivity function. The rate of wage increase determines in turn wage inequality in this ability range.

Now suppose that due to a change in the economic environment (e.g., a change in relative product prices), the workers with abilities a to b match with more-sophisticated firms, so that every worker is now employed by a more-sophisticated firm. Under the circumstances the relative wage gap between any pair of workers with different ability levels is now larger than it was before. For this reason wage inequality is now higher than it was before.

This illustrates an important property of economies of this type: shifts in matching can aggravate or mute wage inequality. For this reason, globalization can impact inequality through its influence on the assortative matching of workers with firms. Ideas of this type have been applied to models with heterogeneous workers in order to study the impact of trade on wages. They make it possible to examine the impact of trade on wage inequality at different parts of the distribution, comparing for example inequality at the top-end with inequality at the bottom-end of the distribution (see below).

3. Labor Market Frictions

There is little doubt that labor markets are subjected to a variety of frictions that prevent instantaneous adjustment of employment and wages. Some of these frictions are designed by governments, such as minimum wages or firing costs, other are ingrained in a functioning economy, such as the cost of finding a job or the cost of switching jobs. The latter may arise in turn from costs of moving to a different location, to a different industry, or to a different occupation. Additionally, in many countries labor unions play a major role in wage determination, be it within firms, within industries, or at the country level.

Many studies examined the influence of labor market frictions—which vary substantially across countries—on foreign trade and wages, shedding light on wage inequality (see Helpman,

2011, chap. 5 for a review). Two key implications of such frictions are that unemployment emerges as a natural outcome, providing business firms have leverage over labor compensation. Details of the mechanism differ, depending on the form of the labor market frictions, but the qualitative outcomes are often comparable.

Search and matching in labor markets, of the type developed by Mortensen and Pissarides (1994) and Diamond (1982a,b) for the study of macroeconomic determinants of unemployment, is a prime ingredient in recent studies of trade and wages. In this framework firms post vacancies and unemployed workers search for jobs. Workers are matched with vacancies, but only some workers succeed in finding a job and only some vacancies are successfully filled. The degree of success of the matching process depends on characteristics of the labor market; in more-efficient markets more matches are realized and some markets favor workers more. Matched firms and workers engage in wage bargaining. Failure to reach an agreement is costly to the workers and the firms, because it raises the number of unfilled vacancies for firms and the number of unemployed for workers. As a result every party has an incentive to reach an agreement. Understandably, wage bargaining takes place in the shadow of these costs, which consequently impact the wage agreement. International trade modifies the choices available to firms and the employment opportunities available to workers. Through changes in these options trade modifies wages and employment.

Skills play a noteworthy role in shaping individual earnings, notwithstanding the fact that luck matters too. Yet skills are difficult to measure, and they depend on a host of individual characteristics such as ability, talent, schooling, experience, and the like. Since the work of Mincer (1974), however, three observable worker characteristics—education, experience and gender—have been used for explaining differences in wages across individuals. As successful as this approach has been, it explains only a fraction of the variation in wages. The residual fraction—which cannot be explained by observable worker characteristics—is referred to as “residual inequality” (see Katz and Murphy, 1992). Residual inequality increased over time and the sources of this increase are still debated. The next two sections discuss the impact of trade on wages through observable worker characteristics, while the following section discusses the impact of trade on residual inequality.

VII. Observable Attributes

Needless to say, it is necessary to think about workers as being heterogeneous in order to study how trade impacts the wages of different types of workers. Whereas most of the discussion has so far focused on two types of workers, skilled and unskilled, the matching mechanism introduced in the previous section can encompass richer patterns of worker heterogeneity in order to study wage inequality across the entire spectrum of wages. Besides, matching can take a variety of forms. Workers can be matched with capital equipment, with managers, with firms, or with sectors, and the consequences of trade for wages depend on these particulars. In this section I discuss workers who have observable attributes and firms that have technologies that cannot be modified. In particular, firms cannot invest in R&D in order to enhance their productivity. The possibility of technology upgrading is explored in the next section.

Costinot and Vogel (2010) developed a variant of the factor proportions trade model that features multiple sectors and multiple types of workers, in which markets are competitive and workers are matched with sectors.²⁴ One interpretation of their model is that sectors produce intermediate inputs that are traded internationally, and every country uses these intermediate inputs to assemble its own final consumer goods. Another interpretation is that workers are matched with tasks (i.e., a sector is relabeled to be a task), and the tasks are combined to produce a final product. In the latter case a country relies on tasks performed in other countries for producing its own consumer goods. This interpretation might be appropriate, for example, for trade in business services. Be this as it may, workers differ in a single dimension, call it ability (they call it skill), and the productivity of a worker with a given ability varies across sectors. Critically, the productivity of a worker with a specific ability level depends only on her sector of employment and it does not depend on how many or what type of other workers are employed in this industry. In addition, sectors can be ordered by a single characteristic, say technological sophistication, so that the natural logarithm of output per worker, which depends on the worker's ability and the sector's sophistication, exhibits complementarity; i.e., the productivity function is log supermodular. In these circumstances there is positive assortative matching: higher-ability workers are matched with more-sophisticated sectors in every country.

²⁴Ohnsorge and Trefler (2007) is a predecessor that studies trade and wages with an assignment model in which workers sort across sectors.

In this type of world, international trade has an unambiguous effect on wage inequality when the countries differ in factor endowments in a way that exhibits a clear ranking of relative ability abundance.²⁵ This means the following: Suppose that the world consists of two countries and for any two ability levels one country has relatively more workers with the higher ability (this is known as the monotone likelihood ratio property). In this case we can state unambiguously that this country is high-ability abundant compared to the other country. The opening of trade between a high-ability abundant country and a low-ability abundant country improves matches for all workers in the former and worsens them for all workers in the latter. That is, in the skill-abundant country trade leads to a reallocation of labor across sectors so that every worker is employed in a more-sophisticated industry, and the opposite happens in the other country. As explained in the previous section, when workers are matched with more-sophisticated sectors (or firms) the gap in wages between every higher-ability and every lower-ability worker rises, leading to more wage inequality. It follows that trade raises wage inequality in the high-ability abundant country and lowers wage inequality in the low-ability abundant country. These results are similar in spirit to the implications of the simple two-sector factor proportions trade model with low-skilled and high-skilled workers. Except that here linearity of the technology prevents the mixing of different types of workers in a single sector.

Trade alters matching in more complex situations as well. Of particular interest is trade between countries that differ in the *diversity* of factor endowments (what Costinot and Vogel call North–North trade).²⁶ For concreteness, suppose that there is an ability level such that one country is high-ability abundant above this skill level and low-ability abundant below this skill level, compared to the other country. Then it is reasonable to think about the former country as having a more diverse factor endowment. To be sure, this is a particular way in which endowments differ across countries, but it proves to be instructive for the purpose at hand. In these circumstances trade worsens the matches of low-ability workers and improves the matches of high-ability workers in the country with a more diverse factor endowment, and the opposite occurs in the

²⁵Generally speaking, with many ability levels it is possible for one country to have relatively more higher-ability workers at the lower end of the ability distribution but relatively fewer higher-ability workers at the upper end of the ability distribution, as well as more complicated patterns of relative abundance of abilities.

²⁶Grossman and Maggi (2000) provided the first theoretical analysis of the impact of diversity on trade flows; Bombardini, Gallipoli and Pupato (2012) provided empirical evidence. See Grossman (2013) for a review of this literature.

country with the less diverse factor endowment. As a consequence, wage inequality declines at the lower end of the wage distribution and rises at the upper end of the wage distribution in the more diverse country, and the opposite happens in the less diverse country. This illustrates that globalization does not have to raise or reduce inequality in a country along the entire wage spectrum; it can impact inequality differently at different ability intervals.

The last point comes out even stronger in Grossman, Helpman and Kircher (forthcoming), who study a two-sector economy that employs two factors of production: workers and managers. Factor intensity differs across sectors, so that one sector is worker intensive and the other is manager intensive. Unlike the standard model, however, both workers and managers are heterogeneous, with varying ability levels. As a result, there are two types of matches: matches of inputs with sectors and matches between inputs within sectors. In other words, certain types of workers and managers are employed in the worker-intensive sector and the other workers and managers are employed in the manager-intensive sector. At the same time, within every sector workers match with managers. To distinguish the matching of inputs with sectors from the matching of inputs within sectors, the term *sorting* is used for the former and the term *matching* is used for the latter.

The total factor productivity of a firm that employs a manager with workers who share a common ability level is higher the higher the ability of either the manager or the workers; the productivity function is log supermodular and it depends only on the attributes of the workers and the manager, notwithstanding the fact that there is diminishing marginal productivity in the number of workers per manager.²⁷ Furthermore, while individual firms have no incentive to mix the abilities of managers or workers, every sector employs workers and managers with multiple abilities.

What types of workers and managers are employed in each sector depends on technological features and product prices. Through product prices international trade affects the sorting of inputs to sectors and thereby the matching pattern between workers and managers within every sector. The matching pattern determines in turn the inequality of workers' wages and managers' earnings. While log supermodularity of productivity ensures positive assortative matching within sectors, sorting across sectors does not necessarily satisfy PAM. In other words, in every sector firms with

²⁷A firm's production function is separable in quantities and abilities of inputs in a way that makes total factor productivity dependent only on abilities, and it exhibits constant returns to scale in quantities.

better managers employ better workers, but if, say, a group of the best managers sort into the exporting sector it is nevertheless possible for a group of the least-able workers to sort into this sector too. Sufficient conditions for one or the other sorting pattern depend on cross-sectoral comparisons of factor intensities and features of the productivity functions. In this environment elements from the factor proportions theory interact with matching and sorting in determining the impact of globalization on inequality.²⁸

To illustrate the richness of this framework consider an economy in which the best workers and the best managers are employed in one industry and the least-able workers and managers are employed in the other. Also suppose that factor intensities do not differ much between the sectors. How will inequality change when globalization changes product prices? Suppose that it raises the price of the product produced by the industry with the least-able factors of production. Then inequality of earnings across sectors declines, as would be predicted by the factor-specific trade model (see Helpman, 2011, section 3.3).²⁹ There is indeed an element of factor specificity in this framework that emanates from the sorting pattern, because the least-able factors of production are somewhat specific to their sector of employment as are the most-able factors of production. This semi-specificity is responsible for the narrowing of the inequality of earnings between the sectors.³⁰ On the other side there is re-matching of workers and managers in every industry. More workers and managers are attracted to the low-ability sector in response to the price hike. But this means that the range of abilities expands together with employment in the low-ability sector and contracts in the high-ability sector, because the former sector attracts the least-able workers and managers from the latter sector and these employees are more able than the low-ability sector's original employees. As a result, matching improves for one factor of production, either workers or managers, in both sectors, and matching deteriorates in both sectors for the other factor of production (see Proposition 6). Consequently, within-sector compensation inequality rises in both sectors for the input whose matching has improved and declines in both sectors for the input whose

²⁸Grossman, Helpman and Kircher (forthcoming) report in their online appendix that the average compensations of managers and workers were positively correlated across 12 Brazilian manufacturing sectors in 1994. One might conclude that at least in this case an equilibrium with PAM across sectors is more relevant for an analysis of the impact of globalization on inequality than one with the opposite sorting pattern.

²⁹The original sector-specific model was developed by Jones (1971).

³⁰In the sector-specific model the sector-specific inputs cannot change their sector of employment, while here they can. Except that here the ability level of an input affects its sectoral comparative advantage, hence the semi-specificity.

matching has deteriorated. In short, globalization leads to a negative correlation between the within-sector inequality of wages and managerial earnings, and it narrows the earning disparities between sectors.³¹

Sectoral differences in factor intensity open the door to interesting variations. Suppose, as above, that globalization raises the price of goods manufactured by the sector with the least-able employees. Then the semi-specificity of inputs still leads to a reduction of income disparity between sectors, the low-ability sector expands by hiring more workers and managers whose abilities are higher than the abilities of its original employees, and the high-ability sector contracts by shaking off its least-able workers and managers. Now, however, the resulting re-matching of workers and managers within sectors does not have to improve the matches of one party across the board. If indeed matches improve for labor in one sector and deteriorate in the other, then the former sector has to be labor intensive. And similarly for managers.³² In this case factor intensity interacts meaningfully with matching, unveiling the sector in which a factor's matching improves. Improved matching raises the within-sector inequality of a factor's compensation, while deteriorated matching reduces its within-sector inequality. For this reason globalization raises wage inequality in the labor-intensive sector and dampens it in the management-intensive sector, so that within-sector inequality is negatively correlated between workers and managers, between workers in different sectors, and between managers in different sectors (see Proposition 7).

Matching between workers and managers is also analyzed in Antràs, Garicano and Rossi-Hansberg (2006). In their framework there is a managerial hierarchy. Every manager has a team of workers. The ability of a worker determines the range of production problems he can handle. If he encounters a problem that is outside this range, he forwards it to his manager and she solves the problem if it happens to be in the range that she can handle. Solving a problem requires a fixed amount of time on the part of the manager. Every person can handle problems whose upper bound is determined by her ability, and therefore more-able individuals can solve all the problems that less-able individuals can solve, and more. A problem without a solution prevents

³¹Grossman, Helpman and Kircher (forthcoming) report in the online appendix that the change in the inequality indexes of worker wages and management earnings between 1986 and 1994, a time interval that covers a major trade liberalization episode, were weakly negatively correlated across 12 Brazilian manufacturing sectors, in line with this prediction.

³²Naturally, wherever matching improves for labor it deteriorates for managers, and vice versa.

the worker from contributing to output. This structure produces a complementarity between the ability of the worker and the ability of the manager.

Individuals are heterogeneous in ability and ability is continuously distributed. There is a single sector that produces homogeneous output by teams of workers and managers. The economy is competitive, yielding positive assortative matching between workers and managers. That is, better workers are matched with better managers. And since higher-ability individuals have a comparative advantage in management, the highest-ability individuals sort into management and the rest become production workers. The resulting matches generate a wage distribution among production workers and a distribution of earnings among managers.

Starting with two closed economies, say North and South, Antràs, Garicano and Rossi-Hansberg examine the impact of globalization on wage and earnings inequality. In autarky workers match with managers that reside in their own country, whereas in a globalized world, workers in one country can match with managers in another country. As a result, globalization leads to a rematching of workers and managers around the globe, a form of offshoring, which changes the distribution of income.

When the ability distributions are uniform between zero and an upper bound, and one country has a larger upper bound, it is natural to refer to the latter as North and to the former as South. When other things are the same in these countries, globalization can lead all individuals in South to be employed by northern managers. And because these workers have the lowest abilities, they work for the lowest-ability managers in North. Alternatively, depending on parameter values, some southern individuals can be managers in a globalized world. In this event the least-able individuals in South are employed by southern managers, while the most able individuals are employed by the best northern managers. However, in all cases southern workers improve their matches.

Focusing on the absolute difference in wages of the highest- and lowest-ability workers as a measure of inequality, Antràs, Garicano and Rossi-Hansberg show that globalization raises wage inequality within the group of southern workers. It also raises wage inequality within the group of northern workers if management consumes little time and the skill gap between North and South is large. Otherwise globalization reduces wage inequality among northern workers. Moreover, globalization reduces earnings inequality among southern managers but has ambiguous effects on earnings inequality among northern managers.

The role of matching between firms and workers in shaping wage inequality in a Melitz-style economy was analyzed by Sampson (2014). Recall that the original model was developed to explain differences between exporting and nonexporting firms within industries, using variation in productivity across firms as an essential ingredient. But labor was treated as a homogeneous input and therefore all workers were equally paid. By including workers with diverse abilities, Sampson's economy generates a wage distribution and this distribution differs between globalized and nonglobalized economies. Unlike the competitive environments discussed above, here only labor markets are competitive, while firms engage in monopolistic competition in product markets. A firm's total factor productivity depends on its technology and its workers' ability level, and the productivity function exhibits log supermodularity. As a result there is positive assortative matching between firms and workers; better firms—with more-productive technologies—match with more-able workers. The matching pattern shapes wage inequality. Inequality of wages within an ability interval is higher when the workers are matched with more-sophisticated (higher-productivity) firms.

To understand the impact of trade on inequality in this economy, it is necessary to account for the fact that trade changes the distribution of productivity of active firms. While at entry every firm draws its technology from the same pool, independently of the degree of globalization, the productivity level of the least-productive active firms *does* depend on the degree of trade openness. The fraction of exporting firms depends on the fixed cost of exporting and when this cost is low all firms export. Consequently, the productivity distribution of incumbent firms varies with the fixed cost of exporting.

In a world of *symmetric* countries with a low fixed export cost, all firms export and trade leads to an increase in the productivity level of the least-productive firms that remain active in the industry, bringing about a rightward shift in the productivity distribution of incumbents. This shift leads in turn to a rematching of workers and firms that improves the matches of all workers. As a result, wage inequality rises. Moreover, the rightward shift in the technology distribution is larger the larger the fixed cost of exporting, and therefore wage inequality is larger the larger the fixed cost of exporting. Although the case in which all firms export is not the most realistic, it helps isolate a mechanism through which globalization raises wage inequality: the rightward shift in the distribution of active technologies. The conclusion is that "...at sufficiently high levels of trade

integration [that inspire all firms to export] wage inequality is always greater over all workers than in autarky” (Sampson, 2014, p. 176).

When the fixed export cost is not low enough to induce all firms to export, there is an additional element that drives inequality: discontinuity in labor demand by firms around the productivity level that makes exports marginally profitable. This discontinuity arises because firms that do export at this productivity level have to be measurably larger in order to cover the fixed export cost, which leads to a break in the slope of the matching function and to shifts in matching in response to this critical productivity level. Although analytical results are not available for this case, simulations corroborate the intuition that globalization has now more nuanced effects on inequality than in the case in which all firms export. First, trade raises inequality among workers employed by exporters, i.e., in the upper tail of the wage distribution. Second, it raises inequality at the lower end of the wage distribution when trade raises the productivity level of the least-productive incumbents. But trade may also *reduce* the productivity level of the least-productive incumbents, in which case trade integration would reduce inequality among the low-ability workers. Third, inequality rises when variable trade costs decline, but it follows an inverted U-shape as the fixed cost of exporting declines, i.e., it rises initially and declines eventually.³³ This is reminiscent of the result in Helpman, Itskhoki and Redding (2010), and very much for the same reasons to be discussed below.³⁴

Analytical frameworks that embrace rich heterogeneity among workers are needed for understanding recent developments that exhibit ample patterns of change in earnings inequality. To illustrate, in the United States the ratio of the fifth to the first decile of the earnings distribution increased from 2.137 in 2000 to 2.146 in 2007, while the ratio of the ninth to the fifth decile increased from 2.240 to 2.397, pointing to a rise in inequality above and below middle income levels.³⁵ Similar rises of inequality between 2000 and 2007 were reported for Ireland, Japan and Korea. In contrast, in France both ratios declined: the ratio of the fifth to the first decile declined from 1.561 to 1.521, and the ratio of the ninth to the fifth decile declined from 2.112 to 2.093.

³³This is in contrast to the case in which the fixed export cost is low enough to induce all incumbent firms to export, a case in which variable trade costs have no effect on inequality and inequality is higher the larger the fixed cost of exporting.

³⁴Sampson (2014) also discusses an alternative formulation, in which firms choose how much to invest in research and development in order to acquire a desired technology, with higher investment securing a better technology. This is discussed later in this section.

³⁵These and the following data are from OECD.StatExtracts accessed on February 28, 2014.

Moreover, in Canada and the UK the ratio of the fifth to the first decile declined while the ratio of the ninth to the fifth decile increased, and in Germany and Sweden the former ratio increased while the latter declined. Evidence of this type confirms the need for analytical frameworks capable of shedding light on inequality in different segments of the distribution.

Sorting and matching have also been used to study the impact of globalization on growth and inequality, the main idea being that both growth and inequality respond to trade exposure and therefore both are endogenous in the long run. This view, developed in Grossman and Helpman (2016), contrasts with the alternative views that either inequality influences growth or that growth influences inequality (see Helpman, 2004, chap. 6). This is not to deny that during the transition to a long-run growth trajectory current inequality can temporarily affect future growth or that current growth can temporarily affect future inequality; the argument is instead that eventually growth and inequality are jointly determined by fundamental characteristics of the world economy. These characteristics include innovation technologies, specialized resources used by these technologies, the ability to convert research and development experience into usable knowledge, and the like.

In Grossman and Helpman (2016) individuals differ by ability and they can be employed either in manufacturing or in innovation, i.e., they can be idea users or idea producers. In manufacturing, workers are matched with heterogeneous firms that produce varieties of intermediate inputs and these firms engage in monopolistic competition. In the idea-creating sector, workers are matched with heterogeneous research labs that generate ideas for new types of intermediate inputs. An entrepreneur in the manufacturing sector purchases the services of innovators to develop a new type of intermediate input, but the productivity of the technology to manufacture this input becomes known only after the cost of R&D services has been sunk. An entrepreneur in the innovating sector rents a lab for her research projects and the productivity of the lab becomes known only after the cost of the lab has been sunk. In other words, entrepreneurs in both sectors face risky investments. The range of intermediate inputs available for manufacturing evolves over time as new varieties are developed.

The functioning of the manufacturing sector is similar to Sampson (2014), except that there are no fixed costs of either operating or exporting. For that reason every firm that has a technology for a specialized intermediate input remains active, and it exports to other countries when trade is feasible. In the innovation sector the number of new intermediate inputs created by a lab depends

on the productivity of the lab, the ability of the workers employed by the lab, the number of these workers, and—similarly to Romer (1990) and Grossman and Helpman (1991)—the stock of useful knowledge available to the country. The stock of useful knowledge evolves over time according to the country's cumulative innovation experience, and, if the country also engages in foreign trade, according to the cumulative innovation experience of its trade partners. The degree to which a country benefits from R&D investment in other countries may vary across trade partners, and while the pattern of these spillovers can be rich, it is exogenous.

The productivity functions of manufacturing technologies and research labs are log supermodular, exhibiting complementarity between workers' abilities and employers' productivity or sophistication levels. As a result there is positive assortative matching between workers and employers within both the manufacturing sector and the innovation sector. Grossman and Helpman also assume that for every pair of workers the higher-ability worker has a comparative advantage in innovation. Under these circumstances workers with abilities above a threshold sort into innovation and workers with abilities below this threshold sort into manufacturing. This threshold is endogenous, however, and it plays a key role in clarifying the resulting relationship between growth and inequality. A lower threshold implies that more workers are employed in innovation, leading to faster growth, and it provides incentives for workers and employers to rematch. A critical result is that matches improve for all workers, independently of where they are employed, thereby leading to a pervasive rise in wage inequality. These relationships between the threshold, the growth rate, and wage inequality are helpful in understanding the main results that follow.

First, globalization accelerates growth in all countries and brings about convergence of growth rates, in line with the endogenous growth results in Grossman and Helpman (1991). When countries trade they benefit from R&D spillovers from other countries, which raises their stocks of useful knowledge and reduces innovation costs. In consequence, the ability thresholds above which workers sort into innovation decline, leading to faster growth and more wage inequality in every country. That is, globalization raises both growth and inequality. Second, ad-valorem variable trade costs, such as tariffs, do not alter growth rates or inequality within countries, although they impact relative wages across countries. Third, in a globalized world an increase in R&D spillovers between any pair of countries accelerates growth in the entire world economy and raises wage inequality everywhere. Fourth, in a globalized world an increase in any country's R&D subsidy raises its wage inequality relative to the other countries and brings about faster growth

everywhere. These results underline the high degree of interdependence between countries in a globalized world. With R&D spillovers, globalization leads not only to trade interdependence but also to interdependence in growth and inequality.³⁶

A detailed quantitative study of wage determinants in the U.S. economy, which incorporates workers with multiple attributes that sort into multiple occupations, is provided in Burstein, Morales and Vogel (2016). In their analytical framework workers differ by education and gender, and labor productivity varies across occupations as a function of worker type and the type of equipment combined with the worker's labor input. Computer usage provides an explicit channel through which technology impacts wage inequality, because the spread of personal computers was uneven across occupations as well as across workers with different characteristics within occupations; it was biased toward more-educated workers and toward women. A rapidly falling price of computers, a reflection of technological change, encouraged growth in the adoption of computers in the workplace, which affected in turn the composition of labor demand and the structure of wages. On the supply side a rising relative supply of more-educated workers also affected the composition of employment and wages.

Between 1984 and 2003 the log of the skill premium, defined as the average wage of workers with a college degree relative to those without a college degree, increased by 15.1 log points (see Table 3 in Burstein, Morales and Vogel, 2016).³⁷ This skill premium would have declined by 11.4 log points due to changes in labor supply had there been no changes in labor demand. Yet shifts in labor demand increased the skill premium by 26.5 log points. About 60% of this demand effect was due to computer usage, and 18.5% was due to shifts in demand for different occupations. Labor productivity accounts for the remaining sources of demand shifts, where labor productivity is estimated as a residual. The large impact of computers on the skill premium emanated from two sources: the comparative advantage of educated workers in computer use, and the comparative advantage of educated workers in computer-intensive occupations.

³⁶See Helpman (2004, chap. 5) for a review of estimates of international R&D spillovers, which are quite large.

³⁷This is a bit lower than the college wage premium in Autor's (2014) data, where the college wage premium is defined as wages of workers with a college degree relative to workers with a high school degree. In Autor's data the college wage premium was 58% in 1984 and 86% in 2003, which yields $\log(1.86) - \log(1.58) = 0.164$. I thank Jonathan Vogel for clarifying this point.

In order to assess the impact of globalization on the skill premium, it is necessary to assess how trade impacted labor demand via computer usage and occupational employment. Burstein, Morales and Vogel found that if the American economy had had no access to international markets for equipment (including computers), then between 1984 and 2003 the rise in the skill premium would have been lower by 2.1 percentage points, and if the U.S. economy could not have traded occupation services the skill premium would have been lower by 1.3 percentage points (see their Tables 7 and 8). Each of these estimates represents only a fraction, 13% and 27% respectively, of the contribution of computers and occupations to the rise in the skill premium. In other words, international trade had a modest impact on wage inequality.

VIII. Observable Attributes and Technology Choice

The previous section described the impact of globalization on wage inequality in the presence of the matching of heterogeneous workers with heterogeneous firms. The nature of these matches determined the slope of the wage function, and foreign trade impacted these matches through changes in the sorting of workers into different activities and the distribution of technologies in those activities. Nonetheless, firms could not invest in order to improve their technologies. Although firms acquired manufacturing technologies, the cost of a technology was uniquely determined by market conditions. The outcome of the investment could be uncertain, yet there was no way to invest more in order to draw a technology from a more attractive distribution. In this section I discuss wage inequality that arises when workers are heterogeneous and firms can choose among technologies of varying quality.

Yeaple (2005) provided an early analysis. In his model workers vary by ability. There is a traditional sector supplying a homogeneous good in which there is no technology choice and labor productivity rises with ability. In contrast, two technologies are available in the advanced sector that supplies varieties of a differentiated product and labor productivity rises with worker ability for each technology. One of these technologies surpasses the other in terms of productivity, but the better technology requires higher fixed costs of operation. A firm that enters the industry in order to produce a brand of the differentiated product can choose one technology or the other. Finally, higher-ability workers have a comparative advantage in the advanced sector and within this sector they have a comparative advantage operating the better technology. This leads the least-

able workers to sort into the traditional sector, the most-able workers to sort into the advanced sector in order to operate the better technology, and workers with mid-range abilities to sort into the advanced sector in order to operate the inferior technology.³⁸ Under the circumstances wage inequality at the upper end of the distribution is determined by the labor productivity of the better technology, wage inequality at the lower end of the distribution is determined by the labor productivity of the traditional-sector technology, and in the middle range wage inequality is determined by the labor productivity of the inferior technology in the differentiated product sector. Inequality across these groups depends in turn on the comparative advantage of workers across the three technologies.

When two symmetric countries trade with each other and there are fixed and variable trade costs, there is no trade in traditional goods. With low enough fixed export costs there is trade in differentiated products, however, and if both technologies are employed, then all firms that invest in the better technology export. Although there are circumstances in which all firms— independently of their technology—export, the interesting case arises when not all of them do. Whenever only a fraction of the firms exports, those who adopted the poorer technology serve only the domestic market.³⁹ That is, with selection into exporting there is a stark contrast between exporters and nonexporters: the former employ a better technology and hire more-able workers.

A reduction in variable trade costs prompts more firms to adopt the better technology in the differentiated product sector, the most-able workers among those operating the worse technology switch employment to firms operating the better technology, and the least-able workers among those who operate the inferior technology switch employment to the traditional sector. Relative wage changes conform to changes in employment: they rise for high-ability workers operating the advanced technology, decline for middle-ability workers operating the inferior technology, and do not change for workers manufacturing traditional goods.⁴⁰ The result is wage and employment polarization. Since labor market polarization has been documented in the United States and 16 European countries (see Goos, Manning and Salomons, 2009, and Autor, 2014), this

³⁸Depending on the distribution of skills and the available technologies both technologies are used in the advanced sector or only one of them. The discussion in the main text proceeds under the assumption that both technologies are used.

³⁹This outcome requires restrictions on the structure of the fixed costs.

⁴⁰This means that wages rise also for workers who switch from the inferior to the superior technology and decline for workers who switch from the inferior technology to the traditional sector.

analysis identifies a mechanism through which globalization might have contributed to this outcome, although there is no good evidence to support the claim that globalization played a big role in causing polarization.⁴¹

Sampson (2014) extended this analysis, allowing for a continuous distribution of worker abilities and a continuum of technology choices but no traditional sector. In this case there is positive assortative matching between workers and firms when the production function is log supermodular. Since the choice of technology is endogenous, worker heterogeneity ensures heterogeneity of firms. In a trading world there is selection into exporting with firm heterogeneity among exporters and among nonexporters. Exporters are more productive than nonexporters and they employ higher-ability workers. A decline in variable trade costs in a world of symmetric countries induces some lower-productivity firms to export and some of the highest-ability workers employed by nonexporters to switch employment to exporting firms. While inequality of wages does not change among workers employed by exporters before and after the trade shock, nor among workers employed by nonexporters before and after the trade shock, it rises among the workers who switch from nonexporters to exporters and it widens the wage gap between workers employed by exporters and nonexporters.⁴²

Similarly to Yeaple (2005), Bustos (2011b) also assumed that firms in the differentiated product industry can adopt one of two technologies: a high-fixed-cost technology, measured in units of final output, that is efficient at large volumes of output, and a low-fixed-cost technology that is efficient at low volumes of output. Unlike Yeaple, however, she had no homogeneous sector. In addition, she assumed that there are only two types of workers, high-ability skilled workers and low-ability unskilled workers; that both types of workers are essential in production; and that the technologies differ in factor intensities: the large-scale-efficient technology uses more-skilled relative to unskilled workers.

These features were combined with firm heterogeneity. Every entrant to the industry acquires a random firm-specific productivity level that affects proportionately its variable production costs. Higher-productivity firms gain the same variable cost advantage in both technologies. As a result, higher-productivity firms have a comparative advantage in the use of the

⁴¹This type of polarization is sometimes described as the “hollowing out of the middle class.”

⁴²No polarization arises in this case because there is no traditional sector. Without a traditional sector the lowest productivity level of firms in the differentiated product sector does not depend on foreign trade.

large-scale-efficient technology. Naturally, whether a firm chooses to use this technology depends on how productive the firm is and how large the skill premium is. An exceedingly high wage of skilled workers relative to the unskilled can make the variable cost of production with the large-scale-efficient technology so high as to offset the advantage this technology has over the small-scale-efficient technology at high output levels. When the skill premium is not that high, both technologies are used by incumbents and high-productivity firms adopt the large-scale-efficient technology.

Fixed export costs lead to selection of the most-productive firms into exporting. And with suitable parameter restrictions some firms exit the industry after entry, the least-productive incumbents serve only the domestic market and adopt the small-scale-efficient technology, while the remaining incumbents export. Among the exporters lower-productivity firms adopt the small-scale-efficient technology while the higher-productivity firms adopt the large-scale-efficient technology.

Bustos used this framework to study the impact of MERCOSUR—the free trade area established in 1991 between Argentina, Brazil, Paraguay, Uruguay and Venezuela—on Argentinean firms.⁴³ In her model a reduction in variable trade costs leads to exit from the industry of some of the least-productive firms and it induces exporters who use the small-scale-efficient technology to switch to the large-scale-efficient technology. It also leads some of the nonexporters (at the upper scale of the nonexporters' productivity distribution) to start exporting, with the most successful among them adopting the large-scale-efficient technology. Evidently, these within-industry reallocations raise the relative demand for skilled workers, both because a larger fraction of firms adopt the large-scale-efficient technology and because market shares are reallocated from firms that use the low-skill-intensive technology to firms that use the high-skill-intensive technology. As a result, the skill premium rises.

Brazilian tariff cuts were used to assess the response of the share of skilled workers in the employment of Argentinian firms (Brazil was Argentina's largest trade partner). These tariff cuts varied across sectors and provided a source of variation that differentially affected firms in different industries. Among firms above median size 76% exported, while only 38% of firms

⁴³Bustos (2011a) studied the impact of MERCOSUR on the technology upgrading of Argentinean firms. She found that most of the upgrading occurred within firms that switched from nonexporting to exporting status in response to the free trade agreement.

exported among those below median size. The estimates suggest that the average reduction of the tariffs, which amounted to 23%, brought about an 8% reduction in the share of skilled workers in firms below the median and an increase of 6% in the share of skilled workers in firms above the median. These figures represent a significant shift in labor composition across firms. The skill composition was affected both among production and non-production workers, with the strongest effect on the share of skilled workers among nonproduction workers taking place within firms at the top quartile of the size distribution.

The positive correlation between firm size or productivity and the relative use of skilled workers is not unique to Argentina. It was documented for the United States by Bernard, Jensen, Redding and Schott (2007), for Mexico by Verhoogen (2008), and for Chile by Harrigan and Reshef (2015). In combination with selection into exporting this feature generates an increase in the relative demand for skilled workers within sectors in response to trade liberalization. Under the circumstances globalization can raise the relative employment of skilled workers within industries even when it raises the skill premium. For this reason globalization can lead to a pervasive rise in the skill premium, and therefore in wage inequality, in both developed and developing countries. While the Stolper-Samuelson mechanism magnifies the rise in the skill premium in rich countries and moderates it in poor countries, wage inequality may rise in all of them. The question is whether this theoretical possibility generates quantitatively substantial results, however. Burstein and Vogel (2016) provide an answer.

They constructed a quantitative model of international trade with 60 countries and a region that encompasses the remaining parts of the world's economy, a large number of traded sectors, and service sectors that do not engage in foreign trade. Within sectors firms are heterogeneous and higher-productivity firms employ relatively more-skilled workers (worker heterogeneity is limited to skilled and unskilled workers). There is oligopolistic competition with limit pricing, as in Bertrand, Eaton, Jensen and Kortum (2003), that results in variable markups. The model is calibrated to data in 2005–2007. Holding constant endowments and technologies, they use the model to carry out counterfactual experiments. In one experiment trade costs are increased so as to drive all countries to autarky. The differences between autarky and the 2005–2007 outcomes provide estimates of the impact of globalization.

Importantly, in this counterfactual both skilled and unskilled workers gain from trade; that is, moving from autarky to trade raises the real wages of both types of labor everywhere. Moreover,

in all countries, with the exception of Russia, trade raises the wages of skilled workers proportionately more than the wages of unskilled workers. As a result, the skill premium rises in almost all countries (see their Figure 2). This means that the Stolper-Samuelson effects are overwhelmed by the within-industry reallocations, including selection into exporting. Nonetheless, the rise of the skill premium varies across countries, being larger in more open economies and in net exporters of skill-intensive products. The rise in the skill premium is largest in Lithuania, amounting to 12%, compared to an average of 5.1%. It rises by 2% in the United States and by 0.5% in Brazil.

A comparison of 1975–1977 with 2005–2007 counterfactuals also features a pervasive rise in the skill premium, with the Philippines experiencing the largest rise, and this largest increase is a little more than half the rise in Lithuania’s skill premium from autarky to 2005–2007. This quantitative analysis suggests that the contribution to the rise of wage inequality of within-sectoral heterogeneity, selection into exporting, and factor proportions, is significant, yet modest.

IX. Residual Inequality

Residual inequality refers to the component of earnings that is not explained by workers’ observed characteristics. After estimating a wage equation by regressing log wages on worker characteristics such as education, experience and gender, the inequality of log wages can be decomposed into inequality due to observed characteristics and inequality due to unobserved characteristics. The latter component represents residual inequality. While different measures of dispersion can be used, such as the 90/10 percentile wage ratio, the standard deviation of log wages, or the Theil index of the distribution of wages, the decomposition into an observed and an unobserved part is essential for identifying residual inequality.

Katz and Murphy (1992) pointed out the contribution of residual inequality to the rise in the college wage premium in the 1980s in the United States. But the contribution of residual inequality to the rise in wage dispersion was challenged by labor economists, and particularly by Lemieux (2006), who attributed this rise to an “episodic” event and to compositional changes. In response, Autor, Katz and Kearney (2008) conducted a detailed analysis of the data, concluding that the rise in inequality was not episodic, and especially so in the upper-tail of the wage distribution (the 90/50 percentile ratio), where inequality increased nearly consistently from 1980

to 2005. Moreover, while changes in the composition of the labor force exerted upward pressure on residual wage inequality from the late 1980s to 2005, it was concentrated in the lower part of the wage distribution and “...changes in earning dispersion within narrowly defined demographic groups, remain a key force in the evolution of both upper- and lower-tail U.S. residual wage inequality” (p. 301). They found that between 1973 and 2005 the 90/10 percentile ratio of wages of males increased by 34.6 log points, of which 12.8 log points were attributed to residual inequality and about two-thirds of the latter resulted from the rise of the 90/50 percentile ratio.

Rising residual wage inequality played a major role in rising wage dispersion in other countries too. In Sweden, for example, residual wage inequality accounted for 70% of wage dispersion in 2001 (measured as the standard deviation of log wages) and it contributed 87% to the rise in wage inequality between 2001 and 2007 (see Table 3 in Akerman, Helpman, Itskhoki, Muendler and Redding, 2013). Likewise, residual wage inequality accounted for 59% of wage dispersion in 1994 in Brazil, where it also accounted for 49% of the rise in wage inequality between 1986 and 1995 (see Helpman, Itskhoki, Muendler and Redding, 2016). Evidently, the importance of residual wage inequality is not confined to rich countries only.

A number of studies addressed the potential impact of globalization on residual wage inequality, emphasizing alternative mechanisms of wage determination, such as labor market frictions in the form of fair wages (Egger and Kreckemeier, 2009; Amiti and Davis, 2012), efficiency wages (Davis and Harrigan, 2011), and search and matching (Davidson, Matusz and Shevchenko, 2008; Helpman, Itskhoki and Redding, 2010; Helpman, Itskhoki, Muendler and Redding, 2016). Combined with firm heterogeneity, each one of these mechanisms generates a wage distribution among workers with similar characteristics, because these models predict that within the same industry firms with higher productivity pay a higher wage to the same type of workers. The positive correlation between firm size and wages, known as the wage-size premium, has been known for a long time (see Oi and Idson, 1999). By combining international trade with a wage-size premium mechanism, recent studies have shown that due to selection into exporting, exporting firms—which are larger and more productive than nonexporters—pay significantly higher wages. The relationship between exporting and wages then implies that trade impacts wage gaps between workers employed by firms with different characteristics, as a result of which trade impacts residual wage inequality.

To illustrate, first consider the model of fair wages employed by Amiti and Davis (2012). There, wages are assumed to be an increasing function of profits, and since profits consist of a fraction of revenue minus fixed costs, wages are increasing in a firm's revenue. But revenue depends on a firm's business strategy: it can serve only the domestic market or also export, and it can choose to purchase only domestic intermediate inputs or also imported intermediates. Since both exporting and importing entail fixed costs, firms select into exporting and importing, with the most-productive firms engaging in both. In other words, the most-productive firms pay the highest wages, they export their final output, and they import intermediate inputs. On the other side there are firms that do not import intermediate inputs and do not export their output; these are the least-productive firms that serve only the domestic market. As in Melitz (2003), there is a productivity threshold above which firms stay in the industry. Under these circumstances trade weeds out the least-productive firms and shifts market shares toward producers that engage in foreign trade.

Studying a trade liberalization episode in Indonesia during the 1990s, Amiti and Davis (2012) showed that tariff cuts had different effects on wages paid by firms that were internationally oriented than firms that were domestically oriented. In addition, cuts in tariffs on final goods had different effects from cuts in tariffs on intermediate inputs. Since only 5% of the firms imported inputs and exported a fraction of their output, 10% exported output but did not import inputs, and 14% imported inputs but did not export, the data provided rich variation for examining the differential effects of the tariff cuts on wages. In these data exporters paid 27.5% higher wages than domestic firms, importers of intermediate inputs paid 46.8% higher wages, and firms that both imported intermediates and exported final output paid 66.4% higher wages (see their Table 1A). These wage premia declined to 7.6%, 14.6% and 25.4%, respectively, after controlling for firm employment and industry fixed effects. And moreover, exporters were larger and more productive than domestic firms, importers were larger and more productive than exporters, and firms that imported and exported were the largest and most productive, suggesting selection into trade in line with the theory. Also in line with the theory, they found that reductions in output tariffs increased wages paid by exporters and decreased wages paid by nonexporters. In contrast, cuts in tariffs on intermediate inputs increased wages paid by large importers and reduced wages paid by nonimporting firms. Since their data allowed them to estimate these effects only on average wages paid by the Indonesian firms, some of the wage responses might have been related to the rematching of different types of workers with firms, thereby reflecting changes in worker

composition. While they argued that these compositional effects were small, it remains a concern for evaluating the quantitative outcomes. In any case, Amiti and Davis have not provided a summary estimate for the impact of this trade liberalization episode on overall wage inequality.

Helpman, Itskhoki and Redding (2010) developed a theory of trade and wages that builds on search and matching in the labor market. In this theory firms post vacancies and unemployed workers search for jobs. A fraction of the workers are matched with a fraction of the vacancies, and the productivity of a worker in a specific job is a random draw from a known distribution. Firms can screen workers in order to improve the mix of their workforce, but this process is costly and imperfect. As a result, a firm can improve the average productivity of its workers but it cannot assess the productivity of each one of them. Because firms differ in total factor productivity, more-productive firms have a stronger incentive to engage in screening and they therefore end up with higher labor productivity due both to their core productivity and to the better fit of their workers to the jobs in their firms. Selection into exporting provides firms that have large core productivity levels with the strongest incentives to screen, which skews the final labor productivity distribution in comparison to the distribution of core productivity.

This skewness plays an important role in the determination of wage inequality within sectors populated by heterogeneous firms, because following matching and screening every firm bargains with its workers over wages, and this multilateral wage negotiation leads to higher wages being paid by more-productive firms. The more-productive firms are bigger and the most productive among them select into exporting. As a result, exporters pay higher wages than nonexporters, in line with the evidence. An important implication of this theory is that ex-ante identical workers are paid different wages, with those who have been matched and retained by more-productive firms earning more. In addition, trade liberalization has non-monotonic effects on wage inequality within sectors. Starting from autarky, sectoral wage inequality rises initially in response to reductions in trade impediments. In the process, the fraction of exporting firms rises. But inequality peaks when the fraction of exporters is less than one. Further reductions of trade impediments reduce wage inequality until—when all firms export—inequality reaches the level of a closed economy. While real wages are higher than closed-economy wages when all firms export, wage *inequality* is the same in both cases. It follows that when only a fraction of the firms export, wages are always more unequal than they are in autarky.

This model yields simple equations that describe the variation across firms in an industry of wages, employment, and selection into exporting. Helpman, Itskhoki, Muendler and Redding (2016) extended this theoretical model to allow firms to differ also in the fixed cost of exporting and in the ability to screen workers. They then showed how the resulting equations for the cross sectional variations within industries can be used to estimate employment and wage premia for exporters, and how these estimates can in turn be used to characterize wage inequality. However, their estimation methods require rich data sets, because they require data on individual workers and individual firms, and information on which workers are employed by every firm in the sample. Fortunately, data sets of this type are now available for a number of countries, and they illustrated their methodology with Brazilian data.

Brazil liberalized its economy over a number of years, culminating with a major trade liberalization in 1991 with the formation of the MERCOSUR free trade area. Helpman, Itskhoki, Muendler and Redding (2016) used administrative data for manufacturing industries, which covers the formal sector, to estimate the model for every year starting in 1986 and ending in 1998. These data cover about 6 million workers and over 90 thousand firms. From the workers' data they estimated Mincer wage equations, including an occupation-specific component for every firm. The distribution of these firm-specific and occupation-specific components was then taken to represent residual wage inequality. While the variation of wages across firms accounted for 55% of wage inequality in 1994, the variation across firms of the estimated firm-specific components accounted for 39% of wage inequality (where inequality was measured by the standard deviation of log wages). In comparison, observable worker characteristics contributed only 13% to wage inequality. Using the Maximum Likelihood estimates of the model, they calculated that the standard deviation of log wages predicted by the model was 0.46 in comparison to 0.42 in the data, that the predicted Gini coefficient was 0.25 in comparison to 0.22 in the data, and that the predicted 90/10 percentile ratio was 3.24 in comparison to 2.95 in the data. Since only between 5% and 9% of the firms exported, the overall fit for nonexporters was better than for exporters.

Using the 1994 estimates, Helpman, Itskhoki, Muendler and Redding simulated the impact of trade impediments on inequality. First, they found that reductions in trade impediments first raise inequality until it peaks, followed by declines in inequality in response to further declines in trade frictions, as predicted by the theoretical model. Second, they found that declines in variable trade costs lead to a higher peak of inequality than declines in fixed export costs. When

disregarding the multiple destinations of different exporters, the former peak was 10.7% above autarky while the latter was 7.6% above autarky. Allowing for multiple destinations raised the former peak to 23.3% above autarky and the latter to 19.0% (see their Figures 1 and 3). Evidently, multiple export destinations, which feature in the data, amplify the effects of trade on residual wage inequality.

X. Conclusions

A major conclusion from my review of the literature is that the prevalent view that globalization is primarily responsible for the large increase in the inequality of labor compensation has no basis in the evidence. Yes, globalization impacted the wages of different types of workers to different degrees, and yes, it contributed to an increase in the wages of skilled relative to unskilled workers through multiple channels. Yet, in sum, all these effects explain only a fraction of the rise in wage inequality in rich and poor countries alike.

This conclusion does not deny the fact that foreign trade, foreign investment, and offshoring have adversely affected certain workers. Trefler (2004) found, for example, that the U.S.–Canada free trade agreement of 1989 increased temporarily (for about three years) the level of unemployment in Canada; and Autor, Dorn and Hanson (2013) found that imports from China had large differential employment effects across U.S. commuting zones, where employment declined more in zones whose industries were more exposed to import competition from China. But these consequences, as painful as they were for the affected workers, do not imply necessarily that globalization is responsible for the large wage differentials observed in the data.⁴⁴

⁴⁴ Autor, Dorn and Hanson (2013) also report wage effects, but their estimates of the impact of Chinese imports on wage differentials across commuting zones are perplexing. Despite the fact that imports from China consisted predominantly of manufactured products, they found no difference in wage impacts across these zones in the manufacturing sector for college graduates or for non-college graduates (see their Table 7). At the same time, they found somewhat larger declines in the wages of no-college workers than of college graduates in nonmanufacturing sectors in zones more affected by Chinese imports, which is puzzling. In a more recent study, Chetverikov, Larsen and Palmer (2016) found differences across commuting zones in wages of workers in different quantiles of the wage distribution, with more adverse effects on low-wage workers in the zones more affected by import competition from China. Yet the largest differential wage decline at the bottom of the wage distribution amounted to only 2.6% over a decade for zones with the average rise in exposure to imports from China. Using individual-level data, Autor, Dorn and Hanson (2014) also found differential impacts across commuting

While the finding that globalization has had only moderate effects on wage inequality may disappoint some observers, the development of research tools in this area has been nothing but gratifying. To begin with, major methodological advances have been made. While studies of trade and wages focused initially on effects arising through the neoclassical channel, subsequent work introduced a host of additional considerations, such as selection into exporting, variation across firms in productivity and factor intensity, technology upgrading, matching, and residual inequality. In addition, new data sets became available to study the relationship between trade and wages, and these data sets were employed to examine trade liberalization episodes in individual countries as well as to explore quantitatively the interactions between multiple countries. All in all more than twenty years of research have greatly enhanced our understanding of this important issue and they have produced a host of quantitative estimates. And while this cumulative knowledge may not be the last word on the subject, it would be a shame to conduct the debate on the impact of globalization on inequality without utilizing this body of work.

zones on workers with different earnings levels, where higher earners were affected less negatively than lower earners. The point remains, however, that the size of these effects is just not large enough to explain the observed rise in inequality.

References

- Akerman, Anders, Elhanan Helpman, Oleg Itskhoki, Marc-Andreas Muendler & Stephen Redding. (2013), 'Sources of Wage Inequality', *American Economic Review (Papers and Proceedings)*, 103: 214–219.
- Amiti, Mary & Donald R. Davis. (2012), 'Trade, Firms, and Wages: Theory and Evidence', *Review of Economic Studies*, 79: 1–36.
- Anand, Sudhir & Paul Segal. (2015), 'The Global Distribution of Income', Chapter 11 in Atkinson, Anthony B. and François Bourguignon (eds), *Handbook of Income Distribution*, vol. 2A (Amsterdam: North–Holland).
- Antràs, Pol, Luis Garicano & Esteban Rossi-Hansberg. (2006), 'Offshoring in a Knowledge Economy', *Quarterly Journal of Economics*, 121: 31–77.
- Atkinson, Anthony B., Thomas Piketty & Emmanuel Saez. (2011), 'Top Incomes in the Long Run of History', *Journal of Economic Literature*, 49(1): 3–71.
- Autor, David H. (2014), 'Skills, Education, and the Rise of Earnings Inequality Among the “other 99 percent”', *Science*, 344(6186): 843–51.
- Autor, David H., David Dorn & Gordon H. Hanson. (2013), 'The China Syndrome: Local Labor Market Effects of Import Competition in the United States', *American Economic Review*, 103(6): 2121–2168.
- Autor, David H., David Dorn, Gordon H. Hanson & Jae Song. (2014), 'Trade Adjustment: Worker Level Evidence', *Quarterly Journal of Economics*, 129(4): 1799–1860.
- Autor, David H., Lawrence F. Katz & Melissa S. Kearney. (2008), 'Trends in U.S. Wage Inequality: Re-assessing the Revisionists', *Review of Economics and Statistics*, 90: 300–23.
- Autor, David H., Lawrence F. Katz & Alan Krueger. (1998), 'Computing Inequality: Have Computers Changed the Labor Market? *Quarterly Journal of Economics*, 113 (4): 1169–1214.
- Autor, David H., Alan Manning & Christopher L. Smith. (2016), 'The Contribution of the Minimum Wage to US Wage Inequality over Three Decades: A Reassessment', *American Economic Journal: Applied Economics*, 8: 58–99.
- Becker, Gary. (1973), 'A Theory of Marriage I', *Journal of Political Economy*, 81: 813–46.

- Berman, Eli, John Bound & Zvi Griliches. (1994), ‘Changes in the Demand for Skilled Labor in U.S. Manufacturing Industries: Evidence from the Annual Survey of Manufacturing’, *Quarterly Journal of Economics*, 109: 367–98.
- Berman, Eli, John Bound & Stephen Machin. (1998), ‘Implications of Skill-Biased Technological Change: International Evidence’, *Quarterly Journal of Economics*, 113: 1245–79.
- Berman, Eli & Stephen Machin. (2000), ‘Skill-Biased Technology Transfer around the World’, *Oxford Review of Economic Policy*, 16: 12–22.
- Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, & Samuel Kortum. (2003), ‘Plants and Productivity in International Trade’, *American Economic Review*, 93: 1268–90.
- Bernard, Andrew B. & J. Bradford Jensen. (1995), ‘Exporters, Jobs, and Wages in U.S. Manufacturing, 1976–1987’, *Brookings Papers on Economic Activity: Microeconomics*, 67–112.
- Bernard, Andrew B. & J. Bradford Jensen. (1999), ‘Exceptional Exporter Performance: Cause, Effect, or Both?’ *Journal of International Economics*, 47: 1–25.
- Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding & Peter K. Schott. (2007), ‘Firms in International Trade.’ *Journal of Economic Perspectives*, 21: 105–130.
- Bombardini, Matilde, Giovanni Gallipoli & Germán Pupato. (2012), ‘Skill Dispersion and Trade Flows’, *American Economic Review*, 102: 2327–48.
- Borjas, George J., Richard B. Freeman & Lawrence F. Katz. (1997), ‘How Much do Immigration and Trade Affect Labor Market Outcome?’ *Brookings Papers on Economic Activity*, 1: 1–90.
- Bourguignon, François. (2015), *The Globalization of Inequality* (Princeton: Princeton University Press).
- Bourguignon, François & Christian Morrisson. (2002), ‘Inequality Among World Citizens: 1820–1992’, *American Economic Review*, 92: 727–44.
- Burstein, Ariel & Jonathan Vogel. (2016), ‘International Trade, Technology, and the Skill Premium’, *Journal of Political Economy* (forthcoming).
- Burstein, Ariel, Eduardo Morales & Jonathan Vogel. (2016), ‘Changes in Between-Group Inequality: Computers, Occupations, and International Trade’, mimeo, September 13.

- Bustos, Paula. (2011a), 'Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinean Firms', *American Economic Review*, 101: 304–40.
- Bustos, Paula. (2011b), 'The Impact of Trade Liberalization on Skill Upgrading: Evidence from Argentina', Barcelona BGSE working paper: 559.
- Card, David & John E. DiNardo. (2002), 'Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles', *Journal of Labor Economics*, 20: 733–83.
- Chetty, Raj. (2009), 'Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods', *Annual Review of Economics*, 1: 451–488.
- Chetverikov, Denis, Bradley Larsen & Christopher Parmer. (2016), 'IV Quantile Regression for Group-Level Treatments, with an Application to the Distributional Effects of Trade', *Econometrica*, 84 (2): 809–833.
- Costinot, Arnaud & Jonathan Vogel. (2010), 'Matching and Inequality in the World Economy', *Journal of Political Economy*, 118: 747–85.
- Davidson, Carl, Steven J. Matusz & Andrei Shevchenko. (2008), 'Globalization and Firm Level Adjustment with Imperfect Labor Markets', *Journal of International Economics*, 75: 295–309.
- Davis, Donald R. & James Harrigan. (2011), 'Good Jobs, Bad Jobs, and Trade Liberalization', *Journal of International Economics*, 84: 26–36.
- Diamond, Peter A. (1982a), 'Demand Management in Search Equilibrium', *Journal of Political Economy*, 90: 881–94.
- Diamond, Peter A. (1982b), 'Wage Determination and Efficiency in Search Equilibrium', *Review of Economic Studies*, 49: 217–27.
- Eeckhout, Jan & Philipp Kircher. (2016), 'Assortative Matching with Large Firms', mimeo.
- Egger, Hartmut & Udo Kreickemeier. (2009), 'Firm Heterogeneity and the Labor Market Effects of Trade Liberalization', *International Economic Review*, 50: 187–216.
- Estevadeordal, Antoni, Brian Frantz & Alan M. Taylor. (2003), 'The Rise and Fall of World Trade: 1870–1939', *Quarterly Journal of Economics*, 118: 359–407.
- Feenstra, Robert C. (2015), *Advanced International Trade: Theory and Evidence* (Princeton, NJ: Princeton University Press).

- Feenstra, Robert C. & Gordon H. Hanson. (1996), 'Foreign Investment, Outsourcing and Relative Wages', in Feenstra, Robert C., Gene M. Grossman and Douglas A. Irwin (eds.), *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati* (Cambridge, MA: The MIT Press).
- Feenstra, Robert C. & Gordon H. Hanson. (1997), 'Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras.' *Journal of International Economics*, 42: 371–393.
- Feenstra, Robert C. & Gordon H. Hanson. (1999), 'Productivity Measurement and the Impact of Trade and Technology on Wages: Estimates for the U.S., 1972–1990', *Quarterly Journal of Economics*, 114: 907–40.
- Feenstra, Robert C. & Gordon H. Hanson. (2003), 'Global Production Sharing and Rising Inequality: A Survey of Trade and Wages', in Choi, Kwan and James Harrigan (eds), *Handbook of International Trade* (New York: Basil Blackwell).
- Findlay, Ronald & Kevin H. O'Rourke. (2007), *Power and Plenty: Trade, War, and the World Economy in the Second Millennium* (Princeton: Princeton University Press).
- Goldberg, Pinelopi K. & Nina Pavcnik. (2007), 'Distributional Effects of Globalization in Developing Countries', *Journal of Economic Literature*, 45: 39–82.
- Goos, Maarten, Alan Manning & Anna Salomons. (2009), 'Job Polarization in Europe', *American Economic Review*, 99: 58–63.
- Grossman, Gene M. (2013), 'Heterogeneous Workers and International Trade', *Review of World Economics*, 149: 211–45.
- Grossman, Gene M. & Elhanan Helpman. (1991), *Innovation and Growth in the Global Economy* (Cambridge, MA: The MIT Press).
- Grossman, Gene M. & Elhanan Helpman. (2016), 'Growth, Trade, and Inequality', http://scholar.harvard.edu/files/helpman/files/growth_trade_and_inequality_063016.pdf.
- Grossman, Gene M., Elhanan Helpman & Philipp Kircher. (2015), 'Matching, Sorting, and the Distributional Effects of International Trade', *Journal of Political Economy*, forthcoming.
- Grossman, Gene M. & Giovanni Maggi. (2000), 'Diversity and Trade', *American Economic Review* 90 (5): 1255–1275.
- Harrigan, James & Ariel Reshef. (2015), 'Skill-Biased Heterogeneous Firms, Trade Liberalization and the Skill Premium', *Canadian Journal of Economics*, 48: 1024–66.

- Helpman, Elhanan. (2004), *The Mystery of Economic Growth* (Cambridge, MA: Harvard University Press).
- Helpman, Elhanan. (2011), *Understanding Global Trade* (Cambridge, MA: Harvard University Press).
- Helpman, Elhanan, Oleg Itskhoki & Stephen J. Redding. (2010), 'Inequality and Unemployment in a Global Economy,' *Econometrica*, 78: 1239–83.
- Helpman, Elhanan, Oleg Itskhoki & Stephen J. Redding. (2016), 'Trade and Inequality: From Theory to Estimation', *Review of Economic Studies*, forthcoming.
- Jones, Ronald W. (1965), "The Structure of Simple General Equilibrium Models', *Journal of Political Economy*, 73: 557–72.
- Jones, Ronald W. (1971), 'A Three-Factor Model in Theory, Trade and History.' In Jagdish N. Bhagwati, Ronald W. Jones, Robert A. Mundell and Jaroslav Vanek (eds), *Trade, Balance of Payments and Growth: Papers in International Economics in Honor of Charles P. Kindleberger* (Amsterdam: North-Holland).
- Jones, Ronald W. & Jose A. Scheinkman. (1977), 'The Relevance of the Two-Sector Production Model in Trade Theory', *Journal of Political Economy*, 85: 909–935.
- Katz, Lawrence F. & David H. Autor. (1999), 'Changes in the Wage Structure and Earnings inequality', in Orly Ashenfelter and David Card (eds), *Handbook of Labor Economics*, Vol. 3, 1463–555 (Amsterdam: North-Holland).
- Katz, Lawrence F. & Kevin M. Murphy. (1992), 'Changes in Relative Wages, 1963–1987: Supply and Demand Factors', *Quarterly Journal of Economics*, 107: 35–78.
- Krugman, Paul R. (1995), 'Growing World Trade: Causes and Consequences', *Brookings Papers on Economic Activity*, 1: 327–362.
- Krugman, Paul R. (2000), 'Technology, Trade and Factor Prices', *Journal of International Economics*, 50: 51–72.
- Krugman, Paul R. (2008), 'Trade and Wages, Reconsidered', *Brookings Papers on Economic Activity*, 2: 103–38.
- Lawrence, Robert & Matthew J. Slaughter. (1993), 'International Trade and American Wages in the 1980s: Giant Sucking Sound or Small Hiccup?' *Brookings Papers on Economic Activity*, 1: 161–211.

- Leamer, Edward E. (1998), 'In Search of Stolper-Samuelson Linkages between International Trade and Lower Wages', in Susan M. Collins (ed), *Imports, Exports, and the American Worker* (Washington, D.C.: Brookings Institution Press).
- Leamer, Edward E. (2000), 'What's the Use of Factor Content?' *Journal of International Economics*, 50: 17–50.
- McCormick, Michael. (2001), *Origins of the European Economy: Communications and Commerce, AD 300–900* (New York: Cambridge University Press).
- Lemieux, Thomas. (2006), 'Increased Residual Wage Inequality: Composition Effect, Noisy Data, or Rising Demand for Skill', *American Economic Review*, 96: 461–498.
- Machin, Stephen & John van Reenen. (1998), 'Technology and Changes in Skill Structure: Evidence from Seven OECD Countries', *Quarterly Journal of Economics*, 113: 1215–44.
- Maddison, Angus. (2001), *The World Economy: A Millennial Perspective* (Paris: OECD).
- Melitz, Marc J. (2003), 'The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity', *Econometrica*, 71: 1695–1725.
- Melitz, Marc J. & Stephen Redding. (2014), 'Heterogeneous Firms and Trade', in Gopinath, Gita, Elhanan Helpman and Kenneth Rogoff (eds), *Handbook of International Economics* (Amsterdam: North Holland).
- Mincer, Jacob A. (1974), *Schooling, Experience, and Earnings* (New York: The National Bureau of Economic Research).
- Mortensen, Dale T. & Christopher A. Pissarides. (1994), 'Job Creation and Job Destruction in the Theory of Unemployment', *Review of Economic Studies*, 61: 397–415.
- OECD. (2015), *In It Together: Why Less Inequality Benefits All* (Paris: OECD Publishing).
- Ohnsorge, Franziska & Daniel Trefler. (2007), 'Sorting It Out: International Trade with Heterogeneous Workers', *Journal of Political Economy*, 115: 868–92.
- Oi, Walter Y. & T. L. Idson. (1999), 'Firm Size and Wages', in Orly Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, vol. 3 (Amsterdam: Elsevier).
- Rodrik, Dani. (2015), *Economics Rules: The Rights and Wrongs of the Dismal Science* (New York: W. W. Norton & Company).
- Romer, Paul M. (1990), 'Endogenous Technological Change', *Journal of Political Economy*, 98: S71-S102.

- Sachs, Jeffrey D. & Howard Shatz. (1994), 'Trade and Jobs in U.S. Manufacturing', *Brookings Papers on Economic Activity*, 2: 1–84.
- Sampson, Thomas. (2014), 'Selection into Trade and Wage Inequality', *American Economic Journal: Microeconomics*, 6 (3): 157–202.
- Slaughter, Matthew J. (2000), 'What are the Results of Product-Price Studies and What Can We Learn from Their Differences?' in Feenstra, Robert C. (ed), *The Effects of International Trade on Wages* (Chicago: The University of Chicago Press).
- Stolper, Wolfgang W. & Paul A. Samuelson. (1941), 'Protection and Real Wages', *Review of Economic Studies*, IX: 58–73.
- Trefler, Daniel. (2004), 'The Long and Short of the Canada-U.S. Free Trade Agreement', *American Economic Review*, 94: 870-895.
- Van Zanden, Jan Luiten, Joerg Baten, Peter Foldvari & Bas van Leeuwen. (2014), 'The Changing Shape of Global Inequality 1820–200: Exploring a New Dataset', *The Review of Income and Wealth*, 60(2): 279–297.
- Verhoogen, Eric. (2008), 'Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector: Theory and Evidence from an Exchange Rate Shock', *Quarterly Journal of Economics*, 123: 489–530.
- Yeaple, Stephen R. (2005), 'A Simple Model of Firm Heterogeneity, International Trade, and Wages', *Journal of International Economics*, 65: 1–20.
- Zhu, Susan Chun & Daniel Trefler. (2005), 'Trade and Inequality in Developing Countries: A General Equilibrium Analysis', *Journal of International Economics*, 65: 21–48.