Recent Perspectives on Trade and Inequality

Ann Harrison,¹ John McLaren,² and Margaret McMillan³

¹Department of Agricultural and Resource Economics, University of California, Berkeley, California 94720, and The World Bank, Washington, DC 20433
²Department of Economics, University of Virginia, Charlottesville, Virginia 22904-4182; email: jmclaren@virginia.edu
³Department of Economics, Tufts University, Medford, Massachusetts 02155, and IFPRI, Washington, DC 20006-1002

Abstract

The 1990s dealt a blow to traditional Heckscher-Ohlin analysis of the relationship between trade and income inequality, as it became clear that rising inequality in low-income countries and other features of the data were inconsistent with that model. As a result, economists moved away from trade as a plausible explanation for rising income inequality. In recent years, however, a number of new mechanisms have been explored through which trade can affect (and usually increase) income inequality. These include within-industry effects due to heterogeneous firms, the effects of offshoring of tasks, effects on incomplete contracting, and the effects of labor-market frictions. A number of these mechanisms have received substantial empirical support.

Keywords
income inequality, offshoring, globalization, wages
1. INTRODUCTION

One of the most robust trends in the past three decades of the twentieth century has been a rise in within-country inequality in a wide range of countries. This rise in inequality—whether measured in income, wages, wage premia, or assets—has been observed in both the developed and developing worlds. Within the United States, Latin America, Asia, and Africa, the gap between individuals has widened considerably.

One plausible explanation for this increasing inequality is the rise in globalization. Whether measured in trade flows, tariffs, capital flows, or offshoring, globalization has increased markedly in both developed and developing countries. Trade between developed and developing countries has increased substantially, and poles of growth have shifted to the developing world. These parallel developments have naturally led to speculation that the increase in inequality is a result of increased exposure to international trade. Until the 1990s, the leading framework for understanding the possible link between trade and inequality was the Heckscher-Ohlin (HO) model, which, in its simple form, predicts that countries export goods that use intensively the factor with which they are most abundantly supplied. One implication of this framework is that trade increases the real return to the factor that is relatively abundant in each country and lowers the real return to the other factor—known as the Stolper-Samuelson theorem. This means that in developed countries, with an abundance of skilled labor, wages of skilled workers should increase relative to unskilled workers, and inequality should rise with trade. The opposite was expected to happen in developing countries that were well endowed with unskilled labor: Inequality should have declined with trade.

A number of studies published between 1990 and 2010 dealt a serious blow to this theory by documenting an increase in inequality in developing countries that frequently paralleled major trade reforms. Countries exhibiting this trend include Mexico, Colombia, Argentina, Brazil, Chile, India, and China (e.g., see Harrison & Hanson 1999, Goldberg & Pavcnik 2007a,b, Topalova 2007). Although the evidence providing a direct link between trade reforms and rising inequality is available only for some countries (such as India and Mexico), the preponderance of the evidence of rising inequality in developing countries in a period of rapid globalization is nevertheless at odds with the simple predictions of the HO framework. An additional problem for the HO theory has been widespread evidence of within-industry increases in demand for skilled workers (Lawrence & Slaughter 1993). For example, both inequality and the demand for skilled workers have increased in the services sector of the United States in which, prior to the 1990s, there was almost no international trade or offshore activity.

These findings led many economists to drop trade as a candidate for explaining rising inequality and to look for other factors. One leading explanation for trends in inequality is skill-biased technological change, which means changes in technology (such as the increasing use of computers) that increase the demand for skilled workers. Other factors that have been cited by economists include the weakening of labor-market institutions such as unions and the declining real value of minimum wages, differential access to schooling, and immigration. Overall, for a substantial period of time, most labor and trade economists were skeptical of assigning too great an importance to trade-based explanations for the increase in inequality.

That may be changing. The emergence of stylized facts at odds with existing trade theory has led to new theoretical developments focusing on heterogeneous firms and
bargaining, trade in tasks, labor-market frictions, and incomplete contracts. These new theories provide insights into the effects of trade on income and wage inequality. This more recent literature, which has emerged in the past decade, is the focus of our review. We demonstrate that there are now a number of ways to explain how trade could contribute to rising within-industry inequality as well as rising inequality in countries at all income levels. However, the empirical literature has not kept pace with the theoretical developments, in part because they are so new. Researchers will need to sort through these different theories to identify which are most consistent with the data.

For the purpose of this review, trade is broadly defined to include trade in goods and services and foreign direct investment. Much has been written about how to define inequality, and we do not have the space to go into those details here. For our purposes, it is sufficient to note two important facts. First, income-based measures of inequality are subject to all the same caveats as income-based measures of poverty (see Deaton 2005 for a review of these issues). Second, this review focuses only on inequality within countries as measured by income and wages; it does not focus on inequality across countries. For discussions of trends in inequality across countries—global inequality—the interested reader is referred to Ravallion (2001, 2003), Milanovic (2005), and Sala-i-Martin (2002).

The rest of this review is organized as follows. Section 2 reviews the theoretical literature on trade and inequality beginning with the older literature but emphasizing the new developments that are more consistent with recent empirical evidence. Section 3 reviews the empirical literature on trade and inequality, again beginning with the older literature but emphasizing recent work using new data sets and innovative approaches. Section 4 comments on directions for future work.

2. THEORY

2.1. Brief Synopsis of Earlier Work

First, we present a whirlwind synopsis of the theory of trade and income inequality before 2003 (after which work in the area seems to have accelerated because of interest in heterogeneous firms). Following that, we launch into more recent work. The mainspring of theory behind empirical work on trade and distribution in the 1990s was the classic comparative-advantage framework. In particular, the distinction between distributive effects in an HO model and in specific-factors models was a key focus.

In an HO model, each factor of production is able to move costlessly between industries (but not across countries). As a result, each factor earns the same income regardless of the industry that employs it, and trade affects income inequality by changing the prices of factors. In a two-factor version of the model, this means that trade increases the real return to the factor that is relatively abundant in each country and lowers the real return to the other factor—the Stolper-Samuelson theorem. If the two factors are skilled and unskilled labor, that means that trade increases income inequality in rich countries (by raising the real return to abundant skilled labor and lowering the return to scarce unskilled labor) and lowers income inequality in poor countries. In many-factor models, trade on average raises the prices of factors that are more abundant in each country relative to less abundant factors (see Deardorff 1982 for a general treatment).

By contrast, in a specific-factors model, one or more factors of production cannot change industries at all. As a result, trade tends to lower the real incomes of factors in
import-competing industries and raise those in export industries (Jones 1971 is a classic reference). For example, if human capital is industry specific, trade will raise the incomes of workers in exporting industries at the expense of workers in import-competing industries. As a result, trade will increase income inequality if export-sector workers tend to have higher wages and reduce it otherwise.

Applied economists over the years have noted problems with both of these simple approaches, particularly the HO framework, as a guide to the income-inequality effects of trade. For example, Harrison & McMillan (2007) collect a number of the more important issues, including the likelihood that different countries produce different goods (which invalidates the Stolper-Samuelson theorem) and the presence of labor-market frictions. Accordingly, a number of important qualifications have been added to this basic framework.

2.1.1. Trade in tasks. Feenstra & Hanson (1996) study a model of offshoring, or the practice by which a firm producing in one country allocates some tasks to workers in another country. In their model, a single good is produced by a competitive industry, with each firm hiring skilled and unskilled workers to perform a continuum of tasks. The tasks can be ranked on the basis of their skill intensity, and a complete set of tasks must be combined with capital to produce output. There are two countries, with different relative supplies of skilled and unskilled workers. Because skilled workers are relatively inexpensive in the skill-rich country, cost minimization calls for each firm to choose a cutoff task, allocating tasks that are more skill intensive than the cutoff to workers in the skill-rich country and tasks that are less skill intensive than the cutoff to workers in the skill-poor country. In other words, if we think of the firms as headquartered in the skill-rich country, then they offshore less skill-intensive tasks to the skill-poor country (or, equivalently, the skill-rich country imports unskilled-intensive tasks from the skill-poor country). Now if the environment changes so that it becomes easier to offshore from the skill-rich to the skill-poor country (modeled by Feenstra & Hanson as a movement of capital from the skill-rich to skill-poor country), the initial cutoff task is replaced by a new cutoff task that is more unskilled intensive. Thus a range of tasks is moved from the skill-rich country to the skill-poor country. As the tasks thus reallocated are the least skill intensive that were being performed in the skill-rich country, but are more skill intensive than the tasks initially done in the skill-poor country, the result is that labor demand becomes more skill intensive in both countries at the same time. Consequently, the equilibrium skill premium rises in both countries. Recall that the simple HO model predicted that trade in goods would raise income inequality in rich countries but lower it in poor ones. By contrast, the Feenstra-Hanson offshoring model predicts that trade in tasks will raise income inequality in both countries. This is a striking result, not least because of abundant empirical work suggesting a rise in income inequality accompanying trade liberalization in countries across the income spectrum in the 1980s.

1In the popular press, this is often called outsourcing, but we follow the usage of the research literature in calling it offshoring to distinguish it from outsourcing in industrial organization.

2Zhu & Trefler (2005) show that the insight provided by Feenstra & Hanson can apply in a model with only goods trade. If the North has a comparative advantage in skill-intensive goods, and technological progress allows the South to grow relative to the North, then a range of the North’s least skill-intensive goods will shift to the South, where they will become the most skill-intensive goods, raising skill premia in both regions. Matsuyama (2007) shows that similar effects can be obtained in a model in which transport costs are modeled as a separate sector, which uses skilled and unskilled labor and is skilled-labor intensive relative to goods production.
2.1.2. Search frictions and unemployment. Davidson et al. (1999) incorporate worker search frictions and unemployment into a standard trade model, showing that such frictions can lead to a substantive revision of the distributional effects of trade. For example, the Stolper-Samuelson theorem does not extend to an environment of that sort when formulated as a statement about the incomes of employed factors, but it does extend to such an environment when formulated as a statement about the expected lifetime income of searching factors. A wide range of effects of search frictions on labor-market outcomes in trade models is gathered in Davidson & Matusz (2009).

2.1.3. Trade and innovation. A small cluster of theoretical work shows that innovation can be an important channel through which trade affects income distribution, in ways that are very different from a comparative-advantage approach. For example, Dinopoulos & Segerstrom (1999) study a two-country growth model with a continuum of industries, in which firms compete through research and development (R&D) for technological supremacy. In each industry, the firm with the best technology captures the entire market, but its price is limited by the marginal cost of the next-best-available technology. At any moment, a large number of firms conduct R&D to obtain a breakthrough and become the new leader. Each country exports the products for which the industry leader happens to be, at the moment, one of its domestic firms, and each country initially maintains a uniform tariff against anything its consumers might import from the other country. The tariffs cut into the market leader’s profit margins, reducing the jackpot that results from being the market leader, and thus reducing the incentive for any firm to do R&D to become the market leader. As a result, trade liberalization increases R&D, and growth, in both countries. Now, with regard to income inequality, if R&D is skilled-labor intensive relative to manufacturing, given its reliance on scientists and engineers, then trade liberalization will raise the relative demand for skilled labor in both countries, increasing income inequality all around. A related approach is provided by Neary (2003, section 9), in which identical countries with a large number of Cournot oligopolies open to trade. Each oligopolist now has an incentive to do R&D to lower its marginal cost and obtain an advantage over its foreign competitor, resulting, once again, in a rise in skill-intensive R&D spending and a rise in income inequality. (Intriguing evidence for this mechanism for Brazil is presented in Nelson 2008.) A related argument is developed by Thoenig & Verdier (2003) in a model of leapfrogging R&D. Note that these R&D-based theories are fully consistent with North-North trade and with complement approaches such as HO and Feenstra-Hanson that are based on North-South trade.

Now we turn to more recent developments in the theory.

2.2. Heterogeneous Firms and Bargaining

An important element was introduced to trade theory by Melitz (2003), who incorporated heterogeneous-firms monopolistic competition, following an approach pioneered by Hopenhayn (1992), into a model of international trade. The approach has had considerable influence on a wide scope of trade topics, and income inequality effects are no exception.

To explore the effect of heterogeneous firms on trade and inequality, we first review the features of the basic model. That model can be summarized as follows. Consumers have constant-elasticity-of-substitution preferences over a continuum of potential products.
Anyone can choose to become an entrepreneur by incurring a fixed cost $f_e$, which can be interpreted as the cost of developing a new product. Once this has been done, the entrepreneur can produce the output, with a production function given by

$$ q = (l - f) \phi, $$

where $l$ is the labor employed per period, $q$ is the output produced per period, $f$ is a fixed labor requirement per period, and $\phi$ is the marginal product of labor. The fixed cost $f$ is a constant of known value across firms and time, but the productivity parameter $\phi$ is a random variable, constant across time for any one firm but taking different values from one firm to another. Importantly, $\phi$ is something that the entrepreneur can learn only after incurring the fixed cost $f_e$. As a result, a certain fraction of entrepreneurs exit the market as soon as they have entered because their realization of $\phi$ is too low for them to be able to break even given the fixed production cost $f$.

In autarky, equilibrium is determined by two values: the number of firms entering and paying $f_e$ and a cutoff productivity $\phi^*_s$ for staying. These two variables need to take values such that two conditions hold. First, the zero cutoff profit condition requires that variable profits for a firm with a productivity parameter exactly equal to $\phi^*_s$ are equal to $f$, so that any firm with a realization $\phi < \phi^*_s$ will exit, and any firm with a realization $\phi > \phi^*_s$ will stay in the industry and make positive profits. Second, the free-entry condition requires that expected profits net of $f$ for any entrant who has not yet learned her value of $\phi$ are equal to $f_e$, taking into account the possibility that the firm will choose to exit right away. This ensures that entrepreneurs’ ex ante profits are equal to zero.

In the open-economy version of the model, there are $n + 1$ identical countries, and any firm can export to any of them by paying an additional fixed cost $f_{ex}$. In addition, there is an iceberg transport cost, meaning that a fraction of any shipment is lost in transit. Owing to the fixed cost of exporting, it is not worthwhile to export a small amount of any product, so only highly productive firms export at all. Therefore, equilibrium is characterized not only by a number of entrepreneurs entering and a cutoff productivity level for staying, but also by a cutoff productivity level for exporting. Denote the latter two by $\phi^*$ and $\phi^*_e$, so that a firm with $\phi < \phi^*$ will exit without producing anything; a firm with $\phi^* < \phi < \phi^*_e$ will stay and produce but not export; and a firm with $\phi > \phi^*_e$ will stay and export. A crucial finding of the model is that $\phi^*_s < \phi^*$, so firms that survive under trade are more productive than the firms that survive under autarky. A way of understanding the mechanism behind this is as follows. Suppose for the moment that the cutoff for firm exit and the number of firms entering do not change when trade is opened. Now each entrepreneur contemplating paying $f_e$ to create a product knows that in addition to the prospects available under autarky, there is the new possibility that if $\phi$ turns out to be high enough, the entrepreneur will also be able to earn more profits by exporting. Because of this, the expected profit will now rise and will be greater than $f_e$. Therefore, prospective entrepreneurs will see a strictly positive expected profit from creating a new product, and the free-entry condition will be violated. If the cutoff for remaining does not change, this requires an increase in the number of entrepreneurs entering. But then there will be more competition; each firm’s share of domestic demand will fall; and the variable profit of any firm that does not export will fall. Therefore, some marginal firms whose variable profits were close to the fixed production cost $f$ will drop out; in other words, $\phi^* < \phi^*_e$.

Therefore, free trade raises productivity. Now nothing in this argument has anything to do per se with income inequality. The labor market is frictionless and all workers are
identical, so all workers receive the same wage. The only possibility for income inequality is in profits because different firms earn different levels of profits ex post, but in a model with only risk-neutral individuals and no modeled financial market, the same equilibrium would be obtained if either (a) firms are self-financed by entrepreneurs out of wage earnings, so that each entrepreneur keeps the profit from his own project, some getting rich and others losing their investment completely, or (b) start-up firms are financed by the sale of equity, with each citizen buying shares of each start-up, receiving exactly the same share of ex post profits, and receiving a zero rate of return on the whole portfolio. The model is not set up with a focus on income distribution and so does not provide a theory of income distribution. We now turn to two prominent examples of models that do focus on income distribution, taking Melitz as a point of departure.

Egger & Kreickemeier (2009) explore a Melitz-type model with a significant form of market friction: Workers care about receiving fair wages. The underlying theory is adopted from Akerlof & Yellen (1990), who argue that workers’ motivation to provide effort depends on the perceived fairness of the wages they are paid, apart from any direct incentives regarding performance and shirking. This is one version of an efficiency-wage argument, and similar to others of the genre, it features equilibrium unemployment in general because even in conditions of excess labor supply an employer has an incentive to not lower the wage, for fear of reducing her workers’ effort level. In addition, the sense of fairness employed here includes an assumption that workers who work at more productive and profitable firms feel entitled to a higher income as a result, so this model also implies that wages will differ from firm to firm. Thus this model generates wage inequality, and this inequality is affected by trade.

The particular formulation of fairness used here makes use of a reference wage, a hypothetical wage against which a worker compares the wage he actually receives in evaluating its fairness. For the purposes of the Egger & Kreickemeier model, the reference wage is defined, for any given worker, as

\[ \hat{w}(\phi) = \phi^\theta \{ (1 - U) \bar{w} \}^{1 - \theta}, \]

where \( \hat{w} \) denotes the reference wage; \( \phi \) is the productivity parameter for the firm in which the employer works, modeled exactly as in the Melitz model above; \( U \) is the aggregate unemployment rate; \( \bar{w} \) is the average wage among employed workers; and \( \theta \) is a parameter, common to all workers, indicating how important a worker’s own firm’s productivity is to workers’ evaluation of the fairness of their own wages. The term \( \{ (1 - U) \bar{w} \} \) is the average income of a worker in the economy, taking into account that a fraction \( U \) are unemployed and therefore have a zero wage. A high value of \( \theta \) indicates that workers in productive firms themselves feel entitled to high wages regardless of aggregate conditions, and this tends to lead to a high variance in wages across firms. Workers paid their reference wage or more put in full effort, whereas workers paid less than that reduce their effort in proportion to the shortfall in wages. Consequently, employers never have an incentive to pay a wage different from their workers’ reference wage, and the reference wage acts as if it is a binding minimum wage—except that it varies from firm to firm, and it responds endogenously to a change in the environment as \( U \) and \( w \) change.

This construction is added to the Melitz model together with an assumption that the distribution of the \( \phi \)'s is Pareto so that the probability that \( \phi \) is greater than \( \phi' \) is equal to \( (\phi')^{-k} \), where \( k > 0 \) is an exogenous shape parameter. Parallel to the basic Melitz model, an autarky equilibrium consists of two variables, a productivity cutoff \( \phi^*_a \) and the number of
firms entering, such that (a) entrepreneurs paying \( f_e \) to enter receive zero profits in expectation and (b) entrepreneurs who have entered stay if and only if their draw of \( \phi \) is at least as high as \( \phi^*_e \). The equilibrium features wages that differ from firm to firm according to Equation 2 and also, in general, positive unemployment. Both these features emerge more strongly if \( \theta \) is high. As noted above, a high value for \( \theta \) implies heterogeneous wages as workers in more productive firms will insist on higher wages than workers in more marginal firms. To see why a high value for \( \theta \) also contributes to unemployment, we note that if \( \theta = 0 \), wage heterogeneity disappears, the reference wage becomes the average wage, and the reference wage constraint given in Equation 2 collapses to a vacuous statement that each firm offers the representative wage. This allows the wage to fall until the labor market clears.

Opening the model to trade, we again have a cutoff for staying, \( \phi^*_x > \phi^*_a \), and a cutoff for exporting, \( \phi^*_x > \phi^* \). Once again, marginal firms are squeezed out by the increasing competition, and average productivity rises. However, now two new effects occur. First, unemployment increases. This is the net result of two forces working in opposite directions on the demand for labor: A rise in overall output, which increases demand for labor, and a rise in productivity, which decreases it. Second, the average real wage of employed workers rises. Third, wage inequality, as measured by the ratio of the average wage for employed workers to the lowest wage for employed workers, \( \bar{w}/\hat{w}(\phi^*) \), rises.

This last result is the key one, and it requires explanation. After all, wage inequality results from heterogeneity in firms, and the selection effect of trade (\( \phi^*_x > \phi^*_a \)) that eliminates lower-productivity firms seems as if it should reduce that heterogeneity. Two points can help us understand what drives this result. First, mere truncation of a distribution does not necessarily reduce the inequality in it.\(^3\) In the Pareto case, truncation of the left-hand tail of the distribution merely scales up the distribution, multiplying every moment by a common factor, and leaving every measure of inequality unchanged (this point is discussed at length in Helpman et al. 2010). Consequently, the elimination of less productive firms does nothing to reduce inequality in the distribution of \( \phi \)'s. Furthermore, note that the increased profitability of high-productivity firms does not directly affect wage inequality either because, by Equation 2, the ratio between the wages paid at two firms is a function of the ratio between the \( \phi \)'s at the two firms, not their realized profits. On the other hand, when trade is opened, the more productive firms hire additional workers to serve foreign markets, while the less productive surviving firms shed workers, battered by competition from imports. Consequently, the average wage among the employed, \( \bar{w} \), now is more heavily weighted to high-wage, high-productivity firms than it was previously. This is what guarantees that the ratio of the average employed worker’s wage to the lowest employed worker’s wage rises with trade.

In other words, the way in which wage inequality is affected in this model can be described as follows: The ratio of the 90th-percentile firm’s wage to the 10th-percentile firm’s wage is unchanged by trade, but the ratio of the 90th-percentile employed worker’s wage to the 10th-percentile employed worker’s wage goes up, provided that the 90th-percentile worker is employed in a firm that exports and the 10th-percentile worker is not. The employment share of the high-wage firms has increased relative to the

\(^3\)In a quick example, consider a random variable that takes a value of 1 with probability \( X \), and a value of \( e \) and \( e^3 \) with probability \( (1 - X) / 2 \) each. If \( X \) is close enough to 1, the log variance of this distribution is very close to zero, but truncating the distribution by eliminating the left-hand tail, in other words, eliminating the value 1, results in a log variance equal to 1.
employment share of the low-wage firms. When we include unemployed workers in the discussion, the finding of increased inequality due to trade is strengthened: The fraction of the workforce who earn zero wages goes up, even as the average income per worker rises.

A related approach is explored by Davis & Harrigan (2007), who adapt a more conventional efficiency-wage theory to the Melitz model. They use the monitoring approach of Shapiro & Stiglitz (1984), in which employees can shirk on the job and need to be deterred from doing so by a threat of firing in the event that they are caught. In the original model, every firm was identical and, in particular, possessed the same exogenous probability of catching a shirker in any period. In equilibrium, each firm charges the minimum wage required to deter shirking given the detection probability, and in the aggregate a positive fraction of workers must be unemployed (or else it would be impossible to deter shirking at all, as a shirking worker would just get a new job with another firm right away). In the Davis & Harrigan approach, however, firms, indexed by $i$, differ from each other in the marginal product of labor $\phi_i$, just as in Melitz, but they also differ in the probability $m_i$ of detecting a shirking worker in any one period. Because the minimum wage required to deter shirking depends on $m_i$, this implies that the wage paid will vary from firm to firm, with firms that are good at catching shirkers (high $m_i$) paying low wages and firms that are bad at catching shirkers paying high wages. As a result, good jobs are those with firms that have low detection probabilities. Because a firm's marginal production cost is equal to the wage it must pay divided by $f_i$, firms with low $m_i$ and $f_i$ are the ones that will exit when trade is opened, but as these tend to be high-wage jobs because of the low $m_i$'s, this means that free trade tends to eliminate the good jobs along with the high-marginal-cost firms. As a result, trade actually reduces wage inequality. Note that unless $m_i$ and $\phi_i$ are strongly negatively correlated, the firms with the high wages tend to be those with high marginal costs, which are therefore the smaller ones, and the ones that do not export—the exact opposite of what is predicted by Egger & Kreickemeier (2009), and a prediction at odds with the data. However, allowing for a sufficiently strong negative correlation reverses these correlations, as the authors show in simulations. In that case, trade once again increases wage inequality, disproportionately killing off the bad jobs.

A third heterogeneous-firms approach to trade and wage inequality is found in Helpman et al. (2010). They add a number of additional elements: search frictions, bargaining between workers and employers, idiosyncratic match quality, and employer testing to identify which workers will be the most productive. Workers search for employers and find an employer with a probability that depends on the ratio of vacancies to workers searching (this is a one-period model, so a worker who does not find a job on the first try simply has zero income). Any worker has an idiosyncratic match quality with any given employer; higher-quality matches result in more productivity on the job, and a low-quality match can actually reduce the firm's overall output, so each firm has an interest in hiring only workers who will be good matches. Consequently, when a worker finds an employer who is hiring, the employer subjects the worker to a test that reveals whether the match quality is above or below a given threshold chosen by the firm. Workers who are revealed to be above the threshold are hired and then bargain with the employer for the wage. Workers below the threshold remain unemployed and receive zero income. In equilibrium, more productive firms screen more assiduously than less productive firms, in the sense that they set their threshold for match quality higher. This is because it is costly to set a higher threshold (this is assumed; the technology of test taking that would lead to this property is not modeled), and it is worthwhile only for the highest-productivity firm, with
its high anticipated volume of sales, to incur the high cost of a stringent test. Consequently, a worker who passes the test at a high-productivity firm is revealed to be highly productive at that firm, and the combination of the firm’s productivity with the worker’s high revealed match quality implies that the bargaining surplus between worker and firm is large—so the worker and firm will agree to a high wage. The result is that workers at large, high-productivity firms receive higher wages than workers at small, low-productivity firms. (However, workers are indifferent about applying for work at high- and low-productivity firms. A high-productivity firm pays high wages to the workers it hires, but it does not hire many of the workers who apply. These effects cancel each other out.)

In addition, trade intensifies these effects. It increases the incentive to screen assiduously at high-productivity firms because of the extra volume of sales that will come from exports. It decreases the incentive to screen at marginal surviving firms, which reduce their output and do not export. As a result, trade unambiguously increases wage inequality (and in a much stronger sense than in the Egger and Kreickemeier model, as it actually produces a new wage distribution that dominates the autarky one by second-order stochastic dominance). In addition, trade increases unemployment by increasing the market share of large firms and then making those large firms more picky about hiring. Thus (as in Egger & Kreickemeier 2009) trade further increases income inequality by increasing the fraction of workers receiving zero income.

2.3. New Approaches to Comparative Advantage and Inequality

The heterogeneous-firms literature has provided a number of channels in which trade can affect income inequality even between identical countries. Beyond that, a number of recent papers have re-examined and extended the comparative-advantage approach in ways that allow for a more nuanced view of trade and income inequality than was available before.

2.3.1. Trade in tasks, revisited. Grossman & Rossi-Hansberg (2008) explore the implications of trade in tasks, earlier examined by Feenstra & Hanson (1996). Their emphasis is the possibility of productivity benefits from offshoring, which in principle can make offshoring a Pareto-improving phenomenon. In the simplest version of the model, there are two goods, X and Y, both of which can be produced in a country, denoted Home, by completing a given set of tasks. Some of the tasks need to be performed by high-skilled labor, but others can be performed by unskilled labor. Consider first the production technology if only domestic labor is used. For good j, there is a continuum of measure 1 of tasks of each type that must be completed to produce one unit of output, and for each high-skill task \(a_{Hj}\), units of high-skill labor are required, while for each low-skill task \(a_{Lj}\), units of low-skilled labor are required. Thus a unit of good j requires \(a_{Hj}\) units of high-skilled labor and \(a_{Lj}\) units of low-skilled labor to produce. Assume that \(a_{HX} / a_{LX} > a_{HY} / a_{LY}\), so that good X is skill intensive. If we let Home be a small economy so that the prices of the two goods are set on world markets, and let X be the numéraire, then this determines the income to both kinds of workers as the solution to the two zero-profit conditions:

\[
\begin{align*}
\frac{a_{HX}}{a_{HY}} w_H + \frac{a_{LX}}{a_{LY}} w_L &= 1, \\
\frac{a_{HX}}{a_{HY}} w_H + \frac{a_{LY}}{a_{LY}} w_L &= P,
\end{align*}
\]  

(3)

where \(w_H\) is the wage paid to high-skilled labor, \(w_L\) is the wage paid to low-skilled labor, and \(P\) is the price of good Y. Because X is high-skilled-labor intensive, this pair of linear
equations has a unique solution for the wages $w_H$ and $w_L$. Now allow for producers in Home to import some low-skill tasks from workers in another country, denoted Foreign. Suppose that performing task $i$ in Foreign for good $j$ requires $a_{i,j} \beta(i)$ units of labor, where $\beta(i) > 1$ to reflect the logistical and monitoring problems of performing a task abroad. These problems can be weighed against the cost benefit of employing lower-cost labor, as the low-skill wage in Foreign, $w^*$, is lower than the low-skill wage in Home, $w_L$. A Home firm will offshore a task $i$ to Foreign if $w^* \beta(i) < w_L$ and will source the task domestically otherwise. Without loss of generality, the function $t(\cdot)$ is increasing, so tasks with a higher index are harder to offshore. In that case, there will be a cutoff task, say, $I$, such that all producers of either good in Home will offshore low-skill tasks $i \in [0, \ldots, I]$ and source all tasks $i \in (I, \ldots, 1]$ in Home. As a result, for given factor prices, the low-skill labor costs for a producer in either industry are reduced by a common proportion, say, to a fraction $\Omega(I)$ of their original value [it is mechanical to compute $\Omega(I)$ by integrating the cost savings over $i$, but the details do not concern us here]. This changes Equation 3 to

$$
\begin{align*}
    a_{i,X} w_H + a_{L,X} \Omega(I) w_L &= 1, \\
    a_{i,Y} w_H + a_{L,Y} \Omega(I) w_L &= P. 
\end{align*}
$$

It is immediately clear that setting $w_H$ and $\Omega(I) w_L$ to the values held by the values $w_H$ and $w_L$ in the solution to Equation 3 will now solve Equation 4. As a result, offshoring has now increased the wages of low-skilled workers in Home, by a factor of $1/\Omega(I)$, without changing wages for high-skilled workers in Home—a Pareto improvement. This is, of course, the opposite of what many commentators on globalization would expect, particularly as it is only low-skilled workers whose jobs are being shipped overseas. The point is that low-skilled workers in Home are benefitting from what is in effect an improvement in their productivity. It is as if each blue-collar worker in Home previously had to construct her own chair to sit on to work, but now globalization allows her to hire a low-wage worker overseas to build the chair, allowing the Home blue-collar worker to concentrate on other tasks, get more work done, and earn a higher income as a result.

A few qualifications to this result are in order. First, the finding that offshoring can be Pareto improving through productivity effects is not, strictly speaking, new. It shows up as a special case of the Feenstra & Hanson (1996, p. 101) model, for example, but the mechanism in the Grossman-Rossi-Hansberg model brings it into exceptionally sharp focus. Second, Grossman & Rossi-Hansberg point out that it is mitigated and can be overturned by terms-of-trade effects, if the small-country assumption is relaxed. In particular, when offshoring becomes possible (or when it becomes more cost-effective because of a drop in the parameter $\beta$), the equilibrium is changed in a way that is similar to the effect of increasing the supply of low-skill labor in Home. This increases Home output of the low-skill-intensive good $Y$ relative to the high-skill-intensive good $X$, which in the event that Home is a large country will tend to push the relative price of $Y$ up and $P$, down. This shifts the zero-profit conditions (Equation 4) in a way that pushes $w_H$ up and $w_L$ down, following conventional Stolper-Samuelson logic. If this effect is strong enough, low-skilled workers in Home are hurt by offshoring. Finally, if the model is modified to allow for the possibility of more factors than goods—if, for example, in equilibrium Home produces only good $X$—then this same feature of offshoring, that it acts like an increase in the supply of unskilled labor, will push $w_L$ down even if Home is a small open economy so that there is no terms-of-trade effect. Whether the productivity effect or these labor supply effects dominates is an empirical question.
2.3.2. Continuum of skills. Some recent work has aimed at a richer and more realistic account of income inequality by looking at trade models with a continuum of skill levels and hence a continuum of income levels. Blanchard & Willmann (2008) formulate a model with a continuum of goods indexed by $j \in [0,1]$ and a labor force with a continuum of ability levels, $a \in [0,1]$, exogenously given as realizations of a random variable. To produce product $j$, a worker needs to complete the appropriate education, which costs the worker $c(j,a)$. This is increasing in $j$, so that industries are ordered in increasing order of skill requirement, and decreasing in $a$, so that the cost of acquiring any sort of education is smaller for a person endowed with high ability. Furthermore, $\frac{\partial c(j,a)}{\partial a} < 0$, so the marginal cost of choosing a more difficult industry is lower for a person of higher ability. Once a worker has acquired the skill required to produce $j$, he can produce one unit of it. In equilibrium, goods prices $j \in [0,1]$ induce each worker of ability $a$ to choose an industry $j$ such that the quantity of each good produced is equal to the quantity demanded. The price function must be increasing in $j$ to provide an incentive for workers to acquire the skills required to produce some of each good. The exact shape of the price function is determined as the solution to a differential equation.

This structure allows the authors to look at questions of income distribution that would be unthinkable in a model limited to high-skilled and low-skilled workers only. For example, the authors are interested in the effects of trade on the middle class. They examine one numerical example in which Home has an educational cost function given by

$$c(j,a) = \frac{(1-a)j^2}{a},$$

(5)

and Foreign has an educational cost function given by

$$c(j,a) = \frac{(1-a)2j^3}{a}.$$  

(6)

The consequence is that the cost functions are quite similar except as $j$ gets close to 1, in which region Foreign’s cost function becomes sharply higher than Home’s. Thus Foreign’s educational system has trouble generating the skills required for the most advanced industries. Other than that, the two countries are identical, with a uniform distribution of $a$’s and Leontieff preferences. Solving the equilibrium, we observe sorting down in Home for low-skilled workers, meaning that a worker of a given ability chooses a lower-skill industry than the worker would have chosen under autarky. At the same time, we observe sorting up for Home’s high-skilled workers. In other words, under trade, Home workers flee the middle-range industries. An interpretation is that Foreign’s educational costs discourage Foreign’s high-ability workers from pursuing the high-skill industries, so a disproportionate number of Foreign’s high-ability workers wind up in middle-range industries (a pattern exacerbated by trade with Home, which will lower the price of high-skill products). This pushes down the prices of middle-range goods compared with what would have been observed in Home under autarky, causing Home’s middle-ability workers to flee the middle, with upper-middle-ability workers fleeing upward and lower-middle-ability workers downward. Thus, in Home, trade hollows out the middle class. In addition, the effect of trade on welfare is nonmonotonic: Low-ability and high-ability Home workers benefit from trade, but because of the crash in the prices of medium-level goods, a range of middle-ability Home workers is hurt. Obviously, none of this discussion would have been possible in a model limited to high-skilled and low-skilled workers.
A closely related paper is by Costinot & Vogel (2010), who also look at a model with a continuum of goods, each of which is produced with labor alone and which differ in their skill intensities. Precisely, the output of an industry with skill-intensity index \( s \) is equal to \( A(s, \sigma) \) per worker, for a worker of skill level \( s \), where \( A(s, \sigma) \) is increasing in \( s \) and satisfies

\[
\frac{A(s', \sigma')}{A(s, \sigma)} > \frac{A(s', \sigma)}{A(s, \sigma)}
\]

for any \( s, s', \sigma, \) and \( \sigma' \) such that \( s' > s \) and \( \sigma' > \sigma \), so skill is disproportionally valuable in high-skill-intensive industries. This assumption is called complementarities in production. There is an exogenous supply of each of a continuum of different skill levels in each country, represented by the function \( V(s) \) for Home and \( V^*(s) \) for Foreign. Equilibrium is again a schedule of prices such that the way workers choose to sort themselves across industries, given that price schedule, creates supply that matches consumer demand for each good. The condition given in Equation 7 ensures that workers in each skill level choose one and only one industry and that higher-skilled workers match themselves in equilibrium to higher-skill-intensive industries.

With this framework, the authors are able to look at a number of interesting possible effects of trade on income distribution. First, they have a simple and elegant generalization of Stolper-Samuelson. If Home is skill abundant relative to Foreign, which means that \( V(s_0) > V^*(s_0) \) whenever \( s_0 > s \), then trade increases income inequality in Home, meaning that

\[
\frac{w'(s'_0)}{w'(s'_0)} > \frac{w(s'_0)}{w(s'_0)}
\]

whenever \( s_0 > s \), where \( w(s) \) denotes the wage paid to a Home worker of skill level \( s \) under autarky, and \( w'(s) \) is the corresponding wage under free trade. The opposite effect is found in Foreign. In addition, they analyze a simple concept of offshoring: Suppose that the technology in Home is superior to the technology in Foreign, in that the \( A(s, \sigma) \) function in Home is a scalar multiple of the function in Foreign. Suppose that under free trade, workers produce in Home with Home’s technology and workers in Foreign produce with Foreign’s technology. However, when offshoring is allowed, a producer in Home can hire workers in Foreign to produce output using Home’s superior technology. Costinot & Vogel show that this is equivalent to increasing the labor supply of Foreign across the board, and as a result it pushes down the wages of low-skilled workers in both countries, pushing up the wages of high-skilled workers in both countries, and raising income inequality in both countries in the sense of inequality (Equation 8). This is, of course, an interpretation of offshoring that is close to the Feenstra-Hanson view.

These are both generalizations of earlier results on North-South trade. Perhaps the most interesting point, however, involves findings on North-North trade. Suppose that Home’s economy is more diverse than Foreign’s, in the sense that there is a cutoff skill level \( s' \) such that among workers with skill levels less than \( s' \) Home is low-skill abundant relative to Foreign, but among workers with skill levels above \( s' \) Home is high-skill abundant relative to Foreign. In other words, compared with Foreign, Home has fatter tails in its skill distribution, rather than a difference in average skill abundance. Then when we let the two countries trade, low-skilled Home workers sort down; high-skilled workers sort up; and wages of middle-income Home workers fall relative to workers at
both ends of the spectrum. In other words, this is the hollowing out of the middle class studied by Blanchard & Willmann, arrived at by a somewhat different mechanism.

2.4. Labor-Market Frictions

A number of recent papers explore trade’s impact on income distribution in the presence of labor-market frictions. Mitra & Ranjan (2007), for example, apply models of search unemployment to examine the impact of offshoring. For reasons similar to the mechanism in Grossman & Rossi-Hansberg (2008), they find that offshoring in a given industry can lower domestic unemployment in that industry. The point is, once again, that offshoring can create a productivity benefit for domestic labor that induces domestic firms to increase the rate at which they create vacancies for domestic employment. In the long run, this reduces unemployment. Anderson (2009) studies a model in which workers must choose in which sector to acquire skills, becoming a specific factor after that choice is made; opening trade increases income inequality by increasing income differentials across industries.

A different approach to labor-market frictions is pursued in Artuç et al. (2008, 2010). In those papers, the authors assume that a worker can switch industries at any time, but must incur two costs. The first is a common cost, a parameter constant across time and the same for all workers. The second is idiosyncratic and time varying and can be negative. For example, a worker may become bored of her work, have an altercation with a supervisor, or need to move geographically for personal reasons to a part of the country where the industry she was in does not exist. Alternatively, the worker may be at the moment really enjoying her work or have children who are attached to their school friends, making a move costly. These idiosyncratic, time-varying costs are important because they allow for a model that generates an important fact in the data: Gross flows of workers across industries are an order of magnitude greater than net flows. At any given moment, between any two industries, one tends to see large numbers of workers moving in opposite directions at the same time.

Building these features into a rational-expectations model, one finds a number of implications for trade and income inequality. First, the effect of trade on the distribution of wages can be very different from the effect of trade on the distribution of lifetime incomes. It is easy to construct an example (and with realistic parameter values) of a trade liberalization that lowers real wages for the import-competing industry in the short run and the long run but that increases the expected lifetime utility of all workers in the import-competing sector. This is because of option value: Each worker in the import-competing sector knows that there is a positive probability that in a given number of years he will choose to move to one of the other sectors. Because trade liberalization raises the real wage in those industries, the value of that option has now gone up. [Similar issues arise in the search literature, as discussed at length in Davidson et al. 1999 (see Davidson & Matusz 2009, chapter 8, for an applied example).]

Second, announcing trade liberalization in advance tends to soften the blow for workers in the import-competing industry and also reduce the benefit to workers in the export industry. This is because of anticipatory movement of workers out of the import-competing industry, pulling up wages there before the liberalization occurs, and pulling down wages in the export sector. In the limit, with enough advance warning, all workers are guaranteed to have the same sign of net benefit from the liberalization, but this could be positive or negative.
2.5. Consumer Effects and Incomplete Contracts

We now look at two strands of literature that have not been explored much but could capture important pieces of the relationship between trade and inequality.

2.5.1. A consumer-side approach. A very different and potentially promising approach is taken by Fajgelbaum et al. (2009). They focus on the consumer-side effects of trade on income inequality. The model is built on two sectors, a competitive numéraire sector producing a homogeneous good and a monopolistically competitive sector in which products are differentiated horizontally (as in standard monopolistic competition models) but also vertically, so consumers can choose different varieties and also different qualities of differentiated products. Both sectors use only labor to produce output. Workers differ in their productivity according to an exogenous distribution. They all have the same utility function, which is nonhomothetic: Consumers with higher income demand higher-quality goods. If two economies open to trade, not only will the number of firms in each country and the product diversity available to each consumer be affected, as in standard monopolistic competition models, but the quality composition of goods also will be affected in complicated ways that depend on income inequality. The authors study an example with two countries that are identical except that one of them has more productive workers on average. When the cost of transporting high-quality goods falls, the number of high-quality firms rises, benefitting affluent consumers in both countries. This draws resources away from low-quality goods, reducing the product diversity available to low-income consumers and, for some parameter values, lowering their welfare.

The novelty in this model is that trade does not affect the distribution of income at all, in terms of the numéraire. That is fixed in each country by the exogenous distribution of worker productivities. It does, however, affect the distribution of real incomes because consumers at different income levels consume different goods. In this sense, it is a consumer-side account of trade and income distribution, whereas previous approaches work through the factor markets. Because in reality the rich and the poor certainly do consume different bundles of commodities and different qualities of goods, this channel may be an important one to explore in the future.

2.5.2. Implicit contracts. For people who do not live in an Arrow-Debreu economy, good luck can translate into high income and bad luck can lead to poverty. For this reason, risk-sharing institutions can have an enormous effect on income inequality, and to the extent that trade affects those institutions, they can be an additional channel through which trade affects inequality. An early exploration of this idea is by Matusz (1985), who incorporates a simple form of incomplete contracting from the macroeconomics literature into an HO model. In that model, firms in one industry suffer random, idiosyncratic productivity shocks. Employers are risk neutral and workers are risk averse. Employers would like to be able to offer employment contracts to workers that put the workers to work in a high-productivity state and lay them off in a low-productivity state, with a payment to the worker that does not depend on the state, but they are prevented from doing so because employers cannot credibly commit to paying the worker anything in a state in which the worker is not producing output. As a result, the low-productivity state has (inefficient) positive employment with a positive probability, with a positive wage paid only when the worker is employed. That the wage will be zero in the event of a layoff implies that firms
must pay a risk premium to workers, which lowers expected profits. Matusz shows that, in this sort of model, a weakened Stolper-Samuelson theorem holds—over a significant portion of the parameter space, trade raises the welfare of workers and lowers the returns to capital if the economy is labor abundant (and vice versa if it is labor scarce). However, even when that familiar relationship holds, it is possible that trade increases the unemployment rate in the implicit-contracts sector, with the wage for employed workers rising enough to give an increase in expected utility to workers. It is further possible that the aggregate unemployment rate rises even when the industry’s unemployment rate does not because a larger fraction of workers is drawn out of the full-employment sector and into the sector with implicit contracts and positive unemployment. In both cases, the point is that although the average worker’s welfare is increased by trade, because of implicit contracts, both unemployment and wage inequality can rise.

The idea has been extended to the invisible handshake studied by labor economists, the idea that a risk-neutral employer may offer wages smoothed over states of nature to a risk-averse worker, in effect selling insurance at the same time as the employer buys labor. However, as these arrangements are implicit contracts and based on shocks that are not observable to third parties, they depend on reputation built out of repeated interactions. The only punishment available to deter an employer from reneging on her wage commitment today is the loss of the worker tomorrow. Consequently, if employers are not sufficiently patient, only imperfect insurance can credibly be offered, and in that case an employer will cut wages in lean times. The upshot is that the more impatient an employer is, the more volatile individual workers’ wages will be, and the more variance will be observed in the cross section among observationally identical workers.

Bertrand (2004) follows the implications of this thinking in the context of international trade, showing that when firms face liquidity constraints and can exit due to bankruptcy, an import shock can make employers effectively less patient by raising the probability of bankruptcy and raising their effective discount rates. Thus a rise in imports in a given industry can increase wage inequality within that industry (and, by the same token, opening an export opportunity can reduce it). Karabay & McLaren (2010) examine invisible handshakes in a two-country general-equilibrium model with both goods trade and offshoring of tasks (in the primitive sense that they look at autarky versus free goods trade, and free goods trade versus complete integration of world goods and factor markets). Even though there is no bankruptcy in the model, trade has large effects on wage volatility through implicit contracts. An exporting sector sees a rise in its output price due to trade, which raises the amount the employer loses if the worker walks away because of a wage dispute. Therefore, the penalty to reneging on a wage promise is steeper, allowing the employer to make promises of stable wages with more credibility. Consequently, wage inequality falls within an export sector, with the opposite effect in an import-competing sector. At the same time, offshoring from a labor-scarce economy to a labor-abundant one makes it easy for an employer to find a new worker to replace one he has lost, thus reducing the punishment to reneging and making it harder for employers to promise stable wages credibly. As a result, an offshoring industry sees an increase in wage inequality, ceteris paribus. Putting these together, the model predicts that implicit-contract effects produce a net increase in wage inequality in a labor-scarce economy due to both forms of globalization together and produce the opposite result in a labor-abundant economy. (Of course, in practice these effects are combined with all the other effects on wage inequality highlighted above, so it could be difficult to disentangle the effect empirically.)
2.6. Summary of Theory Developments

The main thrusts of the theory can be summarized as follows. The older theory offered two stories: Trade affects inequality either by affecting the skill premium (in the Stolper-Samuelson theorem) or by affecting industry premia (in the specific-factors model). It was hard to rationalize how trade could raise inequality everywhere in the world at the same time, or inequality within any group of similarly skilled workers all doing the same job. It was also hard to see how North-North trade could affect inequality at all. But now we have stories that predict rising skill premia across countries as a result of North-South trade in tasks and even as a result of North-North trade in goods due to R&D effects or the skill bias of the transport sector. We have high-dimensional models that go beyond the skill premium to analyze the effect of trade on the middle class and distinguish between wage inequality and inequality in lifetime consumption through explicitly dynamic models of labor adjustment. We are also able to analyze the effects of trade on inequality among observationally identical workers doing the same job in the same industry, through heterogeneous-firms models or implicit-contracts models. This rich set of stories helps in describing the effects of trade on income distribution in the real world.

3. EMPIRICAL WORK

An immense empirical literature exists on the possible linkages between trade and inequality. Most of this literature predates the new trade models and focuses on testing the implications of the HO framework for trade-inequality linkages. A number of literature surveys also review this work (see, for example, Feenstra & Hanson 2001; Goldberg & Pavcnik 2004, 2007a). As in Section 2, we begin Section 3 with only a brief review of the empirical literature associated with the older trade models and then move on to review empirical papers that test the newer theories linking trade to inequality described in Section 2.

3.1. Earlier Work: Tests of the Heckscher-Ohlin and Specific-Factors Models

An excellent summary of the literature on trade and wages is provided by Feenstra (2000) in the introduction to The Impact of International Trade on Wages. Feenstra begins by documenting a sharp increase in the ratio of the wages of nonproduction workers to production workers between 1982 and 1994. Summarizing the papers in the volume, Feenstra concludes that there is some role for international trade in affecting the wages earned by American workers. Goldberg & Pavcnik (2007a), Feenstra (2008a,b), and others conclude that stylized facts on the evolution of inequality within developing countries as they open to trade are not consistent with a naïve view of the HO model. Davis & Mishra (2007) go further and argue that “Stolper-Samuelson is dead.” They write that the use of trade theory to suggest that liberalization will raise the wages of the unskilled in unskilled-labor-abundant countries is “worse than wrong—it is dangerous.” Davis & Mishra show that such arguments are based on a narrow interpretation of the Stolper-Samuelson theorem. In particular, the Stolper-Samuelson theorem holds only if all countries produce all goods, if the goods imported from abroad and produced domestically are close substitutes, or if comparative advantage can be fixed vis-à-vis all trading partners.
As an illustration, a poor country in a world with many factors and many goods may no longer have a comparative advantage in producing low-skill goods. This is easy to understand in the context of three countries; consider, for example, the United States, Mexico, and China. Although Mexico might have a comparative advantage in producing low-skill goods in trade with the United States, its comparative advantage switches vis-à-vis trade with China.

In part to address these and other shortcomings of the HO framework for explaining the rise in inequality within both developing and developed countries, as well as within industries, empirical investigations have branched out into a number of directions, including firm-level analysis, new approaches to trade in tasks and offshoring, and, to a limited degree, implicit contracts.

3.2. Empirical Work on Heterogeneous Firms and Bargaining

Because the literature on firm heterogeneity is so new, and because the data demands for testing these theories are quite high, few studies are available in this area. To take firm and/or worker heterogeneity into account properly, information at both the firm and individual employee level is typically required, suggesting the need for matched employee-employer data sets. In this section, we review several recent papers that have succeeded in contributing to this literature.

The pioneering work on trade and income inequality with heterogeneous firms actually predates the theory. Bernard & Jensen (1997) study the Annual Survey of Manufactures from the U.S. Census Bureau to decompose the large rise in average skilled wage premia that occurred over the 1980s. They show that a substantial fraction of the increase occurred between plants, in other words, by intra-industry shifts in the allocation of workers from plants with lower skill premia to firms with higher premia. This between-plant effect is larger than the within-plant effect (which is merely a rise in the skill premium for any one firm over time). Indeed, by some measures, the between-plant effect is completely dominant (Bernard & Jensen 1997, table 5). Furthermore, it occurs entirely among firms that export and vanishes when the sample is restricted to firms with only domestic sales. This was an early indicator that trade might cause an increase in wage inequality through within-industry effects, a mechanism very different from HO, and alerted the field that firm heterogeneity may have something important to do with the effect of trade on inequality.

More recently, Menezes-Filho & Muendler (2007) combine insights from the Melitz (2003) model with worker heterogeneity to provide a compelling empirical example of the importance of some of the more recent theoretical breakthroughs. These authors link worker-level panel data with firm-level and industry-level data to obtain a rich data set that allows them to test many implications of the most-advanced trade models (e.g., heterogeneous-firm models that incorporate heterogeneous labor) for Brazil. In so doing, the authors are able to assess the impact on jobs of Brazil’s trade liberalization during the 1990s while controlling for a number of worker-specific, firm-specific, industry-specific, and economy-wide structural reforms. Menezes-Filho & Muendler’s data set allows them to follow workers throughout the liberalization period and observe the path of their employment histories in greater detail than previous studies. They are particularly interested in the effects of trade liberalization on employment status, type of employment (formal or informal), and job reallocations.
Menezes-Filho & Muendler’s results show that exporting firms and firms in industries with a comparative advantage shed workers more frequently. Moreover, these same firms also hire workers less frequently than the average firm. Thus, on net, trade liberalization leads to net employment losses in these firms. This is surprising given the standard predictions of international trade models that would indicate that these sectors and firms would potentially hire more workers when liberalization occurs. Menezes-Filho & Muendler also show that tariff reductions and increased import penetration are associated with an increase in the likelihood of a worker transitioning into informality and unemployment, as well as with a lower probability of a transition from informality back to formal employment. Furthermore, they find that trade liberalization in Brazil has been associated with longer reallocation times for workers moving from one formal-sector job to another formal-sector job. Their results are robust to different levels of exposure to trade, firm-level productivity, and worker heterogeneity, as well as other general trends that occurred in the country during the period studied, such as skill-biased technological change and labor-market reforms.

Frias et al. (2009) use matched employer-employee data from Mexico to examine the wage premia paid by exporters in the Mexican manufacturing sector. Wage premia are defined as wages above what workers would earn elsewhere in the labor market. Because of the nature of their data, Frias et al. are able to decompose plant-level wages into a component that reflects skill composition and a component that reflects wage premia. Their identification strategy is based on the peso devaluation of 1994 that they argue differentially affected incentives to export within industries. Comparing across plants within industries, they find that roughly two-thirds of the higher level of wages in larger, more productive plants are explained by wage premia and that nearly the entire differential within-industry wage changes induced by the shock to exports is explained by wage premia and not by skills. The authors conclude that sorting on individual ability is not responsible for the well-documented correlation between exporting and wages.

As the first contribution in the literature to account for both firm heterogeneity and intermediate trade in their analyses, Amiti & Davis (2008) offer a theoretical and empirical examination of the impact of tariff cuts on workers’ wages that account for the extent of a firm’s engagement in international trade. Using Indonesian manufacturing census data for 1991–2000, a period that encompasses Indonesian trade liberalization, the authors develop a general equilibrium model to estimate this relationship. They find that the impact of a given tariff change on a firm’s workers’ wages is dependent on that firm’s role in the global economy. That is, a 10-percentage-point decrease in output tariffs will lower wages of import-competing firms by 3% but will raise wages at exporting firms by 3%. Likewise, a decrease in input tariffs by 10 percentage points will increase wages by 12% at firms that rely on imports but will have an insignificant impact on the wages of firms that rely on a domestic supply.

As Amiti & Davis (2008) summarize, their findings show that trade liberalization raises wages for workers at firms that are most globalized and lowers wages at firms that are either marginalized in the global economy or oriented toward the domestic market. This provides some confirmation for the ideas in theoretical work such as Egger & Kreickemeier (2009) and Helpman et al. (2010) that hypothesizes a relationship between firm-specific wages and the firm’s response to globalization.

Bustos (2007) posits that an examination of wage inequality focused on the interaction between trade and technology—as opposed to selecting one explanation in preference over
the other—will offer a better explanation of the relationship between global trade and inequality. Bustos therefore presents a model of the relationship between trade liberalization and demand for technology and skill among firms in developing countries that accounts for firm heterogeneity. She tests this model using panel data from Argentine manufacturing firms. The data set spans 1992–1996 and therefore captures a period of trade and capital account liberalization in Argentina.

Bustos finds a strong relationship between exporting and increases in technology investment and skill upgrading. Specifically she finds that, prior to trade liberalization, continuing exporters and foreign-owned firms employed higher-skilled labor than those firms that were domestically owned and that had never exported. Those firms that began exporting after liberalization upgraded worker skill more quickly than those firms that remained exclusively in the domestic market; they also upgraded technology more quickly than all other firms. Furthermore, after trade liberalization, new and continuing exporters as well as foreign-owned firms spent 53%–69% more on technology than their domestic nonexporting counterparts. Those firms that invested more in technology upgrading also realized a faster increase in skilled labor. Bustos concludes that, due to the consequences of rising demand for technological investment, trade liberalization can have a strong impact on worker-skill upgrading.

Although much of the literature looks at the impact of the act of exporting on firm behavior (for example, Bustos 2007, Verhoogen 2008), Brambilla et al. (2010) focus on how the destination of those exports explains firm behavior. These authors develop and then test an integrated theory of export destinations and skills. Exploring the linkages among exports, export destinations, and the use of skilled labor by firms, they theorize that firms exporting to high-income destinations will hire a higher proportion of high-skilled workers and will pay them higher wages than firms that either export to low- or middle-income countries or sell their products domestically. They test their theory with 1998–2000 panel data from Argentine manufacturing firms. Their findings show that Argentine firms exporting to high-income countries are associated with higher-skilled workers and higher average wages than firms that either do not export or export to middle-income countries. However, they find no significant difference in firms’ use of skilled labor between those firms selling their product domestically and those exporting to low- and middle-income countries. The authors reason that these results are due to the similarities in the domestic and export markets among low- and middle-income countries. Their theory and results are in line with Verhoogen’s (2008) findings on the quality upgrades of exporting firms and Bustos’ (2007, 2011) work on technology and skill-upgrading behavior of exporting firms. The results further suggest that nonhomotheticities in demand are important for income distribution (as suggested in theory, although not with the same mechanism, in Fajgelbaum et al. 2009).

To sum up, the emerging work on trade and inequality with firm-level data appears to be confirming a central role for between-firm effects in governing the relationship between trade and inequality, and the available results seem to support the thrust of theoretical models in Section 2.2 that predict a rise in inequality with more openness.

3.3. New Empirical Work on Trade in Tasks

Much new empirical work focuses on the fragmentation of the production process, or offshoring. Recall from Section 2.3.1 that one of the pioneering theory models in this area...
was provided by Feenstra & Hanson (1996); not surprisingly, they also pioneered the empirical work (Feenstra & Hanson 1996, 1997, 1999). In that model, offshoring increases the relative demand for skilled labor in both countries involved because the offshored tasks are more skill intensive than those previously performed in the country to which they were offshored, but they are less skill intensive than those in the country that is doing the offshoring.

Feenstra & Hanson (1999) test whether their explanation for the increase in the demand for skill, based on more offshoring, is consistent with the pattern of increasing wage inequality in the United States. They consider the alternative hypothesis that skill-biased technological change accounted for the observed increase in wage inequality. They proxy for technical change with the fraction of high-technology equipment in each industry’s capital stock, and they measure offshoring with the intermediate inputs imported by each industry. They use a two-step procedure, first to identify the impact of offshoring and high-technology investments on productivity and prices and then to trace through the induced productivity and price changes to calculate production and nonproduction wages.

Using data for the U.S. manufacturing sector between 1979 and 1990, Feenstra & Hanson (1999) find that 25% of the increase in the relative wage of nonproduction workers was explained by offshoring and about 30% by technological change. They conclude that both offshoring and the increased use of high-technology capital are important in explaining the increase in the relative wage of skilled workers. They also examine the impact on real wages as distinct from relative wages, which are the focus of measures of inequality. They find that the real wages of production workers were probably unaffected by offshoring activities, while the real wages of nonproduction workers increased by 1–2 percentage points. Sitchinava (2008) updates the Feenstra & Hanson (1999) paper to 1996 data and also takes into account the possibility of service offshoring. Sitchinava finds that most of the increase in the relative wages of nonproduction relative to production workers can be explained by technical change, which is proxied with the share of high-technology equipment in the capital stock.

Adapting Feenstra & Hanson (1999) to measure service offshoring, Amiti & Wei (2009) provide evidence for the effects of both service and material offshoring on domestic productivity growth. Using U.S. Bureau of Labor Statistics data from 1992 to 2000, they find that service offshoring has a significant positive effect on labor productivity growth, accounting for approximately 10% of average growth in this factor. Although material offshoring also has a positive effect, it is smaller in magnitude—accounting for 5% of average growth in labor productivity—and is not significant across all specifications. They conclude that service offshoring does have a positive impact on labor productivity growth in the United States and speculate that the smaller and less significant values for material offshoring may be due to possible decreasing returns from scale and over time from this sector.

Liu & Trefler (2008) analyze the impact of not only offshoring but also inshoring—the sale of services produced in the United States to unaffiliated buyers in China and India—across several indicators: workers’ change of occupation and industry, weeks spent unemployed as a share of total weeks in the labor force, and earnings. They find that the total net effect of inshoring and offshoring is positive. However, for those workers in industries exposed to offshoring and those workers who are less educated, the effect is either less positive or negative.
A different approach is explored by Ebenstein et al. (2009), who focus on the effects of trade across different types of task, as measured by the routineness of different occupations. Typically, highly routine occupations are associated with workers who have lower educational attainment, whereas less routine occupations are associated with higher skills and educational attainment.

Why should routineness matter? Grossman & Rossi-Hansberg (2008) posit that improvements in technology make offshoring less costly. Cost reductions are much more likely for routine tasks, which are more easily codified and can be communicated and consequently transferred to overseas affiliates. Ebenstein et al. test this hypothesis by assessing the empirical relationship between the routine nature of a task and offshoring. Their measure of routine is based on Autor et al. (2003), who describe routine jobs as “tasks that can be expressed using procedural or ‘rules-based’ logic, that is codified in a fully specified sequence of logical programming commands (‘If-Then-Do’ statements) that designate unambiguously what actions the machine will perform and in what sequence at each contingency to achieve the desired result.” Although Autor et al. (2003) use routineness to designate which jobs can be easily performed by computers, the jobs that are classified as routine also include the jobs in manufacturing that we typically think of as being offshorable. These jobs include attaching hands to faces of watches, sewing fasteners and decorative trimming to articles, and, although not mentioned explicitly in their paper, service tasks that we think of as offshorable such as answering telephones.

We can contrast this occupation-based approach to a well-known alternative, which is to examine the changes in import penetration across industries [what Goldberg and Pavcnik (2007a) refer to as the “differential exposure approach”]. A difficulty with that approach is that, in the event that workers can change industries more easily than occupations, it will miss the main effect; industry premia will be largely arbitrated away, but premia to each occupation can be large and much affected by trade. Ebenstein et al. address this problem by calculating an occupation-specific measure of offshoring, import competition, and export activity and show that, although international trade has not had large effects on industry premia, it has had large, significant effects on occupation-specific wages for routine workers.

Ebenstein et al. (2009) merge Current Population Survey data on U.S. wage earners from 1983 to 2002 with data on import competition, export activity, and offshoring employment of U.S. multinational firms to show that the impact of offshoring on labor-market outcomes depends both on the location of offshore activity and on the routineness of the task performed by the worker. Expansion in offshore employment in low-income locations is associated with wage reductions for routine workers. However, offshore activity in high-income locations is positively correlated with routine wages. These associations, which are significantly stronger in the 1990s relative to the previous decade, parallel earlier findings by Bernard et al. (2006), who show that U.S. manufacturing plants whose dominant industry is one in which low-wage-country imports are large are more likely to exit and less likely to expand, and these differences are more pronounced for more labor-intensive plants; imports from other countries have a much smaller effect, and sometimes the opposite sign. These two studies suggest that the effect on U.S. income distribution of import competition from, and offshoring to, low-wage countries seems to be qualitatively different from the effect of high-wage countries.

Ebenstein et al. (2009) also find significant effects of import competition on employment reallocation, finding that the largest effect of globalization on low-skilled workers’
income comes from movement from higher-wage industries to lower-wage ones. This parallels earlier findings by Bernard & Jensen (1997), who show that between-industry effects were a large component of increases in the average skill premium in the 1980s, although they were not able to pin down globalization as the cause.

Ebenstein et al. (2009) also find much stronger effects of offshore activities on domestic wages in the later part of the sample period, between 1997 and 2002. Occupation-specific changes in offshoring and trade are associated with significant wage effects, particularly for workers who are in routine occupations. For these workers, a 1-percentage-point increase in low-income offshore affiliate employment is associated with a 0.11% fall in wages. For these same workers, however, increasing affiliate activity in high-income locations is associated with a 0.1% increase in wages. A 1% increase in export shares is associated with a 1% increase in wages, while a 1% increase in import penetration is associated with a 0.46% decline in wages. The effects of these globalization measures are generally small in magnitude and insignificant for individuals who are in the least-routine occupations.

Ebenstein et al. (2009) also find that the net impact of offshoring on wages is a function of the nature of the job: Workers who perform more routine tasks have experienced wage declines as a result of offshoring, whereas workers who perform nonroutine tasks have experienced wage increases. For routine occupations, which are more easily transferred offshore, the net effect on wages is negative, but for the least-routine (skilled) occupations, the net effects are positive. Recalling the Grossman & Rossi-Hansberg (2008) model of Section 2.3.1, one might interpret these findings roughly as implying that the labor-supply effect of offshoring exceeds the productivity effect for the lower-skilled workers who specialize in the routine tasks that are more easily offshored.

Hummels et al. (2010) analyze the relationship between offshoring and workers’ wages and employment opportunities. They use a matched worker-firm data set that encompasses the Danish labor force from 1995 to 2006 as well as a data set on offshoring at the firm level. They estimate the impact of exogenous shocks to offshoring and exporting on firm characteristics and on the wages of individual workers. Furthermore, they assess the dependency of these estimates on the education and occupational characteristics of the workers so as to understand the relative sensitivity of types of workers and occupations to offshoring. They find that exogenous import shocks will have significant and opposite effects on skilled and unskilled labor wages: Skilled labor wages will increase by 8.5%, while unskilled labor wages will decrease by 7.3%. In contrast, shocks to exporting will increase both skilled and unskilled labor wages, but low- and medium-skilled workers will see a greater increase.

The examination by Hummels et al. of the role of occupational characteristics in wage inequality reveals that workers who are exposed to unsafe working conditions and workers in the natural sciences and engineering will see their wages fall in the case of offshoring shocks, while workers in the social sciences and language industries will be less affected by those shocks.

Hummels et al. (2010) also consider the relationship between displaced workers and offshoring, finding that workers displaced by offshoring generally experience greater and more persistent wage and earnings loss than workers displaced for other reasons. Although initially both low- and high-skilled displaced workers experience wage loss, this loss is smaller and less persistent for high-skilled workers. A year after losing their jobs to offshoring, skilled workers will have lost 19% of their predisplacement earnings (which
accounts for both lost hours and lowered wages once the worker has re-entered the workforce) as compared with the 28% loss experienced by unskilled workers. Additionally, Hummels et al. find that it is essential to control for endogeneity of trade events in such an analysis.

An additional firm-level study by Sethupathy (2009) shows that, over the period in which the North American Free Trade Agreement came into force, lowering costs of offshoring within North America, U.S. firms that already offshored to Mexico significantly increased (a) their offshoring to Mexico, (b) their operating profits per U.S. worker, and (c) the wages they paid to their U.S. workers—without, apparently, reducing their U.S. workforce. This suggests that some of the productivity benefits of the Grossman & Rossi-Hansberg (2008) model have been realized by those firms, and have been captured by those workers, but it should be noted that this finding does not imply any benefit to workers outside of those firms (unlike the Grossman-Rossi-Hansberg model).

In sum, recalling the account in Section 2.3.1, there are two main stories that have emerged from the theoretical literature: first, that offshoring can raise wage inequality in both countries, as in Feenstra & Hanson (1996), and, second, that offshoring can raise the real wages of unskilled workers by enhancing their effective productivity, as in Grossman & Rossi-Hansberg (2008). These two predictions are distinct, but not mutually exclusive. There is now fairly strong evidence for the first story, the income-distribution effect, from multiple data sets and approaches. The second story has done less well when confronted with data, at least when the focus is on offshoring to low-wage countries.4

3.4. Implicit Contracts

Empirical tests of implicit-contract models with trade are rare. Bertrand (2004) points out that if implicit contracts are effective, then a worker’s wages will be affected by labor-market conditions such as local unemployment rates at the time the worker joined the firm, but not by subsequent labor-market conditions. The reason is that the worker and firm will bargain for their optimal implicit contract at the beginning of their relationship, at which point the current unemployment rate will have an effect on the worker’s bargaining power and hence on the wage agreed to, but if implicit contracts are strong and provide good insurance to the worker, for the remainder of the job, wages will simply follow the agreed-upon wage regardless of subsequent labor-market conditions. She finds that (a) a worker’s current wage is affected by initial conditions at the beginning of the job, independent of current labor-market conditions, and (b) in industries hit by a rise in import penetration, the current wage is much more dependent on the current labor-market conditions compared with other industries. Together, these findings suggest that implicit contracts are important and that import competition indeed weakens them, as predicted by her model.

3.5. Labor-Market Frictions

Krishna & Senses (2009) offer an empirical study of the impact of openness to trade on domestic income. Whereas previous studies examine the impact on wage growth or wage premia, Krishna & Senses examine the impact on wage volatility. Using longitudinal

---

4It is worth noting that in Bernard et al. (2006) and Ebenstein et al. (2009), Mexico is not classified as a low-wage country.
earnings data on workers from three panels (spanning 1993–2003) of the Survey of Income and Program Participation, they estimate the relationship between labor income risk (defined as the variance of unpredictable changes in earnings) and import penetration, a measure of industry exposure to international trade. They find that a 10% increase in import penetration will increase the standard deviation in persistent (as opposed to transitory) income shocks by 20%–25% for all workers. Their results are both statistically and economically significant.

Krishna & Senses (2009) also estimate subsets of their data to identify the impact of openness to trade on particular industries and on workers who have changed employment from one industry to another; they find higher income risk among workers who have switched from one industry to another. Among those who switched, income risk was higher among those who moved to nonmanufacturing sectors than those who switched within manufacturing sectors. In light of their findings, Krishna & Senses (2009) conclude that the impact on labor income risk needs to be taken into account when calculating the costs of openness to international trade.

Robustness tests by Krishna & Senses (2009) reveal that controlling for offshoring causes the coefficient on import penetration to increase. In addition, the offshoring variable is negative and significant, suggesting that an increase in offshoring in a particular industry is associated with a decrease in income risk in that industry.

Artuç et al. (2010) and Artuç & McLaren (2010) estimate and simulate the dynamic model of labor adjustment developed in Cameron et al. (2007) and Chaudhuri & McLaren (2007) to assess the distributional effects of trade shocks. The former study uses U.S. Current Population Survey data and the latter uses 2004–2006 data from the Household Employment Survey of the Turkish Statistical Institute. The studies estimate both the average cost of switching industries and the variance of idiosyncratic switching costs and use the estimates to simulate a trade shock to the manufacturing sector. In both cases, the authors find that, due to the high costs of switching from one industry to another, the economy takes a decade to reach the new steady state after liberalization. During this time, workers move from the manufacturing sector to other sectors, wages in the manufacturing sector first drop then rise as labor supply to that sector falls, and wages in other sectors at first rise and then fall as labor supply to those sectors rises. However, throughout these fluctuations, the real wage of the manufacturing sector remains below that of the tariff steady state, while the nonmanufacturing sector real wage remains above it.

Importantly, the distributional effects of the trade shock on lifetime expected utility are much smaller than the effects on wages, once each worker’s future possible mobility and option value are taken into account. In particular, in many specifications import-competing workers’ lifetime welfare rises despite a drop in their wages because each manufacturing worker understands that there is a probability each year that she will choose to enter the expanding export sector and benefit from the increased real wages there. A dynamic approach with a full accounting of option value therefore complicates the welfare analysis of income inequality. For example, recall that Ebenstein et al. (2009) argue that offshoring to low-wage countries has pushed large numbers of U.S. workers from high-wage manufacturing jobs into lower-wage service-sector jobs. In a dynamic model, these workers may nonetheless benefit from such offshoring because each manufacturing worker knows that with some probability each year he will move into the service sector.
anyway; the value of this option is enhanced by any measure that raises the real wage in that sector.

4. DIRECTIONS FOR FUTURE RESEARCH

Research in the 1990s undermined the simple HO theory linking trade and inequality, and economists pointed at nontrade factors, such as technology and labor institutions or laws, to explain rising inequality. However, now a number of other channels have been discovered that have led to a vigorous resurgence of the idea that trade can lead to a rise in inequality—with the new features that it can do so through North-North trade, in countries of the South, and within each industry and within each class of workers. These theoretical developments have been fed by empirical work in important respects and are now in turn giving rise to a rich new empirical literature, partly because of the increasing access to firm-level data across an increasing range of countries.

A number of natural directions for future work suggest themselves. The new ideas on consumer-side effects (Fajgelbaum et al. 2009) and on higher-dimensional income-inequality effects (Section 2.3.2) await empirical exploration (with the exception of Broda & Romalis 2009 for the former), whereas the interactions of trade with imperfect contracting have had little empirical attention. Most empirical work still focuses on the manufacturing sector, which for most countries covers a minority of the workforce, while major effects of globalization may make themselves felt in the remaining sectors (Ebenstein et al. 2009). Most work on inequality has been less focused on whether real incomes have increased or fallen due to trade than on the variance of incomes, and it is still difficult to disentangle trade effects from technology shocks. There has been little attention in theory or in the data given to the effects of trade on income inequality across age categories (Artuç 2009), and the empirical analysis of unemployment lags far behind the theory (Davidson & Matusz 2009). All these can usefully be placed on the agenda for the coming years.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Amiti M, Wei S-J. 2009. Service offshoring and productivity: evidence from the US. World Econ. 32:203–20


*www.annualreviews.org* • *Recent Perspectives on Trade and Inequality* 287


Matusz SJ. 1985. The Heckscher-Ohlin-Samuelson model with implicit contracts. Q. J. Econ. 100(4):1313–29

