The Review *of* Economics *and* Statistics

VOL. LXXXVIII | NOVEMBER 2006 | NUMBER 4

INCREASING RETURNS, IMPERFECT COMPETITION, AND FACTOR PRICES

Paolo Epifani and Gino Gancia*

Abstract—We show how, in general equilibrium models featuring increasing returns, imperfect competition, and endogenous markups, changes in the scale of economic activity affect the income distribution across factors. Whenever final goods are gross substitutes (gross complements), a scale expansion raises (lowers) the relative reward of the scarce factor or the factor used intensively in the sector characterized by a higher degree of product differentiation and higher fixed costs. Under very reasonable hypotheses, our theory suggests that scale is skill-biased. This result provides a micro foundation for the secular increase in the relative demand for skilled labor. Moreover, it constitutes an important link among major explanations for the rise in wage inequality: skill-biased technical change, capital-skill complementarities, and international trade. We provide new evidence on the mechanism underlying the skill bias of scale.

I. Introduction

UNDERSTANDING the effects of changes in market size on factor rewards is of central importance in many contexts. It is well recognized that international trade, technical progress, and factor accumulation are all vehicles for market expansion. Yet, despite the interest in the distributional effects of each of these phenomena, very little effort has been devoted to studying the distributional consequences of the increase in the scale of economic activity they all bring about. This is the goal of our paper. In particular, we study the effects of a market size expansion in a two-sector, two-factor, general equilibrium model with increasing returns, imperfect competition, and endogenous markups. Our main result is that, under fairly general conditions, scale is nonneutral on income distribution.

Given that there is no unified theory of imperfect competition, we derive our main results within three widely used models: contestable markets (Baumol, Panzar, & Willig, 1982), quantity competition (Cournot), and price competition with differentiated products (Lancaster, 1979). These models share a number of reasonable characteristics: the presence of firm-level fixed costs, free entry (no extra profits) and, most important, the property that the degree of competition is endogenous and varies with market size. In particular, as in Krugman (1979), in these models a scale expansion involves a procompetitive effect, which forces firms to lower their markups and increase their output to cover the fixed costs. This is a key feature for our purpose.

We allow the two sectors to differ in factor intensity, degree of product differentiation, and fixed and marginal costs. On the demand side, the elasticity of substitution in consumption between final goods is allowed to differ from 1. Under these assumptions, we show that any increase in market size is generally nonneutral on relative factor rewards. More precisely, whenever final goods are gross substitutes, a scale expansion raises the relative reward of the factor used intensively in the less competitive sector. This is the sector characterized by a combination of smaller employment of factors, higher degree of product differentiation, and higher fixed costs. An interesting implication of our result is that, in the absence of sectoral asymmetries in technology or demand, a scale expansion benefits the factor that is scarcer in absolute terms (that is, in smaller supply). The reverse happens when final goods are gross complements, and it is only in the knife-edge case of a unitary elasticity of substitution that scale is always neutral on income distribution.

The reason for this result is that, in the models we study, equilibrium economies of scale fall with the degree of competition. This is a natural implication of oligopolistic models approaching perfect competition as market size tends to infinity. As a consequence, a less competitive sector has more to gain from market enlargement, in that a larger market would effectively increase its productivity relative to the rest of the economy and hence expand (reduce) its income share if the elasticity of substitution in consumption is greater (less) than 1.

Our characterization of the factor bias of scale has several theoretical implications. Given that intra-industry trade between similar countries can be isomorphic to an increase in market size, our theory suggests which factor stands to gain more from it. In doing so, it fills a gap in the new trade theory, where the distributional implications of two-way trade in goods with similar factor intensity are often overlooked (see for example, Helpman & Krugman, 1985). Likewise, our theory suggests that technical progress, by

Received for publication August 4, 2004. Revision accepted for publication November 10, 2005.

* Boconi University and CESPRI, and CREI and UPF, respectively.

We thank Daron Acemoglu, Philippe Aghion, Giovanni Bruno, Antonio Ciccone, Torsten Persson, Dani Rodrik, Jaume Ventura, Fabrizio Zilibotti, two anonymous referees, and seminar participants at CRE1, Universitat Pompeu Fabra, IIES, Stockholm University, CESPRI, Bocconi University, and the Conference on Economic Growth and Distribution (Lucca, 2004), for comments. Remaining errors are our own. Gino Gancia thanks the Spanish Ministerio de Ciencia y Tecnologia (grant SEC2002-03816) and the Barcelona Economics Program of CREA for financial support.

increasing market size, is nonneutral on the income distribution, thus contributing to the recent literature on the factor bias of technical progress (for example, Acemoglu, 2002, 2005). Finally, our results imply that factor demand curves may be upward sloping for low levels of factor employment: if the endowment of some factor is very small, the sector using the factor intensively may be subject to increasing returns so strong that a marginal increase in the factor supply actually raises its reward.

We also provide two extensions of our framework. First, we derive simple conditions for the factor bias of scale within the Dixit-Stiglitz model of monopolistic competition, a very tractable and widely used model featuring constant markups. We show that the results are very similar to those mentioned earlier, although the mechanism is different, in that it does not rely on the procompetitive effect of scale, but rather on external increasing returns arising from a variety effect. Second, we briefly discuss the distributional effects of a biased scale expansion, that is, an endowment increase associated with a change in the factor ratio. This exercise can be isomorphic to trade integration among countries that differ in relative factor scarcity and helps understand how the factor bias of scale may alter the distributional implications of the standard Heckscher-Ohlin model. Interestingly, we show that, if the scale effect is strong enough, a factor that is scarce both in absolute terms and relative to other countries may experience an increase in its relative reward after trade opening. In other words, under certain conditions, the mechanism we emphasize may overturn the Stolper-Samuelson prediction.

A prominent application of our results is in the debate over the causes of the widespread rise in wage inequality that took place since the early 1980s. The theoretical literature has identified three main culprits: skill-biased technical change, capital-skill complementarity, and international trade. Our theory suggests the existence of a neglected link among these explanations, namely, the skill bias of scale. In particular, we review evidence showing that skilled workers, in any country, constitute a minority of the labor force, are employed in sectors where plant-level fixed costs are high, and produce highly differentiated goods that are gross substitutes for less-skill-intensive products. Under these circumstances, our theory implies that scale is skill-biased, thereby providing a micro foundation for the perpetual increase in the relative demand for skilled workers.[1] Moreover, because technical change, as well as factor accumulation and trade integration, implies a market size increase, we conclude that all three are essentially skill-biased phenomena, even in the absence of technology biases, complementarity among inputs, or Stolper-Samuelson effects.

Finally, we provide evidence on the mechanism underlying the skill bias of scale. In particular, we confront our theory with data from the NBER productivity file, a unique database on industry-level inputs and outputs widely used to investigate the determinants of the rise in wage inequality in the United States. We find strong evidence that markups fall when industry size rises and that they fall by more in the skill-intensive industries, where they are higher. In line with our model's predictions, these results suggest that the procompetitive effect of scale expansion is stronger in the skill-intensive industries, which are less competitive and hence benefit more from industry expansion. We conclude by comparing our findings with the related literature on wage inequality.

[1] See, among others, Wood (1998) for evidence on the secular increase in the relative demand for skilled labor.

II. Increasing Returns, Imperfect Competition, and Factor Prices

Consider a country endowed with V_i units of factor i and V_j units of factor j, where two final goods are produced. Consumers have identical homothetic preferences, represented by the following CES utility function:[2]

$$U = \left[\gamma(Y_i)^{\frac{\epsilon-1}{\epsilon}} + (1-\gamma)(Y_j)^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}}, \tag{1}$$

where Y_i (Y_j) stands for consumption of the final good intensive in factor i (j), and ϵ is the elasticity of substitution between the two goods. γ is a parameter capturing the relative importance in consumption of the i-intensive good. The relative demand for the two goods implied by equation (1) is

$$\frac{P_i}{P_j} = \frac{\gamma}{1-\gamma}\left(\frac{Y_j}{Y_i}\right)^{1/\epsilon}, \tag{2}$$

where P_i and P_j are the final prices of goods Y_i and Y_j, respectively.

We focus deliberately on sectoral production functions that are homothetic in the inputs they use, or else the nonneutrality of scale would be merely an assumption. It follows that, as also shown below, a scale expansion that leaves V_i/V_j unchanged can affect relative factor prices (w_i/w_j) only as long as it changes the income shares of sectors. In turn, equation (2) implies that relative sectoral shares are entirely characterized by either the relative price of goods or the relative output:

$$\chi \equiv \frac{P_iY_i}{P_jY_j} = \left(\frac{\gamma}{1-\gamma}\right)^{\epsilon}\left(\frac{P_i}{P_j}\right)^{1-\epsilon} = \frac{\gamma}{1-\gamma}\left(\frac{Y_i}{Y_j}\right)^{(\epsilon-1)/\epsilon}. \tag{3}$$

To see the effect of scale, note that homotheticity of Y_i implies that its dual total cost function C_i takes the separable

[2] We assume CES preferences for tractability, although our results do not depend crucially on the assumption that the price elasticity is constant. More general demand systems would certainly complicate the analysis, but our results would hold at least "locally."

form $C_i(w_i, w_j, Y_i) = \tilde{c}_i(w_i, w_j)e(Y_i)$, with $e'(Y_i) > 0$. Then, assuming zero profits and exploiting the homogeneity property of cost functions, equation (3) can be rewritten as

$$\frac{\tilde{c}_i(w_i/w_j,\, 1)e(Y_i)}{\tilde{c}_j(w_i/w_j,\, 1)e(Y_j)} = \chi = \frac{\gamma}{1-\gamma}\left(\frac{Y_i}{Y_j}\right)^{(\epsilon-1)/\epsilon}.$$

Differentiating it with respect to w_i/w_j, Y_i, and Y_j, we can compute the effect of marginal output changes on relative factor rewards:

$$\begin{aligned} d\left(\frac{w_i}{w_j}\right)\left[\frac{\tilde{c}_{ii}}{\tilde{c}_i} - \frac{\tilde{c}_{ji}}{\tilde{c}_j}\right] + \frac{e'(Y_i)Y_i}{e(Y_i)}\,\hat{Y}_i - \frac{e'(Y_j)Y_j}{e(Y_j)}\,\hat{Y}_j \\ = \left(1 - \frac{1}{\epsilon}\right)(\hat{Y}_i - \hat{Y}_j), \end{aligned} \tag{4}$$

where a hat denotes a proportional variation, $\tilde{c}_{ii} \equiv \partial\tilde{c}_i/\partial(w_i/w_j)$, and $\tilde{c}_{ji} \equiv \partial\tilde{c}_j/\partial(w_i/w_j)$. Note that $\tilde{c}_{ii}/\tilde{c}_i > \tilde{c}_{ji}/\tilde{c}_j$, because Y_i is by assumption intensive in factor i. Note, also, that the factor $e'(Y_i)Y_i/e\,(Y_i)$ is the output elasticity of the total cost function, an inverse measure of returns to scale. It can be shown (see, for example, Hanoch, 1975) that its reciprocal equals the scale elasticity of sectoral output ($e_s^{Y_i}$), that is, the elasticity of output with respect to an equiproportional increase in all inputs:

$$e_s^{Y_i} \equiv \frac{d\,\log\,Y_i(sV_i^i,\, sV_j^i)}{d\log s} = \frac{\hat{Y}_i}{\hat{s}} = \frac{e(Y_i)}{e'(Y_i)Y_i}, \tag{5}$$

where V_i^i and V_j^i are the employments of factors i and j in sector $Y_{i,}$ and $s > 0$ is a scaling parameter, evaluated at $s = 1$. Using equation (5) in (4), we obtain the change in relative factor rewards after a proportional expansion, $\hat{s}$, of all inputs:

$$d\left(\frac{w_i}{w_j}\right) = \left(\frac{\tilde{c}_{ii}}{\tilde{c}_i} - \frac{\tilde{c}_{ji}}{\tilde{c}_j}\right)^{-1}\left(1 - \frac{1}{\epsilon}\right)(e_s^{Y_i} - e_s^{Y_j})\hat{s}. \tag{6}$$

Equation (6) shows that scale affects relative factor prices as long as returns to scale differ across sectors ($e_s^{Y_i} \neq e_s^{Y_j}$) and $\epsilon \neq 1$. The intuition for this result is simple. After a scale increase, output grows relatively more in sectors with stronger increasing returns; if goods are gross substitutes ($\epsilon > 1$), prices react less than quantities, so that the income share of the high-increasing-returns sector expands and so does the relative return of its intensive factor. The reverse happens when goods are gross complements ($\epsilon < 1$); it is only in the knife-edge case of a unitary elasticity of substitution that income shares are always scale-invariant.

What are then the determinants of returns to scale? To address this question, we first note that in models of imperfect competition featuring free entry and fixed costs in production, increasing returns and market power are closely related. Because firms charge a price in excess of marginal costs, the markup function $R(\cdot)$, defined as the ratio of average to marginal revenue, is a measure of monopoly power. Likewise, the function $\theta(\cdot)$, defined as the ratio of average to marginal cost, is a measure of economies of scale internal to firms. When profits are driven down to zero by free entry, in equilibrium the degree of monopoly power must be equal to the degree of economies of scale:[3] $R(\cdot) = \theta(\cdot)$. The reason is that operational profits (that is, the surplus over variable costs) must be just enough to cover fixed costs, and fixed costs generate increasing returns. This immediately suggests that sectors may differ in increasing returns because of differences in market power.

The models we study next explore this possibility and show how the factor bias of scale depends on basic parameters. Before moving on, we want to stress an important point: increasing returns at the firm level matter only as long as the scale of production of a typical firm grows with overall market size. We consider this a realistic property and focus on market structures (the majority) where it holds; however, we will also see that our results extend to some form of increasing returns that are *external* to firms.

To anticipate our main findings, we will see that in the simplest case of contestable markets, where there is a single firm per sector and price equals average cost, increasing returns depend only on the ratio of fixed cost to sectoral output. Clearly, smaller sectors enjoy stronger increasing returns. The Cournot case of competition in quantities will show that in general market power also depends on demand conditions, such as the elasticity of substitution between products. High substitutability implies a very elastic demand that limits the ability of firms to charge high markups, thereby translating into low increasing returns. Price competition with differentiated products (following the ideal variety approach) will demonstrate that the Cournot result is not a special one; further and more importantly, it will illustrate another source of increasing returns common in models with product differentiation: scale economies *external* to firms due to a preference for variety in aggregate. Instead of modifying our previous findings, this new element will just reinforce them. As a comparison, we will also show that in the Dixit-Stiglitz (1977) model of monopolistic competition, where firm size is constant, only this latter effect survives.

We now turn to the detailed analysis of specific cases. To preserve the highest transparency, we limit our study to the simplest specific-factors model, where $\chi = w_iV_i/w_jV_j$. As shown above, similar results can in fact be derived from any homothetic sectoral production functions, provided that the factor intensity differs across sectors.[4]

A. *Contestable Markets*

We start with one of the simplest forms of imperfect competition: contestable markets, in which the threat of

[3] See Helpman and Krugman (1985) for a formal derivation.

[4] To have a sense of how our results carry over to the case of nonextreme factor intensities, in the Appendix we illustrate the contestable markets model in the case of Cobb-Douglas production functions.

entry drives down prices to average costs even if goods are produced by monopolists. Assume that there are many potential competitors (indexed by υ) who can produce good Y_i with the same technology. In particular, the total cost function of each producer in sector i entails a fixed requirement, F_i, and a constant marginal requirement, c_i, of efficiency units of factor i:

$$C_i(\upsilon) = [F_i + c_i y_i(\upsilon)]w_i, \tag{7}$$

where $y_i(\upsilon)$ is the amount produced by a single firm, and w_i is the reward of one unit of factor i.

A contestable market equilibrium is defined by the following conditions: market clearing [that is, $\sum_\upsilon y_i(\upsilon) = Y_i$], feasibility (meaning that no firm is taking losses) and sustainability (requiring that no firm can profitably undercut the market price). An implication of these conditions is that any good must be produced by a single monopolist and priced at average cost. Then, imposing full employment,

$$F_i + c_i Y_i = V_i,$$

we can immediately solve for the sectoral output:

$$Y_i = \frac{V_i - F_i}{c_i}. \tag{8}$$

Analogous conditions apply to sector j. Substituting equation (8) (and the analog for sector j) into equation (3) and recalling that $\chi = w_i V_i / w_j V_j$, we can express the relative factor rewards as:

$$\frac{w_i}{w_j} = \frac{\gamma}{1-\gamma}\left(\frac{V_j}{V_i}\right)^{1/\epsilon}\left(\frac{c_j}{c_i}\cdot\frac{1 - F_i/V_i}{1 - F_j/V_j}\right)^{1-\frac{1}{\epsilon}}. \tag{9}$$

Intuitively, the relative price of factor i is higher the higher the relative importance of the i-intensive good in consumption, as captured by γ. Further, when $\epsilon > 1$, relative rewards are decreasing in relative marginal costs (c_i/c_j). In fact, with an elasticity of substitution in consumption greater than 1, a higher relative marginal cost raises the relative price of the final good and reduces its expenditure share, because consumers demand the cheaper good more than proportionally. Finally, the term $(V_j/V_i)^{1/\epsilon}$ captures the standard scarcity effect: ceteris paribus, the relative price of a factor is higher the lower its relative supply.

More interestingly, from equation (9) it is easy to see that whenever goods are gross substitutes (that is, whenever $\epsilon > 1$), an increase in scale that leaves the relative endowment unchanged raises the relative price of factor i as long as

$$\frac{F_i}{V_i} > \frac{F_j}{V_j}. \tag{10}$$

The opposite is true when final goods are gross complements (that is, $\epsilon < 1$). The relative factor reward is always scale-invariant if and only if $\epsilon = 1$.

The reason for this result is the following: The presence of fixed costs introduces firm-level increasing returns that fall with output. With only one firm in each sector, the same increasing returns apply at the sectoral level. From equation (8) the scale elasticity of output is easily computed:

$$e_s^{Y_i} = \frac{1}{1 - F_i/V_i},$$

which is greater than 1 and decreasing in V_i/F_i. Note that in the simplest model of contestable markets, there are no other determinants of market power, and increasing returns thus depend only on endowments and technology (V_i and F_i). Next we will see that in more general models market power and increasing returns also depend on demand parameters.

B. *Quantity Competition*

We consider now a model with product differentiation that includes some elements of strategic interaction: firms producing the same good compete in quantities, taking each other's output as given. As shown by Kreps and Scheinkman (1983), under mild conditions quantity competition can also be interpreted as the outcome of a two-stage model of capacity choice followed by price competition. Because Y_i represents the output of a large macro sector, we think it is realistic to assume that individual firms cannot affect its price. Therefore, we view goods Y_i and Y_j as produced by perfectly competitive firms assembling, at no cost, own-industry differentiated intermediate goods.[5] In particular, we assume that each sector contains a continuum of intermediates of measure 1 and that the production functions for final goods take the following CES form:

$$Y_i = \left(\int_0^1 Y_i(\upsilon)^{\frac{\sigma_i - 1}{\sigma_i}} d\upsilon\right)^{\frac{\sigma_i}{\sigma_i - 1}}, \tag{11}$$

where $Y_i(\upsilon)$ is the total amount of the intermediate good type υ used in the production of good i, and $\sigma_i > 1$ is the elasticity of substitution between any two varieties of intermediates used in sector i. The price for final good Y_i (equal to the average cost) implied by equation (11) is

$$P_i = \left(\int_0^1 p_i(\upsilon)^{1-\sigma_i} d\upsilon\right)^{1/(1-\sigma_i)}, \tag{12}$$

where $p_i(\upsilon)$ is the price of the intermediate good type υ used in the production of good i.

[5] Equivalently, Y_i and Y_j can be interpreted as consumption baskets of i- and j-intensive goods.

Imperfectly competitive firms operate at the more disaggregated level of intermediate industries. Each intermediate υ is a homogeneous good produced by a finite number $n_i(\upsilon)$ of symmetric firms engaging in Cournot competition. Again, the production of each intermediate υ in sector i involves a fixed requirement F_i and a constant marginal requirement c_i so that the total cost function for a producer of variety υ in sector i is still given by equation (7). Profit maximization by intermediate firms, taking the output of other competitors as given, implies the following pricing rule:

$$p_i(\upsilon) = p_i = \left(1 - \frac{1}{\sigma_i n_i(\upsilon)}\right)^{-1} c_i w_i, \tag{13}$$

where the markup depends on the number of competing firms. A free-entry condition in each industry producing any variety v implies zero profits in equilibrium (up to the integer problem):

$$\pi_i(\upsilon) = \left(\frac{c_i y_i(\upsilon)}{\sigma_i n_i(\upsilon) - 1} - F_i\right) w_i = 0.$$

Full employment requires

$$[F_i + c_i y_i(\upsilon)] n_i(\upsilon) = V_i,$$

where V_i is the supply of factor i. Using this condition together with the free-entry condition yields the equilibrium number of firms in each industry and the output produced by each of them:

$$n_i(\upsilon) = n_i = \left(\frac{V_i}{F_i \sigma_i}\right)^{1/2}, \tag{14}$$

$$y_i(\upsilon) = y_i = \frac{1}{c_i}\,[(V_i F_i \sigma_i)^{1/2} - F_i]. \tag{15}$$

Note that a scale increase (that is, an increase in V_i) is associated with a rise in firms' output. This is a direct consequence of the procompetitive effect of a market size expansion, which reduces price–marginal-cost markups and forces firms to increase output to cover fixed costs.

Finally, note that symmetry implies

$$Y_i = Y_i(\upsilon) = n_i y_i, \qquad P_i = p_i(\upsilon) = p_i. \tag{16}$$

The same conditions apply to sector j. The relative factor reward can be found by substituting equations (14), (15), and (16) and the analogous conditions for sector j into equation (3) and recalling that $\chi = w_i V_i / w_j V_j$:

$$\frac{w_i}{w_j} = \frac{\gamma}{1-\gamma}\left(\frac{V_j}{V_i}\right)^{1/\epsilon}\left(\frac{c_j}{c_i} \cdot \frac{1 - (F_i/V_i\sigma_i)^{1/2}}{1 - (F_j/V_j\sigma_j)^{1/2}}\right)^{\frac{\epsilon-1}{\epsilon}}. \tag{17}$$

Equation (17) is almost identical to equation (9). In particular, the relative factor price w_i/w_j depends on the basic parameters ϵ, γ, c_i/c_j, and V_i/V_j as in the previous model. The only notable difference is in the condition for the factor bias of scale: Under Cournot competition, if final goods are gross substitutes (that is, $\epsilon > 1$), then an increase in scale that leaves the relative endowment unchanged raises the relative price of factor i as long as

$$\frac{F_i}{V_i\sigma_i} > \frac{F_j}{V_j\sigma_j}. \tag{18}$$

Again, the reverse is true when final goods are gross substitutes (that is, $\epsilon < 1$), and relative factor rewards are always scale-invariant if and only if $\epsilon = 1$. In contrast with equation (10), the new condition shows that product differentiation (or, equivalently, the elasticity of substitution between varieties within a single sector) also matters for the factor bias of scale: factors used intensively in the production of more differentiated products (low σ_i) tend to benefit more from a market size increase.

The difference from the previous case is easily explained. As before, in equilibrium sectoral increasing returns are proportional to markups. In fact, using equations (13) and (14), it is possible to see that equation (18) holds whenever the markup is higher in the i-intensive sector. However, whereas under contestable markets markups are determined uniquely by technological factors (the ratio of fixed costs to endowments), now they also depend on demand conditions: when a sector produces varieties that are highly substitutable, firms cannot charge high prices, which in turn implies that markups and increasing returns must be low in equilibrium.

The mechanism at work in this model is similar to the one we discussed before. The procompetitive effect implies that firms' output grows with market size. For this reason, sectoral production functions exhibit increasing returns to scale that fall with rising market size, just like that of any single firm. Substituting equations (14) and (15) into (16) to derive an expression for sectoral production functions in terms of parameters, it is straightforward to show that the scale elasticity of sectoral output is

$$e_s^{Y_i} = \frac{1 - \frac{1}{2}(F_i/V_i\sigma_i)^{1/2}}{1 - (F_i/V_i\sigma_i)^{1/2}}, \tag{19}$$

which is greater than 1 and decreasing in $V_i\sigma_i/F_i$. Together with equation (6), equation (19) shows how variable increasing returns at the sectoral level determine the factor bias of scale.

C. *Price Competition*

We consider now the case of price competition with differentiated products, following the *ideal variety* approach of Salop (1979) and Lancaster (1979). We depart from the previous analysis in the choice of market structure within each differentiated industry.

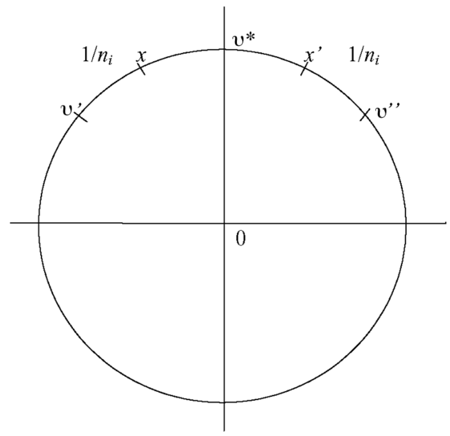
It will prove convenient to work with the limit case where intermediate goods $Y_i(\upsilon)$ are not substitutable, that is, $\lim \sigma_i \to 0$, so that equation (11) becomes a Leontief function. This assumption is not just for analytical convenience, but also to keep the analysis as close to the previous setup as possible. In fact, in the Cournot case firms within the same intermediate industry were producing a homogeneous good; therefore, to study how product differentiation, with its effect on competition, influences the relationship between scale and factor rewards, we needed the parameter σ_i, capturing an exogenous component of firms' market power coming from product (or demand) characteristics specific to each sector. In this section, instead, we use a model where product differentiation arises *within* each intermediate industry and we do not need to study additional effects of product differentiation *between* intermediate industries. Therefore, we simplify the interdependence of intermediate industries by assuming that all the varieties in equation (11) are demanded in the same amount. Our results do not depend on this assumption.

Within each intermediate industry producing $y_i(\upsilon)$ there is a continuum of potential types, and we imagine a one-to-one correspondence between these types and the points on the circumference (of unit length) of a circle, which represents the product space. Competitive firms buying intermediates to assemble the bundle Y_i have preferences over these types; in particular, we assume that each buyer has an ideal type, represented by a specific point on the circle. In order to assemble the final good using a type other than the most preferred, a firm incurs an additional cost that is higher the further away the intermediate is from the ideal type. We model this cost of distance in the product space as a standard iceberg transportation cost: one unit of a type located at arc distance x from the ideal one is equivalent to only e^{-xd_i} units of ideal type. Therefore, if $p_i(\upsilon)$ is the price of the ideal type, the price of an equivalent unit bought by a firm located at distance x will be $p_i(\upsilon)e^{xd_i}$. Note that the function e^{xd_i}, Lancaster's compensation function, parametrizes the degree of product differentiation. As $d_i \to 0$, different types become perfect substitutes, as nobody would be willing to pay any extra cost to buy a specific type of good. We will see shortly that d_i plays in this context the same role as $1/\sigma_i$ in the previous model.

We restrict again the analysis to symmetric equilibria, so that all firms in the same sector set the same price p_i. In particular, we assume that preferences of buyers over different types are uniformly distributed at random on each circular product space. We also assume that sellers are located equidistant from one another on each circle.[6] Given that there is a continuum [0, 1] of these circles, by the law of large numbers every buyer faces the same unit cost of producing good Y_i. Assuming that $n_i(\upsilon)$ firms have entered the market for $y_i(\upsilon)$, we can calculate demand for each firm as follows. Suppose that firm υ^*, represented graphically in figure 1 as a point on the product space of industry υ, sets a price $p_i(\upsilon)$ for its type. A buyer whose ideal type is located at distance $x \in (0, 1/n_i(\upsilon))$ from υ^* is indifferent between purchasing from firm υ^* and purchasing from its closest neighbor on the circle υ' if

$$p_i(\upsilon^*)e^{xd_i} = p_i(\upsilon')e^{d_i[1/n_i(\upsilon)-x]}. \tag{20}$$

FIGURE 1.—PRICE COMPETITION WITH DIFFERENTIATED PRODUCTS



Therefore, given the prices $p_i(\upsilon)$, equation (20) implicitly defines the market width for any single firm: all buyers whose ideal type is within the arc distance x from type υ^* are customers of firm υ^*. Note that, in general, an increase in the price set by a firm will have two effects. First, as shown in equation (20), it reduces the measure of customers who buy that type, and second, it reduces the quantity demanded by the remaining customers. The Leontief assumption cancels the second effect, so that demand for each firm can be derived from (20) as

$$D_i(\upsilon) = 2x\bar{Y}_i = \left[\frac{1}{n_i(\upsilon)} + \log\left(\frac{p_i(\upsilon')}{p_i(\upsilon)}\right)^{1/d_i}\right]\bar{Y}_i, \tag{21}$$

where $\bar{Y}_i$ is the aggregate sectoral output that would be produced if optimal types were always used. Profit maximization given equation (21) and the already introduced cost function (7) yield a familiar pricing formula (after imposing symmetry):

$$p_i(\upsilon) = p_i = \left(1 - \frac{d_i}{n_i(\upsilon)}\right)^{-1} c_i w_i. \tag{22}$$

Note that, as in the Cournot case, the markup over marginal cost decreases with $n_i(\upsilon)$. Moreover, setting $d_i = 1/\sigma_i$, equation (22) reduces exactly to equation (13). Hence, in our specification, a firm's behavior under Cournot competition within differentiated industries is isomorphic to that under price competition with differentiated products in each

[6] It can be shown that this is indeed optimal. The reason is that a firm that tries to change its location slightly loses on one side of its market the same number of customers that it gains on the other side. Hence, small changes in location do not alter the quantity demanded.

industry. The rest of the analysis is also similar. In particular, free entry and market clearing still apply, so that the equilibrium number of firms and their output are given by equations (14) and (15) after substituting $d_i = 1/\sigma_i$. This immediately implies that in both models markups and internal increasing returns depend on scale in exactly the same way.

However, there is an important difference in how production of intermediates, $n_i(\upsilon)y_i(\upsilon)$, translates into output of the final good Y_i and thus [from equation (3)], into the price of factors. In the Cournot case, the benefit of having a larger number of firms lies in the procompetitive effect and therefore in a better exploitation of scale economies that are internal to firms, whereas now there is an additional benefit of scale in that buyers will be on average closer to their ideal type. This is a source of increasing returns at sectoral level. To see this, note that output of final goods, Y_i, equals the total amount of intermediates produced in any industry υ, $\bar{Y}_i = n_i(\upsilon)y_i(\upsilon)$, less the cost of the mean distance from the ideal type:

$$Y_i = \frac{\bar{Y}_i}{2n_i \int_0^{1/2n_i} e^{xd_i}\, dx} = \frac{\bar{Y}_i}{\frac{2n_i}{d_i}\,(e^{d_i/2n_i} - 1)}. \tag{23}$$

Given that $\lim_{d_i/2n_i \to 0} [(2n_i/d_i)\,(e^{d_i/2n_i} - 1)] = 1$, one can see that final output in sector i grows to $\bar{Y}_i$ as the number of available types grows to infinity ($n_i \to \infty$) or types become perfect substitutes ($d_i \to 0$). Given that the additional effect depends on n_i/d_i just like markups, external and internal increasing returns share the same determinants and the new mechanism simply reinforces the scale effect found in the Cournot case. In fact, setting $d_i = 1/\sigma_i$ and using equation (3), we can derive

$$\frac{w_i}{w_j} = \left(\frac{w_i}{w_j}\right)^c \left[\left(\frac{V_j\sigma_j/F_j}{V_i\sigma_i/F_i}\right)^{1/2} \frac{\exp[(4V_j\sigma_j/F_j)^{-1/2}] - 1}{\exp[(4V_i\sigma_i/F_i)^{-1/2}] - 1}\right]^{\frac{\epsilon-1}{\epsilon}},$$

where $(w_i/w_j)^c$ is the relative reward in equation (17). Simple inspection reveals that the condition for the factor bias of scale is identical to the previous case (18).

This third case has illustrated an additional reason why scale can be biased: asymmetries in increasing returns that are *external* to firms and arise from a preference for variety in aggregate. This new effect does not alter our previous conclusions, because it depends on the elasticity of substitution between varieties and the number of firms, just like market power.

D. *Dixit-Stiglitz Monopolistic Competition*

We now briefly consider the Dixit-Stiglitz (1977) model of monopolistic competition, a widely used model and a prominent toolbox in the new trade theory.[7] The purpose of this section is to show that, although we will be able to find a simple condition for scale to be factor-biased consistent with that of previous models, the mechanism behind it differs in an important respect, for it does not rely on the procompetitive effect emphasized so far. As we will see, in this model firm size and markups are fixed exogenously, and the factor bias of scale only depends on asymmetries in external increasing returns due to the variety effect discussed at the end of section II C.

Consider the Cournot model of section II B, and allow the range of varieties produced in each sector, n_i and n_j, to vary (previously it was confined to the unit interval):

$$Y_i = \left[\int_0^{n_i} y_i(\upsilon)^{\frac{\sigma_i - 1}{\sigma_i}}\, d\upsilon\right]^{\frac{\sigma_i}{\sigma_i - 1}}, \tag{24}$$

We assume a potentially infinite measure of producible varieties. Thus, the fixed costs in equation (7) assure that no two firms will find it profitable to produce the same variety and each will be sold by a monopolist. Then, the pricing rule (13) simplifies to

$$p_i(\upsilon) = p_i = \left(1 - \frac{1}{\sigma_i}\right)^{-1} c_i w_i,$$

showing that the markup is now constant and only depends on σ_i. The measure of firms in each sector is determined endogenously by a free-entry condition: new firms (and thus varieties) are created up to the point where profits are driven to zero. Imposing $\pi_i = 0$ yields the scale of production for each firm:

$$y_i(\upsilon) = y_i = \frac{F_i(\sigma_i - 1)}{c_i}. \tag{25}$$

Combined with full employment, $(F_i + c_i y_i)n_i = V_i$, equation (25) gives the equilibrium measure of firms in each sector:

$$n_i = \frac{V_i}{\sigma_i F_i}, \tag{26}$$

which completes the characterization of the equilibrium.

The relative factor reward can be found by substituting n_i (26) and y_i (25) into Y_i (24), and using this and the analogous equation for sector j into equation (3). Recalling that $\chi = w_iV_i/w_jV_j$, we obtain

$$\frac{w_i}{w_j} = \frac{\gamma}{1-\gamma}\frac{V_j}{V_i}\left[\frac{\left(\frac{V_i}{\sigma_i F_i}\right)^{\frac{\sigma_i}{\sigma_i - 1}} \frac{F_i(\sigma_i - 1)}{c_i}}{\left(\frac{V_j}{\sigma_j F_j}\right)^{\frac{\sigma_j}{\sigma_j - 1}} \frac{F_j(\sigma_j - 1)}{c_j}}\right]^{(\epsilon-1)/\epsilon}.$$

[7] This case is studied more in detail in Epifani and Gancia (2004).

Clearly, for $\epsilon > 1$ ($\epsilon < 1$), an increase in scale that leaves the relative endowment V_i/V_j unchanged raises the relative price of factor i as long as $\sigma_i < \sigma_j$ ($\sigma_i > \sigma_j$), whereas in the case $\epsilon = 1$ the relative factor reward is always scale-invariant. Given that the scale of production of each firm is fixed, an increase in market size does not allow one to better exploit scale economies at the firm level. A larger market only translates into a wider range of differentiated products, which is beneficial because the aggregate productivity in equation (24) grows with variety. Using equations (26) and (25) in (24), it is easy to show that the scale elasticity of sectoral output is

$$e_s^{Y_i} = \frac{\sigma_i}{\sigma_i - 1}.$$

Thus, asymmetries in returns to scale only depend on differences between σ_i and σ_j. Yet, even in this case, scale turns out to be biased in favor of the factor used intensively in the sector where markups are higher.

It should also be noted that a constant markup is usually seen as a limit of the otherwise convenient Dixit-Stiglitz formulation. This property is sometimes removed by assuming that demand becomes more elastic when the number of varieties increases, that is, $\sigma_i = f(n_i)$, with $f'(n_i) > 0$, as in Krugman (1979). Using equation (26), it is easy to see that the elasticity of substitution becomes an increasing function of the ratio of endowments to fixed costs: $\sigma_i f^{-1}(\sigma_i) = V_i/F_i$. In this case, the condition for the factor bias of scale reduces to equation (10), just as in the contestable markets model.

E. Discussion

We have shown how in models with increasing returns and imperfect competition the market size affects the income distribution across factors: whenever final goods are gross substitutes (gross complements), a scale expansion tends to raise (lower) the relative reward of the factor used intensively in the sector characterized by smaller factor employment, higher degree of product differentiation, and higher fixed costs. In this section, we pause to discuss some properties of our results, their implications, and the realism of the key assumptions on which they are built.

A first notable implication of the conditions (10) and (18) is that, in the absence of sectoral asymmetries in fixed costs ($F_i = F_j$) or in the degree of substitutability among varieties ($\sigma_i = \sigma_j$), *a scale expansion benefits the scarce factor in the economy.* Second, because the scale elasticity of sectoral outputs converges to 1 for V_i approaching infinity asymptotically, the factor bias of scale vanishes when the scale grows very large. However, this will be the case only once prices have become approximately equal to marginal costs in both sectors. On the contrary, when the endowment of a factor is very low, increasing returns may be so high that the reward of that factor actually rises with its supply. In other words, *the factor demand curve may be at first upward sloping.*[8] For example, in the model of quantity competition, it is easy to show from equation (17) that the relative reward of factor i, w_i/w_j, increases with its supply V_i as long as $(F_i/V_i\sigma_i)^{1/2} > 2/(\epsilon - 1)$. Clearly, this is possible only if goods are gross substitutes, and is more likely the higher is the elasticity of substitution ϵ.[9] Third, although the relative real marginal cost (c_i/c_j) and the bias in demand (γ) affect the level of the relative factor reward, they have no effect on its scale elasticity. The reason is that markups are independent of marginal costs, whereas the bias in consumption only shifts *income* between sectors, which is immaterial for the factor bias of scale, given homotheticity of technologies.

Our theory yields novel predictions on the distributional effects of international trade and technical progress. First, it suggests that factor-augmenting technical progress, by increasing the effective market size of an economy, will tend to increase the marginal product of factors used intensively in the least competitive sectors. Second, and perhaps more important, given that intra-industry trade between similar countries can be isomorphic to an increase in market size, our theory suggests which factor stands to gain more from it. In doing so, it fills a gap in the new trade theory, where the distributional implications of two-way trade in goods with similar factor intensity are usually overlooked (see, for example, Helpman & Krugman, 1985). More generally, because any form of trade entails an increase in the effective size of markets, the distributional mechanism discussed in this paper is likely to be always at work. In stark contrast with the standard Heckscher-Ohlin view, our results suggest that in some cases the *scarce* factor benefits the most from trade. Although the concept of scarcity we refer to is in absolute terms, and not relative to other countries as in the factor-proportions trade theory, it is nonetheless possible to build examples in which the factor bias of scale dominates the Stolper-Samuelson effect, so that the distributional implications of the standard trade theory are overturned.

To elaborate on this point, we now analyze the effects of trade integration between dissimilar countries on relative factor prices in the simplest model with contestable markets (the other models would yield very similar results). Free-trade factor prices can be found by substituting world endowments in equation (9) instead of domestic values.[10] Totally differentiating equation (9), we can decompose the

[8] The possibility of an upward-sloping demand curve due to the endogenous reaction of technology is also emphasized in Acemoglu (2002).

[9] The assumption of an inelastic supply of factors guarantees that our models will always have a unique equilibrium even with an upward-sloping factor demand curve. If both the supply and demand curves are upward sloping, multiple equilibria may arise. In the presence of a dynamic adjustment process, stability of an equilibrium will be an issue if the demand curve is steeper than the supply curve, that is, if factor endowments are very responsive to prices. In this case, short-run adjustment costs may guarantee local stability.

[10] This is true as long as factor price equalization (FPE) holds, which is guaranteed by the specific factor assumption. In more general models where both factors are employed in both sectors, we would require that countries' endowments not be too dissimilar [see Helpman and Krugman (1985) for a definition of the FPE set].

effects of trade integration, interpreted as an increase in both V_i and V_j, on w_i/w_j as follows:

$$\frac{w_i}{w_j} = -\frac{1}{\epsilon}(\hat{V}_i - \hat{V}_j) + \left(1 - \frac{1}{\epsilon}\right)\left(\frac{F_i}{V_i - F_i}\hat{V}_i - \frac{F_j}{V_j - F_j}\hat{V}_j\right), \tag{27}$$

where again a hat denotes a proportional variation. The first term on the right-hand side shows the impact on relative rewards of changes in the relative scarcity of factors, in the absence of any scale effects (that is, when F_i and F_j are very small and/or V_i and V_j very large): according to it, factors becoming relatively scarcer will see their relative price increase, as in the standard Hecksher-Ohlin-Samuelson theory. The second term, in contrast, is the scale effect, showing that when $\epsilon > 1$ scale tends to benefit more the "scarce" factor—or, more precisely, the factor with a lower ratio of employment to fixed cost. When trade integration takes place between identical countries, the relative scarcity does not change ($\hat{V}_i = \hat{V}_j$) and only the scale effect survives (unless $F_i/V_i = F_j/V_j$, in which case scale is immaterial too). Also, as $\epsilon \to \infty$, the relative scarcity effect disappears and the contribution of scale becomes stronger. Thus, if ϵ is high, a small country that is scarce in factor i relative to the rest of the world may experience an increase in the relative reward of factor i after trade opening, provided that factor i is also scarce in the absolute sense of the condition (10).

Finally, we briefly discuss the empirical support of the key assumptions at the root of our results. First, on the production side, we assumed firm-level scale economies that decrease with firm size, for they are generated by fixed costs. This is consistent with recent plant-level evidence. Tybout and Westbrook (1995) use plant-level manufacturing data for Mexico to show that most industries exhibit increasing returns to scale that typically decrease with larger plant sizes. Similarly, Tybout, De Melo, and Corbo (1991) and Krishna and Mitra (1998) find evidence of a reduction in returns to scale in manufacturing plants after trade liberalization in Chile and India, respectively.[11] Second, the market structure in our models involves variable markups. In this respect, the evidence is compelling. Country studies reported in Roberts and Tybout (1996), which use industry– and plant-level manufacturing data for Chile, Colombia, Mexico, Turkey, and Morocco, find that increased competition due to trade liberalization is associated with falling markups. Similar results using a different methodology are found, among others, by Levinsohn (1993) for Turkey, by Krishna and Mitra (1998) for India, and by Harrison (1994) for Côte d'Ivoire; Galí (1995) provides cross-country evidence that markups fall with increasing income. Further, business cycle studies show that markups tend to be countercyclical [see Rotemberg and Woodford (1999) for a survey], which is again consistent with our hypothesis.

[11] See also Tybout (2003) on this point.

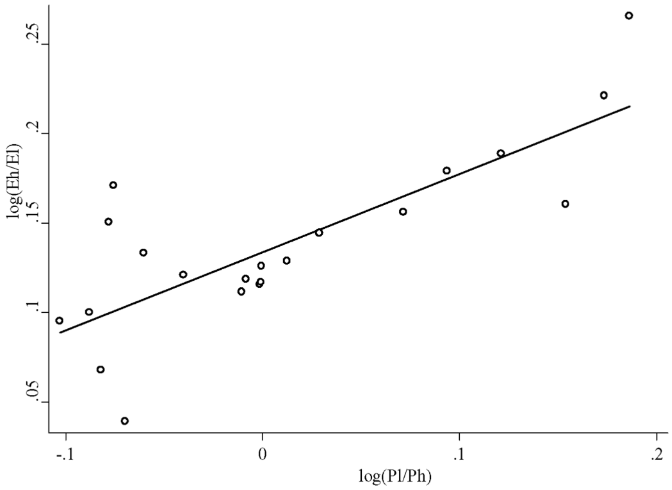
III. An Application to Wage Inequality

A prominent application of our results is in the debate over the causes of the rise in skill premia that has occurred since the early 1980s. The theoretical literature has identified three main culprits: skill-biased technical change, capital-skill complementarity, and international trade. Our theory suggests the existence of a neglected link among these explanations, namely, the skill bias of scale. In this section, we first discuss some available evidence supporting the empirical validity of the assumptions needed for an increase in market size to lead to a higher skill premium. Then, in section III A, we provide a test of the mechanism underlying the skill bias of scale according to our theory, which builds on the procompetitive effect. To this purpose, we use a large panel of U.S. industries to show that markups tend to be higher in skill-intensive sectors and fall with increasing scale, the more so the higher the skill intensity. Finally, in section III B, we briefly discuss the related literature on wage inequality.

If goods produced with different skill intensity are gross substitutes, the conditions (10) and (18) imply that scale is skill-biased when skilled workers are a minority in the total workforce, use technologies with relatively high fixed costs, and produce highly differentiated goods. All these conditions are likely to be met in the real world. As for the latter two, note that skill-intensive productions often involve complex activities, such as R&D and marketing, that raise both fixed costs and the degree of product differentiation. Regarding the share of skilled workers in the total workforce, we can refer to the Barro-Lee database to make a crude cross-country comparison. Identifying skilled workers as those with college education (as in a large part of the empirical literature), we find that in 2000 the percentage of skilled workers ranged from a minimum of 0.1% in Gambia to a maximum of 30.3% in the United States, with New Zealand ranking second with a share of only 16%.

Further, and most important, the model's prediction of sectoral asymmetries in the scale elasticity of output finds support in two recent empirical studies. Antweiler and Trefler (2002), using international trade data for 71 countries and five years, find that skill-intensive sectors, such as petroleum refineries and coal products, Pharmaceuticals, electric and electronic machinery, and nonelectrical machinery, have an average scale elasticity around 1.2, whereas traditional non-skill-intensity sectors, such as apparel, leather, footwear, and food, are characterized by constant returns. Using a different methodology, Morrison Paul and Siegel (1999) estimate returns to scale in U.S. manufacturing industries for the period 1979–1989. Their estimates of

FIGURE 2.—ELASTICITY OF SUBSTITUTION BETWEEN LOW- AND HIGH-SKILL-INTENSITY GOODS



Source: Epifani and Gancia (2004).

sectoral scale economies are strongly positively correlated with the sectoral skill intensity.[12]

For these asymmetries to be consistent with a rise in the skill premium, we also need the elasticity of substitution (between goods produced with different factor intensities) to be greater than 1. In Epifani and Gancia (2004), we show that in the years from 1980 to 2000 the relative expenditure on skill-intensive goods in the United States increased by more than 25%, while the relative price of traditional, low-skill-intensity goods increased by more than 25%, a result broadly consistent with most of the studies on product prices surveyed in Slaughter (2000). In figure 2 we plot the relationship between the log relative expenditure on modern goods, $\log(E_h/E_t)$, and the log relative price of traditional goods, $\log(P_t/P_h)$. The slope coefficient and standard error of the regression line in the figure are 0.44 and 0.08, respectively, with an R-squared of 0.62. The estimated coefficient implies an elasticity of substitution of 1.44, consistent with our assumption.[13] Moreover, indirect evidence also suggests that the elasticity of substitution between low- and high-skill-intensity goods is significantly greater than 1. In particular, in our model the *aggregate* elasticity of substitution in production between skilled and unskilled workers is equivalent to the elasticity of substitution in consumption between goods with low and high skill-intensity. Several studies provide estimates of the former parameter, and most of them are above 1.[14]

Finally, by predicting that the aggregate relative demand for skilled workers is increasing in size, our model provides an explanation for the empirical finding by Antweiler and Trefler (2002) that a 1% scale increase brings about a 0.42% increase in the relative demand for skilled workers. Evidence of skill-biased scale effects is also found by Denny and Fuss (1983) in their study of the telecommunication industry, and by Berman, Bond, and Griliches (1994), Autor, Katz, and Krueger (1998), and Feenstra and Hanson (1999), who all find that skill upgrading is positively and significantly associated with variation in industry size in their studies of wage inequality in the United States.[15]

A. *The Procompetitive Effect of Scale: Evidence from U.S. Industries*

We now provide evidence on the mechanism underlying the skill bias of scale according to our theory. For scale to be skill-biased, our theory requires that the following conditions be satisfied: (a) markups must be higher in the skill-intensive industries; (b) a rise in the size of an industry must bring about a procompetitive effect, which reduces markups; (c) the procompetitive effect must be stronger in the skill-intensive industries. If these conditions are met, a scale increase raises the relative demand for skilled workers provided that the elasticity of substitution between goods with low and high skill-intensity is also greater than 1. This suggests the following simple (and yet demanding) test: upon observing a panel of industry-level data on markups MK_{it}, skill intensity $(H/L)_{it}$, and industry size Y_{it}, we may run the following regression:

$$MK_{it} = \beta_0 Y_{it} + \beta_1 (H/L)_{it} + \beta_2 (H/L)_{it} \cdot Y_{it} + \eta_i + d_t + X'_{it}\alpha + \epsilon_{it}, \tag{28}$$

where i and t index industries and time, respectively, η_i and d_t are industry and time fixed effects, X_{it} is a vector of controls, and ϵ_{it} is a random disturbance. The procompetitive effect of a scale expansion suggests that the expected sign of β_0 is negative; the expected sign of β_1 is instead positive, because our theory implies that markups are higher in the skill-intensive industries. Finally, the interaction term in $(H/L)_{it} \cdot Y_{it}$ allows us to test whether the procompetitive effect of scale expansion is stronger in the skill-intensive industries and the expected sign of β_2 is negative.

[12] See also Epifani and Gancia (2004) on this point.

[13] On controlling for the log of per capita GDP, the coefficient of the relative price is slightly reduced (0.36), but is still significant at the 7% level (with a standard error of 0.19). In contrast, the per capita GDP coefficient is positive (0.02), as expected, but small and imprecisely estimated (its standard error equals 0.05).

[14] Freeman (1986) suggests a value of the elasticity of substitution between more and less educated labor in the range between 1 and 2. Hamermesh and Grant (1979) find a mean estimate of 2.3. Krusell et al. (2000) and Katz and Murphy (1992) report estimates for the U.S. economy of 1.67 and 1.41, respectively.

[15] When our model is interpreted as describing a single sector, it can easily explain the positive association between skill upgrading and variation in industry size. Assume, in particular, that sector i is made of two subsectors, h and l, the former using only skilled workers and the latter only unskilled workers. Then equation (3) implies that in this sector the relative income share of skilled workers, $\chi(h)_i = w_h H_i / w_l L_i$, is proportional to the relative output of the two subsectors, Y_{hi}/Y_{li}. Hence, under the assumptions discussed earlier ($\epsilon > 1$ and a higher markup in subsector h), an increase in the size of sector i brings about an increase in Y_{hi}/Y_{li} and in the relative income share of skilled workers.

Unfortunately, data on industry-level markups are not readily available, because prices and marginal costs are rarely observed. To circumvent this problem, two main approaches can be followed. One is to estimate markups from a structural regression à la Hall (1988). For our purposes, one problem with this approach is that, to estimate markups across industries or over time, either the time or the industry dimension has to be sacrificed, which means that markups have to be assumed constant over time or across industries. In contrast, the test of our theory requires markups to vary both across industries and over time.

Alternatively, markups can be constructed using data on industry sales and total costs. This is the approach advocated by Tybout (2003) and widely used in the empirical literature on the procompetitive effect of trade liberalization in the developing world (see, for example, Roberts & Tybout, 1996). Here, we follow this methodology and use price-cost margins as a proxy for industry markups. Constructed markups have in fact the advantage of being variable both across industries and over time.

We apply our test to data from the NBER Productivity Database by Bartelsman and Gray. As far as we know, this is the most comprehensive and highest-quality database on industry-level inputs and outputs, covering approximately 450 U.S. manufacturing industries at the four-digit SIC level for the period between 1958 and 1996. Moreover, the NBER file has been widely used to investigate the determinants of the recent rise in U.S. wage inequality.[16] Here, we show a novel way of exploiting information in this data set to uncover a potentially relevant mechanism underlying the evolution of wage inequality in the United States.

In our benchmark specifications, price-cost margins are computed as the value of shipments (adjusted for inventory change) less the cost of labor,[17] materials, and energy, divided by the value of shipments. As a proxy for industry size we use the real value of shipments. Finally, following a standard practice in the empirical literature on wage inequality, we proxy skilled workers with nonproduction workers, and therefore our measure of skill intensity is the ratio of nonproduction to production workers. Consistent with our model, this measure of skill intensity is positively correlated with price-cost margins: the simple correlation between the two variables equals 0.3.

Capital-intensive industries generally require higher price-cost margins to cover the cost of capital. Following Roberts and Tybout (1996), we therefore control for the capital-output ratio, $(K/PY)_{it}$. Note also that our definition of price-cost margins implicitly assumes that all capital expenditures constitute fixed costs, which is not, in general, true. To address this problem, we recompute the price-cost margins by directly netting out capital expenditures. We find that our results are independent of whether we treat capital expenditures as a fixed or a variable cost.

We add more controls to the specification of equation (28) in order to isolate the procompetitive effect of scale expansion from other relevant sources of variation in price-cost margins. In particular, if entry is less than perfect in the short run, an increase in industry profitability would both stimulate entry and raise the price-cost margin, thereby inducing a positive association between industry size and price-cost margins. Hence, in order to better isolate the effect of exogenous variation in industry size on the price-cost margins, we control for industry profitability.[18] A way to do this is by using the index of total factor productivity (TFP5) reported in the NBER file. However, this control is likely to be endogenous and may induce a bias in the estimation of our coefficients of interest. Therefore, in order to address the endogeneity bias due to reverse causation between industry size and price-cost margins, we also estimate equation (28) by instrumental variables. Interestingly, the two procedures lead to similar results.

Finally, we include two controls related to import competition. Our model shares with other models the standard implication that foreign competition reduces markups. To capture this effect, we use the ratio of imports to the value of shipments, $(M/PY)_{it}$, as a proxy for the intensity of foreign competition. Our data on U.S. imports by four-digit SIC industry (1972 basis) for the the period from 1958 to 1994 come from the NBER Trade Database by Feenstra. Our model also suggests that the procompetitive effect of foreign competition is stronger in the skill-intensive industries. To capture this effect, we also include the interaction term in $(H/L)_{it} \cdot (M/PY)_{it}$, whose coefficient is therefore expected to be negative.

Our first set of results is reported in table 1. Here, we estimate various specifications of equation (28) by using the fixed-effects within estimator. The dependent variable is the price-cost margin gross of capital expenditures. We always include time dummies to avoid spurious results due to correlation of our covariates with time effects. In column (1), we estimate our baseline regression without controls. Note that the coefficients of the skill intensity and of the interaction term between skill-intensity and industry size have the expected sign and are significant at the 1% level. In contrast, the coefficient of industry size is significant but wrongly signed, suggesting that an expansion in industry size is associated with a rise in price-cost margins. As mentioned earlier, this result is not surprising, for an increase in profitability should stimulate entry, thereby increasing the size of an industry together with the price-cost margin. Therefore, without controlling for variation in in-

[16] See, in particular, Berman et al. (1994), Autor et al. (1998), and Feenstra and Hanson (1999).

[17] We would ideally want to disentangle the fixed from the variable cost of labor, instead of lumping them together in the overall cost of labor. Note, however, that the cost of labor reported in the NBER file does not include the wages of employees in headquarters and support facilities, which represent a relevant share of the overall fixed cost of labor and accounted for more than 10% of total payroll in manufacturing in 1986.

[18] See also Roberts and Tybout (1996) and Hoekman, Kee, and Olarreaga (2004) on this point.

TABLE 1.—PROCOMPETITIVE EFFECT OF SCALE EXPANSION (FIXED EFFECTS)

Indep. variable	(1)	(2)	(3)	(4)	(5)
	FE	FE	FE	FE	RE
Y_{it}	.012**	−.067***	−.089***	−.084***	−.060***
	(.005)	(.006)	(.006)	(.007)	(.006)
$(H/L)_{it}$	.135***	.121***	.125***	.086***	.080***
	(.027)	(.026)	(.026)	(.030)	(.029)
$Y_{it} \cdot (H/L)_{it}$	−.017***	−.016***	−.016***	−.017***	−.013***
	(.004)	(.004)	(.004)	(.004)	(.004)
TFP_{it}		.305***	.299***	.301***	.278***
		(.011)	(.011)	(.012)	(.012)
$(K/PY)_{it}$			−.060***	−.068***	−.050***
			(.006)	(.006)	(.006)
$(M/PY)_{it}$				−.014***	−.016***
				(.003)	(.003)
$(H/L)_{it} \cdot (M/PY)_{it}$				−.014***	−.015***
				(.002)	(.002)
Time dummies	Yes	Yes	Yes	Yes	Yes
Observations	16981	16981	16981	15395	15395
Groups	448	448	448	433	433
R-squared	.18	.22	.22	.21	.21

Dependent variable: price-cost margin (MK_{it}). Notes: All variables in logs. Standard errors in parentheses. ***, **, * = significant at the 1%, 5%, and 10% levels, respectively. Estimation is by fixed effects (within) in columns (1)–(4), and by random effects in column (5). Coefficients of time dummies not reported. Data sources: NBER Productivity Database (by Bartelsman and Gray) and NBER Trade Database (by Feenstra).

dustry profitability, the coefficient β_0 would be upward biased and the procompetitive effect of scale expansion would be underestimated. Indeed, as shown in column (2), when using the TFP index to control for variation in industry profitability, the negative effect of industry size on price-cost margins is restored and is significant beyond the 1% level. Note, also, that the coefficient of TFP is highly significant and large in magnitude, and that the coefficients of the other explicatives have the expected sign and are also significant at the 1% level.

In column (3), we also control for the capital-output ratio, whose coefficient is significant but wrong signed.[19] The coefficients of our main variables are the same order of magnitude and are significant at the 1% level. Finally, in column (4) we add the two covariates that control for the effects of foreign competition on price-cost margins: import penetration (the ratio of imports to the value of shipments) and the interaction term between skill intensity and import penetration. As expected, the coefficients of both variables are negative and significant at the 1% level, and the other coefficients are unaffected. This suggests that import competition reduces markups and that this effect is stronger in the skill-intensive industries.

Because the within estimator uses only temporal variation to estimate the coefficients, in column (5) we complement our analysis by rerunning our previous specification using the random-effects estimator. Note that the coefficients of all covariates (except for the capital-output ratio) have the expected sign and are significant beyond the 1% level. They are also similar to those estimated by fixed effects, which suggests that using also sectional variation to estimate the coefficients does not much affect the results.[20]

Although the results reported in table 1 represent an interesting test of our theory, they leave some important methodological problems unsolved. In particular, though fixed-effects regressions remedy the endogeneity problems that can be traced to the unobservable time-invariant industry heterogeneity, they do not address the simultaneity bias due to mutual interaction between the left- and right-side variables, and in particular between the price-cost margins and industry size. Therefore, we rerun various specifications of equation (28) by using instrumental variables. Table 2 reports the results of the fixed-effects instrumental variables estimation. In all specifications, we instrument all right-side variables using their lagged values as instruments. The choice of the lag structure of instruments is dictated by the Sargan test of overidentifying restrictions. In particular, the test always rejects the null hypothesis of instruments' validity when using close lags of the endogenous covariates as instruments. Some experimentation suggests, however, that the 5th-to-7th lag (or the 6th-to-8th lag) turn out to be appropriate instruments for most endogenous covariates in all specifications.[21] As shown by the p-value of the Sargan test in the bottom line of the upper part of table 2, the exogeneity of these instruments is never rejected.

Using distant lags of endogenous covariates as instruments raises a concern about weak instruments, in which case estimation by instrumental variables would be biased in the same direction as estimation by least squares. There-

[19] A negative coefficient of the capital-output ratio in fixed-effects regressions of the price-cost margins is recurrent in the empirical literature (see, for example, Roberts & Tybout, 1996). As shown in table 2, this anomaly disappears when using instrumental variables.

[20] However, Hausman's specification test strongly suggests that treating unobservable industry heterogeneity as random may lead to misspecification.

[21] Exceptions are the TFP and the real value of shipments, for which we generally use more distant lags.

TABLE 2.—PROCOMPETITIVE EFFECT OF SCALE EXPANSION (IV)

Indep. Variable	(1)	(2)	(3)	(4)
Y_{it}	−.066***	−.063***	−.075***	−.111***
	(.014)	(.014)	(.019)	(.021)
$(H/L)_{it}$	.312***	.260***	.379***	.306**
	(.106)	(.097)	(.128)	(.150)
$Y_{it} \cdot (H/L)_{it}$	−.022**	−.020**	−.032**	−.030**
	(.011)	(.010)	(.014)	(.015)
$(K/PY)_{it}$		.093***	.119***	.119***
		(.021)	(.025)	(.026)
$(M/PY)_{it}$			−.032**	−.053***
			(.015)	(.017)
$(H/L)_{it} \cdot (M/PY)_{it}$			−.001	−.010
			(.008)	(.009)
TFP_{it}				.234***
				(.066)
Time dummies	Yes	Yes	Yes	Yes
P-value Sargan test	.650	.854	.833	.131
F-Statistics of Excluded Instruments in First-Stage Regressions				
Y_{it}	515	510	262	183
$(H/L)_{it}$	102	108	54	33
$Y_{it} \cdot (H/L)_{it}$	190	189	94	60
$(K/PY)_{it}$		221	119	75
$(M/PY)_{it}$			100	62
$(H/L)_{it} \cdot (M/PY)_{it}$			172	111
TFP_{it}				63
Observations	12057	12056	10662	9820
Groups	448	448	433	433
R-squared	.07	.06	.05	.03

Dependent variable: price-cost margin (MK_{it}). Notes: All variables in logs. Standard errors in parentheses. ***, **, * = significant at the 1%, 5% and 10% levels, respectively. Coefficients of time dummies not reported. Estimation is by fixed-effects (within) instrumental variables. All right-side variables are treated as endogenous, using their lagged values as instruments. Time dummies are always used as additional instruments. The bottom half of the table reports the *F*-statistics for the null that excluded instruments do not enter first-stage regressions. Data sources: NBER Productivity Database (by Bartelsman and Gray) and NBER Trade Database (by Feenstra).

fore, in the bottom part of table 2 we report the *F*-statistics for the null hypothesis that excluded instruments are jointly insignificant in the first-stage regressions. Note that in all first-stage regressions the *F*-statistic of the excluded instruments is very high, suggesting that our instruments are not weak.[22]

Bearing in mind that the above tests of instruments validity raise the confidence in our instrumental variables estimates, we can now comment the main results in table 2. In column (1), we estimate equation (28) without controls. Note that the coefficients of all variables suggested by our theory have the expected sign and are highly significant. Moreover, columns (2) to (4) show that adding controls to our baseline regression does not affect the main results. It is remarkable, in particular, that even without controlling for TFP [columns (1)–(3)], the coefficient of industry size, which captures the procompetitive effect of scale expansion, is always negative and highly significant, just as in columns (2)–(4) of table 1, where we do control for TFP. This suggests that controlling for TFP in a *non*-IV regression washes out much of the simultaneity bias due to mutual interaction between price-cost margins and industry size. Finally, note that the coefficient of the capital-output ratio is now positive and highly significant, as expected.

Finally, we test for the robustness of our results with respect to capital costs. In particular, we recompute the price-cost margins by directly netting out capital expenditures, defined as $(r_t + \delta)K_{i\ t-1}$, where $K_{i\ t-1}$ is capital stock, r_t is the real interest rate, and δ is the depreciation rate. Data on U.S. real interest rates come from the World Bank *World Development Indicators.*[23] For the depreciation rate δ, we choose a value equal to 7%, implying that capital expenditures equal, on average, roughly 10% of the capital stock.[24] The main results are reported in table 3, where we rerun various specifications of equation (28), by fixed effects in columns (1)–(3), and by instrumental variables in columns (4)–(7). Note that the pattern of coefficients is very similar to that shown in previous tables. In particular, as in table 1, adding controls to the baseline specification turns the coefficient of industry size negative and highly significant in the fixed-effects regressions. Moreover, as in table 2, the coefficient of industry size is always negative and highly significant in the instrumental variables regressions (even without controlling for TFP). Finally, all the coefficients (except one) of the main variables of interest are significant beyond the 1% level under both estimation procedures.

[22] Staiger and Stock (1997) have in fact shown that two-stage least-squares estimates are unreliable when the first stage *F*-statistic is less than 10.

[23] The U.S. real interest rate has a mean value of 3.75% (with a standard deviation of 2.5%) over the period of analysis.

[24] The depreciation rates used in the empirical studies generally vary from 5% for buildings to 10% for machinery.

TABLE 3.—PROCOMPETITIVE EFFECT OF SCALE EXPANSION (NETTING OUT CAPITAL EXPENDITURES)

Indep. Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	FE	FE	FE	IV	IV	IV	IV
Y_{it}	.136***	−.065***	−.049***	−.089***	−.094***	−.108***	−.160***
	(.008)	(.010)	(.011)	(.024)	(.023)	(.030)	(.031)
$(H/L)_{it}$	.151***	.122***	.047	.702***	.649***	.853***	.626***
	(.043)	(.041)	(.048)	(.196)	(.168)	(.226)	(.224)
$Y_{it} \cdot (H/L)_{it}$	−.024***	−.019***	−.020***	−.061***	−.062***	−.088***	−.068***
	(.006)	(.006)	(.006)	(.020)	(.018)	(.023)	(.022)
$(K/PY)_{it}$		−.327***	−.339***		−.106***	−.067	−.102**
		(.010)	(.010)		(.038)	(.044)	(.041)
$(M/PY)_{it}$			−.014***			.027	.030
			(.005)			(.024)	(.024)
$(H/L)_{it} \cdot (M/PY)_{it}$			−.025***			.007	.006
			(.003)			(.014)	(.013)
TFP_{it}		.329***	.329***				.323***
		(.017)	(.019)				(.068)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16818	16818	15245	11899	11898	10517	10517
Groups	448	448	433	448	448	433	433
R-squared	.13	.21	.21	.05	.11	.09	.15
P-value Sargan test				.155	.313	.184	.121
F-statistics of Excluded Instruments in First-Stage Regressions							
Y_{it}				520	498	250	219
$(H/L)_{it}$				90	96	47	41
$Y_{it} \cdot (H/L)_{it}$				182	182	90	77
$(K/PY)_{it}$					197	81	76
$(M/PY)_{it}$						98	86
$(H/L)_{it} \cdot (M/PY)_{it}$						170	147
TFP_{it}							124

Dependent variable: price-cost margin net of capital expenditures. Notes: All variables in logs. Standard errors in parentheses. ***, **, * = significant at the 1%, 5%, and 10% levels, respectively. Coefficients of time dummies not reported. Estimation is by fixed effects (within) in columns (1) to (3), and by instrumental variables in columns (4) to (7). In IV regressions, all right-side variables are treated as endogenous, using their lagged values as instruments. Time dummies are always used as additional instruments. The bottom half of the table reports the F-statistics for the null that excluded instruments do not enter first-stage regressions. Data sources: NBER Productivity Database (by Bartelsman and Gray), NBER Trade Database (by Feenstra), and World Development Indicators (World Bank).

To conclude, the evidence on U.S. industries suggests that a scale expansion brings about a procompetitive effect, which reduces markups; moreover, the procompetitive effect is stronger in the skill-intensive industries, where markups are higher. These are the mechanics of our theory.

B. Related Literature

A few recent papers have identified alternative and more specific channels through which larger markets may be associated with a higher demand for skill. Neary (2002) shows that in the presence of oligopolistic markets, increased competition encourages strategic overinvestment by incumbent firms in order to deter entry. This raises the ratio of fixed to variable costs and, assuming that fixed costs are skill-intensive, also the skill premium. In Ekholm and Midelfart (2005), firms can choose between two technologies: a skill-intensive technology with high fixed costs and low marginal costs, and an unskilled-intensive technology with low fixed costs and high marginal costs. They then show that a trade-induced expansion in market size raises the relative profitability of the skill-intensive technology, thereby raising the skill premium. A limit of this model, where fixed costs are skill-intensive and there is free entry, is that it tends to imply counterfactually that markups should rise with skill premia. Yeaple (2005) also builds a model where firms can choose between two technologies with different ratios of fixed to marginal costs. In addition, he also allows for worker skill heterogeneity. Interestingly, he finds that firms with higher fixed costs end up hiring relatively more skilled workers, which is consistent with our assumption that fixed costs are higher in the skill-intensive industries. Otherwise, the two models are very different, in that Yeaple's model builds on a trade-induced selection effect and is mainly aimed at explaining within-group wage inequality, whereas our model builds on a procompetitive effect and is more suited for explaining average skill premia. Dinopulous and Segerstrom (1999) argue that, in models of endogenous technical change, trade can affect the skill premium by changing the reward to innovation: if trade, by expanding the market for new technologies, raises the reward to innovation and the R&D sector is skill-intensive, then it will naturally push up the skill premium. This interesting explanation seems, however, unlikely to be a major driving force behind the dramatic shifts in the demand for skill, given the small size of the R&D sector (approximately 2% of GDP in the United States) and its stability through time. Finally, in Epifani and Gancia (2004), we show that in the presence of an elasticity of substitution in consumption greater than 1 and stronger increasing returns in the skill-intensive sector, trade integration, even among identical countries, is skill-biased. We also provide evidence in support of our main assumptions. However, the model in that paper does not provide a micro foundation for the sectoral asymmetries in the scale elasticity of output.

Our result that scale is skill-biased also provides an important link among major explanations for the worldwide rise in skill premia: skill-biased technical change, capital-skill complementarity, and international trade. According to the first, inequality rose because recent innovations in the production process, such as the widespread introduction of computers, have increased the relative productivity of skilled workers.[25] In this respect, an important implication of our model is that, independent of the specific features of technological improvements, factor-augmenting technical progress may appear skill-biased simply because it raises the total supply of effective labor in the economy and therefore its scale. Similarly, the capital-skill complementarity argument [see Krusell et al. (2000), among others] emphasizes that, because new capital equipment requires skilled labor to operate and displaces unskilled workers, its accumulation raises the relative demand for skilled labor. More generally, we have shown that, even in the absence of capital-skill complementarity (indeed, even in the absence of physical capital, though that is straightforward to incorporate), factor accumulation tends to be skill-biased because it expands the scale of production. Finally, it is often argued that North-South trade liberalization may have increased wage inequality in advanced industrial countries through the well-known Stolper-Samuelson effect. However, the Stolper-Samuelson theorem is silent on the distributional effects of North-North (or South-South) trade, which represents the large majority of world trade. Our model suggests, instead, that any kind of trade integration, by increasing the market size for goods, is potentially skill-biased.

In summary, we add to the literature on the determinants of wage inequality by illustrating a mechanism that, although very simple, is surprisingly more general than the existing ones, in that it applies not only to trade-induced increases in market size but to any scale expansion. Further, and most important, it does not rely on specific assumptions on technology, but rather provides an explanation for why skill-intensive sectors become more productive as an economy grows.

IV. Conclusions

We have shown that, under plausible and fairly general assumptions about market structure, preferences, and technology, scale is nonneutral on factor rewards. The mechanics of our result can be summarized as follows. In the presence of firm-level fixed costs and free entry, economies of scale are endogenous and equal markups. Therefore, less competitive sectors are characterized by higher equilibrium scale economies, which implies that a market size increase brings about a rise in their relative output. As long as final goods are gross substitutes (complements) and sectoral production functions are homothetic in the inputs they use, this translates into a rise (fall) in the relative reward of the factor used intensively in the less competitive sectors. These are sectors characterized by a lower factor employment, higher fixed costs, or a higher degree of product differentiation.

We have also shown that, when applied to low- and high-skill workers, our theory predicts that scale is skill-biased. We have provided evidence on the mechanism underlying the skill bias of scale according to our theory. In particular, using the NBER Productivity Database, we have shown that the evidence on U.S. industries suggests that a rise in industry size reduces markups, and that the fall of markups is greater in the skill-intensive industries, where they are higher. This evidence suggests that the mechanics of skill-biased scale effects may be effectively at work in the real world.

[25] See, among others, Autor et al. (1998) for empirical evidence, and Aghion (2002) for theoretical perspectives.

REFERENCES

Acemoglu, D., "Directed Technical Change," *Review of Economic Studies* 69 (2002), 781–809.

——— "Equilibrium Bias of Technology," MIT mimeograph (2005).

Aghion, P., "Schumpeterian Growth Theory and the Dynamics of Income Inequality," *Econometrica* 70 (2002), 855–882.

Antweiler, W., and D. Trefler, "Increasing Returns and All That: A View from Trade," *American Economic Review* 92 (2002), 93–119.

Autor, D. H., L. F. Katz, and A. B. Krueger, "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113 (1998), 1169–1213.

Baumol, W. J., J. C. Panzar, and R. D. Willig, *Contestable Markets and the Theory of Industry Structure* (San Diego: Harcourt Brace Jovanovich, 1982).

Berman, E., J. Bond, and Z. Griliches, "Changes in the Demand for Skilled Labor within U.S. Manufacturing Industries: Evidence from the Annual Survey of Manufacturing," *Quarterly Journal of Economics* 109 (1994), 367–398.

Denny, M., and M. Fuss, "The Effect of Factor Prices and Technological Change on the Occupational Demand for Labor: Evidence from Canadian Telecommunications," *Journal of Human Resources* 18 (1983), 161–176.

Dixit, A., and J. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review* 67 (1977), 297–308.

Dinopoulos, E., and P. Segerstrom, "A Schumpeterian Model of Protection and Relative Wages," *American Economic Review* 89 (1999), 450–472.

Ekholm, K., and K. H. Midelfart, "Relative Wages and Trade-Induced Changes in Technology," *European Economic Review* 49 (2005), 1637–1663.

Epifani, P., and G. Gancia, "The Skill Bias of World Trade," UPF Working Paper no. 833 (2004).

Feenstra, R. C., and G. H. Hanson, "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979–1990," *Quarterly Journal of Economics* 114 (1999), 907–940.

Freeman, R. B., "Demand for Education" (Chapter 6), in O. Ashenfelter and R. Layard (Eds.) *Handbook of Labor Economics,* Vol. I (Amsterdam: North-Holland, 1986).

Galí, J., "Product Diversity, Endogenous Markups and Development Traps," *Journal of Monetary Economics* 36 (1995), 39–64.

Hall, R. E., "The Relations between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy* 96 (1988), 921–947.

Hamermesh, D., and J. Grant, "Econometric Studies of Labor-Labor Substitution and Their Implications for Policy," *Journal of Human Resources* 14 (1979), 518–542.
Hanoch, G., "The Elasticity of Scale and the Shape of the Average Costs," *American Economic Review* 65 (1975), 492–497.
Hanson, G. H., and A. Harrison, "Trade, Technology and Wage Inequality: Evidence from Mexico," *Industrial and Labor Relations Review* 52 (1999), 271–288.
Harrison, A., "Productivity, Imperfect Competition and Trade Reform," *Journal of International Economics* 36 (1994), 53–73.
Helpman, E., and P. Krugman, *Market Structure and Foreign Trade* (Cambridge, MA: MIT Press, 1985).
Hoekman, B., H. L. Kee, and M. Olarreaga, "Tariffs, Entry Regulation and Markups: Country Size Matters," *Contributions to Macroeconomics* 4:1 (2004), 1260.
Katz, L., and K. Murphy, "Changes in Relative Wages: Supply and Demand Factors," *Quarterly Journal of Economics* 107 (1992), 35–78.
Kreps, D. M., and J. A. Scheinkman, "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *Bell Journal of Economics* 14 (1983), 326–337.
Krishna, P., and D. Mitra, "Trade Liberalization, Market Discipline and Productivity Growth: New Evidence from India," *Journal of Development Economics* 56 (1998), 447–462.
Krugman, P., "Increasing Returns, Monopolistic Competition and International Trade," *Journal of International Economics* 9 (1979), 469–479.
Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica* 68 (2000), 1029–1053.
Lancaster, K., *Variety, Equity and Efficiency* (New York: Columbia University Press, 1979).
Levinsohn, J. A., "Testing the Imports-as-Market-Discipline Hypothesis," *Journal of International Economics* 35 (1993), 1–22.
Morrison Paul, C. J., and D. S. Siegel, "Scale Economies and Industry Agglomeration Externalities: A Dynamic Cost Function Approach," *American Economic Review* 89 (1999), 272–290.
Neary, P., "Foreign Competition and Wage Inequality," *Review of International Economics* 10 (2002), 680–693.
Roberts, M. J., and J. R. Tybout, *Industrial Evolution in Developing Countries* (Oxford: Oxford University Press, 1996).
Rotemberg, J. J., and M. Woodford, "The Cyclical Behavior of Prices and Costs," in J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics,* Vol. 1B (Amsterdam: North-Holland, 1999).
Salop, S. C., "Monopolistic Competition with Outside Goods," *Bell Journal of Economics* 10 (1979), 141–156.
Slaughter, M., "What Are the Results of Product-Price Studies and What Can We Learn from Their Differences?" (pp. 129–145), in R. C. Feenstra (Ed.), *The Impact of International Trade on Wages* (Chicago: University of Chicago Press, 2000).
Staiger, D., and J. H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica* 65 (1997), 557–586.
Tybout, J. R., "Plant- and Firm-Level Evidence on 'New' Trade Theories," in E. K. Choi and J. Harrigan (Eds.), *Handbook of International Economics* (Oxford: Basil Blackwell, 2003).
Tybout, J. R., and M. D. Westbrook, "Trade Liberalization and the Dimensions of Efficiency Change in Mexican Manufacturing Industries," *Journal of International Economics* 39 (1995), 53–78.
Tybout, J. R., J. De Melo, and V. Corbo, "The Effects of Trade Reforms on Scale and Technical Efficiency: New Evidence from Chile," *Journal of International Economics* 31 (1991), 231–250.
Wood, A., "Globalisation and the Rise in Labour Market Inequalities," *The Economic Journal* 108 (1998), 1463–1482.
Yeaple, S. R., "A Simple Model of Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics* 65 (2005), 1–20.

APPENDIX

Contestable Markets With Cobb-Douglas Production Functions

We now extend the simple model of contestable markets to study a situation where each good is a Cobb-Douglas composite of both factors, V_i and V_j. The total cost function of each producer of y_i is thus modified as follows:

$$C_i(\upsilon) = [F_i + c_i y_i\ (\upsilon)]\ w_i^{\gamma_i} w_j^{1-\gamma_i}, \tag{A1}$$

with $1 \geq \gamma_i \geq \gamma_j \geq 0$, that is, y_i is intensive in factor *i*. Demand for factors in each sector can be derived applying Shephard's lemma. Then, defining $\omega \equiv w_i/w_j$ and imposing the factor market-clearing conditions

$$V_i = \gamma_i \omega^{\gamma_i - 1}(F_i + c_i y_i) + \gamma_j \omega^{\gamma_j - 1}(F_j + c_j y_j),$$

$$V_j = (1 - \gamma_i)\omega^{\gamma_i}(F_i + c_i y_i) + (1 - \gamma_j)\omega^{\gamma_j}(F_j + c_j y_j),$$

we can solve for the sectoral output:

$$Y_i = y_i = \frac{(1-\gamma_j)\omega^{1-\gamma_i}V_i - \gamma_j \omega^{-\gamma_i}V_j}{c_i(\gamma_i - \gamma_j)} - \frac{F_i}{c_i}. \tag{A2}$$

An analogous expression gives Y_j. Note that sectoral output is increasing in the supply of its intensive factor and decreasing in that of the other:

$$\frac{\partial Y_i}{\partial V_i} > 0,$$

$$\frac{\partial Y_i}{\partial V_j} < 0.$$

In the jargon of trade economists, these are just Rybczynski derivatives.

To see the factor bias of scale, it suffices to compute the scale elasticity of output and refer to the equation (6), showing, for example, that with $\epsilon > 1$ scale is biased toward the factor used intensively in the sector with a higher scale elasticity. Differentiation of equation (30) yields

$$e_s^{Y_i} = \frac{(1-\gamma_j)\omega^{1-\gamma_i}V_i - \gamma_j\omega^{-\gamma_i}V_j}{(1-\gamma_j)\omega^{1-\gamma_i}V_i - \gamma_j\omega^{-\gamma_i}V_j - F_i(\gamma_i - \gamma_j)}.$$

As in the simpler models in the main text, the scale elasticity of output in sector *i* has the following properties: it is decreasing in V_i and approaches 1 as V_i tends to infinity, it is increasing in F_i, and independent of c_i. Moreover, now $e_s^{Y_i}$ rises with V_j. This simple extension confirms that increasing returns will be stronger in the sector using intensively the scarce factor (in absolute terms) and subject to higher fixed costs.