

Centre de Referència en Economia Analítica

Barcelona Economics Working Paper Series

Working Paper n° 210

Increasing Returns, Imperfect Competition and Factor Prices

Paolo Epifani and Gino Gancia

July, 2004

Increasing Returns, Imperfect Competition and Factor Prices*

Paolo Epifani[†] Gino Gancia[‡]
Parma University and CESPRI CREI and UPF

July 30, 2004

Barcelona Economics WP n° 210

Abstract

We show how, in general equilibrium models featuring increasing returns, imperfect competition and endogenous markups, changes in the scale of economic activity affect 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 hypothesis, our theory suggests that scale is skill-biased. This result provides a microfoundation 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.

JEL Numbers: F12, F16.

Keywords: Endogenous Markups, Pro-competitive Effect, Income Distribution, Trade Models with Imperfect Competition, Wage Inequality.

*We thank Daron Acemoglu, Philippe Aghion, Giovanni Bruno, Antonio Ciccone, Torsten Persson, Jaume Ventura, Fabrizio Zilibotti and seminar participants at CREI, 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.

[†]University of Parma, via Kennedy, 6 - 43100 Parma (Italy). E-mail: paolo.epifani@unipr.it

[‡]Corresponding author: CREI and Universitat Pompeu Fabra, Ramon Trias Fargas, 25-27, 08005, Barcelona (Spain). E-mail: gino.gancia@upf.edu

1 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 for the distributional effects of each of these phenomena, very little effort has been devoted to study 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 non-neutral on income distribution.

Given that there is no unified theory of imperfect competition, we derive our results within three widely used models: contestable markets (Baumol et al, 1982), quantity competition (Cournot) and price competition with differentiated products (Lancaster, 1979). All the models we use 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), these models imply that a scale expansion involves a pro-competitive effect which forces firms to lower 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 terms of factor-intensity, degree of product differentiation, fixed and marginal costs. On the demand side, the elasticity of substitution in consumption between final goods is allowed to differ from one. Under these assumptions, we show that any increase in market size is generally non-neutral 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 scarce factor in the economy. The converse happens when final goods are gross-complements, while it is only in the knife-edge case of 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 the sense 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 (lower) than one.

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 (e.g., Helpman and Krugman, 1985). Likewise, our theory suggests that technical progress, by increasing market size, is non-neutral on income distribution, thus contributing to the recent literature on the factor-bias of technical progress (e.g., Acemoglu, 2002). Finally, our result implies 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.

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 this respect, 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 low skill-intensive products. Under these circumstances, our theory implies that scale is skill-biased, thereby providing a microfoundation for the perpetual increase in the relative demand for skilled workers.¹ Moreover, since technical change, as well as factor accumulation and trade integration all imply a market size increase, we conclude that they 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 US. 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 pro-competitive 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.

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

2 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:

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

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 (1) is:

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

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 non-neutrality of scale would be merely an assumption. It follows that a proportional increase in the expenditure allocated to each sector does not change the relative factor demand, so that scale can affect relative factor prices only as long as it changes the income shares of sectors. More precisely, a scale expansion that leaves V_i/V_j unchanged will increase the relative reward of factor i , w_i/w_j , if and only if it raises the income share of the i -intensive good. In turn, equation (2) implies that sectoral shares are entirely characterized by either the relative price of goods or the relative output:

$$S(i) \equiv \frac{P_i Y_i}{P_j Y_j} = \left(\frac{\gamma}{1-\gamma} \right)^\epsilon \left[\frac{P_i}{P_j} \right]^{1-\epsilon} = \frac{\gamma}{1-\gamma} \left[\frac{Y_i}{Y_j} \right]^{(\epsilon-1)/\epsilon}. \quad (3)$$

Equation (3) shows that scale affects expenditure shares and therefore relative factor prices as long as the scale elasticity of output (i.e., increasing returns) is different across sectors. 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. The converse happens when goods are gross-complements ($\epsilon < 1$), while 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 scale economies? 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. Since 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:² $R(\cdot) = \theta(\cdot)$. The reason is that profits 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 the 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. Finally, 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 preferences for variety in aggregate. Remarkably, instead of modifying our previous findings, this new element will just reinforce them.

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 $S(i) = w_i V_i / w_j V_j$, knowing that similar results can be derived from any homothetic sectoral production functions.

2.1 CONTESTABLE MARKETS

We start with one of the simplest forms of imperfect competition: contestable markets in which the threat of entry drives down prices to average costs even if goods are produced by monopolists. Assume that there are many potential competitors (indexed by v) who can produce good Y_i with the same technology. In particular, the total cost function of each

²See Helpman and Krugman (1985) for a formal derivation.

producer in sector i entails a fixed requirement, F_i , and a constant marginal requirement, c_i , of efficiency units of factor i :

$$TC_i(v) = [F_i + c_i y_i(v)] w_i, \quad (4)$$

where $y_i(v)$ 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 (i.e., $\sum_v y_i(v) = Y_i$), feasibility (meaning that no firm is making 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 sectoral output:

$$Y_i = \frac{V_i - F_i}{c_i}. \quad (5)$$

Analogous conditions apply to sector j . Substituting (5) (and the analogous for sector j) into (3) and recalling that $S(i) = w_i V_i / w_j V_j$, we can express the relative factor rewards as:

$$\frac{w_i}{w_j} = \frac{\gamma}{1 - \gamma} \left(\frac{V_j}{V_i} \right)^{1/\epsilon} \left[\frac{c_j}{c_i} \cdot \frac{1 - F_i/V_i}{1 - F_j/V_j} \right]^{1 - \frac{1}{\epsilon}}. \quad (6)$$

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 one, a higher relative marginal cost raises the relative price of the final good and reduces its expenditure share because consumers demand more than proportionally the cheaper good. 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 (6) it is easy to see that whenever goods are gross substitutes (i.e., if $\epsilon > 1$) an increase in scale that leaves the relative endowment unchanged raises the relative price of factor i as long as:

$$\frac{F_i}{V_i} > \frac{F_j}{V_j}. \quad (7)$$

The converse is true when final goods are gross-complements (i.e., $\epsilon < 1$). Finally, 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 a 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 (5) the scale elasticity of output, is easily computed:

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

which is greater than one 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.

2.2 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. Since Y_i represents 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.³ In particular, we assume that in each sector there is a continuum of intermediates of measure one and that the production functions for final goods take the following CES form:

$$Y_i = \left[\int_0^1 Y_i(v)^{\frac{\sigma_i-1}{\sigma_i}} dv \right]^{\frac{\sigma_i}{\sigma_i-1}}, \quad (8)$$

where $Y_i(v)$ is the total amount of the intermediate good type v used in the production of good i , and $\sigma_i > 1$ is the elasticity of substitution among any two varieties of intermediates used in sector i . The price for final good Y_i (equal to the average cost) implied by (8) is:

$$P_i = \left[\int_0^1 p_i(v)^{1-\sigma_i} dv \right]^{1/(1-\sigma_i)}, \quad (9)$$

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

Imperfectly competitive firms operate at the more disaggregated level of intermediate industries. Each intermediate v is an homogeneous good produced by a finite number $n_i(v)$ of symmetric firms engaging in Cournot competition. Again, the production of each intermediate v in sector i involves a fixed requirement, F_i , and a constant marginal requirement, c_i , so

³Equivalently, Y_i and Y_j can be interpreted as consumption baskets of i - and j -intensive goods and Y as a utility function.

that the total cost function for a producer of variety v in sector i is still given by (4). Profit maximization by intermediate firms, taking output of other competitors as given, implies the following pricing rule:

$$p_i(v) = p_i = \left[1 - \frac{1}{\sigma_i n_i(v)} \right]^{-1} c_i w_i, \quad (10)$$

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):

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

Full employment requires:

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

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:

$$n_i(v) = n_i = \left(\frac{V_i}{F_i \sigma_i} \right)^{1/2}, \quad (11)$$

$$y_i(v) = y_i = \frac{1}{c_i} \left[(V_i F_i \sigma_i)^{1/2} - F_i \right]. \quad (12)$$

Note that a scale increase (i.e., an increase in V_i) is associated with a rise in firms' output. This is a direct consequence of the pro-competitive effect of a market size expansion, which reduces price-marginal cost markups and forces firms to increase output to cover fixed costs. Note also that, as shown by (11), the number of firms grows less than market size. This is the so-called defragmentation effect of a market size expansion (Helpman, 1984): when, due to fiercer competition, the price falls, some firms must exit for the surviving ones to expand their output.

Finally, note that symmetry implies:

$$Y_i = Y_i(v) = n_i y_i, \quad P_i = p_i(v) = p_i. \quad (13)$$

The same conditions apply to sector j . The relative factor reward can be found by substituting (11), (12), (13) and the analogous conditions for sector j into (3) and recalling that $S(i) =$

$w_i V_i / w_j V_j$:

$$\frac{w_i}{w_j} = \frac{\gamma}{1-\gamma} \left(\frac{V_j}{V_i} \right)^{1/\epsilon} \left[\frac{c_j}{c_i} \cdot \frac{1 - (F_i/V_i \sigma_i)^{1/2}}{1 - (F_j/V_j \sigma_j)^{1/2}} \right]^{\frac{\epsilon-1}{\epsilon}}. \quad (14)$$

Equation (14) is almost identical to (6). In particular, the relative factor price (w_i/w_j) depends on 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 (i.e., $\epsilon > 1$), an increase in scale that leaves the relative endowment unchanged raises the relative price of factor i as long as:

$$\frac{F_i}{V_i \sigma_i} > \frac{F_j}{V_j \sigma_j}. \quad (15)$$

Again, the converse is true when final goods are gross-substitutes (i.e., if $\epsilon < 1$), while relative factor rewards are scale-invariant if and only if $\epsilon = 1$. Compared to (7), 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 with the previous case is easily explained. As before, in equilibrium sectoral increasing returns are proportional to markups. In fact, using equations (10) and (11) it is possible see that (15) holds whenever the markup is higher in the i -intensive sector. However, while under contestable markets markups are determined uniquely by technological factors (the ratio of fixed costs to output), 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 pro-competitive and the defragmentation effects imply that firms' output grows with market size. For this reason, sectoral production functions exhibit increasing returns to scale that fall with market size, just like that of any single firm. Substituting (11) and (12) into (13) 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:

$$e_s^{Y_i} = \frac{1 - (1/2) (F_i/V_i \sigma_i)^{1/2}}{1 - (F_i/V_i \sigma_i)^{1/2}}, \quad (16)$$

which is greater than one and decreasing in $V_i \sigma_i / F_i$. Together with equation (3), (16) shows how variable increasing returns at the sectoral level determine the factor bias of scale.

2.3 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 $Y_i(v)$ are not substitutable, i.e., $\lim \sigma_i \rightarrow 0$, so that (8) becomes Leontief. 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 an 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 neutralize any strategic interaction among firms in different intermediate industries by assuming that all the varieties in (8) are demanded in the same amount.⁴ Our results do not depend on this assumption.

Within each intermediate industry producing $y_i(v)$ 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 in an additional cost that is higher the further away the intermediate is located 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 $p_i(v)$ is the price of the ideal type, the price of an equivalent unit bought by a firm located at distance x will be $p_i(v) e^{xd_i}$. Note that the function e^{xd_i} , Lancaster’s compensation function, parametrizes the degree of product differentiation. As $d_i \rightarrow 0$, 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 of $1/\sigma_i$ 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

⁴A more drastic simplification would be to remove the continuum of intermediate industries. In this case, however, firms would not take the price of final goods, P_i and P_j , as given.

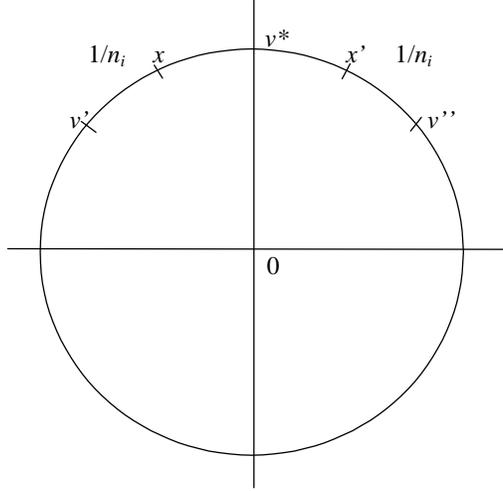


Figure 1: Price competition with differentiated products

uniformly distributed at random on each circular product space. We also assume that sellers are located equidistant from one another on each circle.⁵ 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 $n_i(v)$ firms have entered the market for $y_i(v)$, we can calculate demand for each firm as follows. Suppose that firm v^* , represented graphically in Figure 1 as a point on the product space of industry v , sets a price $p_i(v^*)$ for its type. A buyer whose ideal type is located at distance $x \in (0, 1/n_i(v))$ from v^* is indifferent between purchasing from firm v^* and from its closest neighbor on the circle v' if:

$$p_i(v^*) e^{x d_i} = p_i(v') e^{d_i(1/n_i(v)-x)}. \quad (17)$$

Therefore, given the prices $p_i(v)$, (17) implicitly defines the market width for any single firm: all buyers whose ideal type is within the arc distance x from type v^* are customers of firm v^* . Note that, in general, an increase in the price set by a firm will have two effects. First, as shown in (17), 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 (17) as:

$$D_i(v) = 2x \bar{Y}_i = \left[\frac{1}{n_i(v)} + \log \left(\frac{p_i(v')}{p_i(v)} \right)^{1/d_i} \right] \bar{Y}_i, \quad (18)$$

⁵It can be shown that this is indeed optimal. The reason is that a firm that tries to change slightly its location 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 alert the quantity demanded

where \bar{Y}_i is the aggregate sectoral output that would be produced if optimal types were always used. Profit maximization given (18) and the already introduced cost function (4) yields a familiar pricing formula (after imposing symmetry):

$$p_i(v) = p_i = \left[1 - \frac{d_i}{n_i(v)} \right]^{-1} c_i w_i. \quad (19)$$

Note that, as in the Cournot case, the markup over marginal cost decreases with $n_i(v)$. Moreover, setting $d_i = 1/\sigma_i$, (19) reduces exactly to (10). Hence, in our specification, firm's behavior under Cournot competition within differentiated industries is isomorphic to that under price competition with differentiated products in each 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 (11) and (12) 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, $n_i(v) y_i(v)$, 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 higher number of firms lies in the pro-competitive 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 v , $\bar{Y}_i = n_i(v) y_i(v)$, less the cost of the mean "distance" from the ideal type:

$$Y_i = \frac{\bar{Y}_i}{2n_i \int_0^{1/2n_i} e^{xd_i} dx} = \frac{\bar{Y}_i}{\frac{2n_i}{d_i} (e^{d_i/2n_i} - 1)}. \quad (20)$$

Since $\lim_{d_i/2n_i \rightarrow 0} \left[\frac{2n_i}{d_i} (e^{d_i/2n_i} - 1) \right] = 1$, one can show that final output in sector i grows to \bar{Y}_i as the number of available types grows to infinity ($n_i \rightarrow \infty$) or types become perfect substitutes ($d_i \rightarrow 0$). 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 (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 (14). Simple inspection reveals that the

condition for the factor bias of scale is identical to the previous case (15).

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. Remarkably, this new effect does not alert our previous conclusions because it depends on the elasticity of substitution between varieties and the number of firms, just like market power.⁶

2.4 DISCUSSION

We have shown how in models with increasing returns and imperfect competition market size affects 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 conditions (7) and (15) 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, since the scale elasticity of sectoral outputs converges to one for V_i approaching asymptotically infinity, 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*. For example, in the model of quantity competition, it is easy to show from (14) 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 more likely the higher is the elasticity of substitution ϵ . Third, while 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, while 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. 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.

⁶In fact, in the model of monopolistic competition by Dixit and Stiglitz (1977), where markups and firms' size are constant, only this latter effect would survive. However, a constant markup is usually seen as a limit of the Dixit-Stiglitz otherwise convenient formulation. Sometimes this property is removed by assuming that demand becomes more elastic when the number of varieties increases (as in Krugman, 1979).

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 (e.g., Helpman and Krugman, 1985). More generally, since 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 Stolper-Samuelson effects so that the conventional distributional implications of trade are overturned. Similarly, our results 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.

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, since 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 et al. (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.⁷ 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 Levinshon (1993) for Turkey, Krishna and Mitra (1998) for India, and Harrison (1994) for Cote d'Ivoire, while Gali' (1995) provides cross-country evidence that markups fall with income. Further, business cycles studies show that markups tend to be countercyclical (see Rotemberg and Woodford, 1999, for a survey), which is again consistent with our hypothesis.

3 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 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

⁷See also Tybout (2001) on this point.

skill bias of scale.

If goods produced with different skill-intensity are gross substitutes, conditions (7) and (15) 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 U.S., with New Zealand ranking second with a share of 16% only.

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 Treffer (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 Non-Electrical Machinery, have an average scale elasticity around 1.2, whereas traditional low skill-intensive sectors, such as Apparel, Leather, Footwear and Food, are characterized by constant returns. Using a different methodology, Paul and Siegel (1999) estimate returns to scale in US manufacturing industries for the period 1979-1989. Their estimates of sectoral scale economies are strongly positively correlated with the sectoral skill-intensity.⁸

For these asymmetries to matter, we also need the elasticity of substitution between goods produced with different factor-intensity, ϵ , to be greater than one. In Epifani and Gancia (2002), we show that in the years from 1980 to 2000 the relative expenditure on skill-intensive goods in the US increased by more than 25%, while the relative price of traditional, low skill-intensive 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, $\log(E_h/E_l)$, and the log relative price of traditional goods, $\log(P_l/P_h)$. 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 close to 1.5, consistent with our assumption.⁹ Moreover, indirect evidence also suggests that the elasticity of substitution between low and high skill-intensive goods is significantly greater than one. In particular, in our model the *aggregate*

⁸See also Epifani and Gancia (2002) on this point.

⁹When 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).

elasticity of substitution in production between skilled and unskilled workers is equivalent to the elasticity of substitution in consumption between low and high skill-intensive goods. Several studies provide estimates of the former parameter and most of them are above one.¹⁰

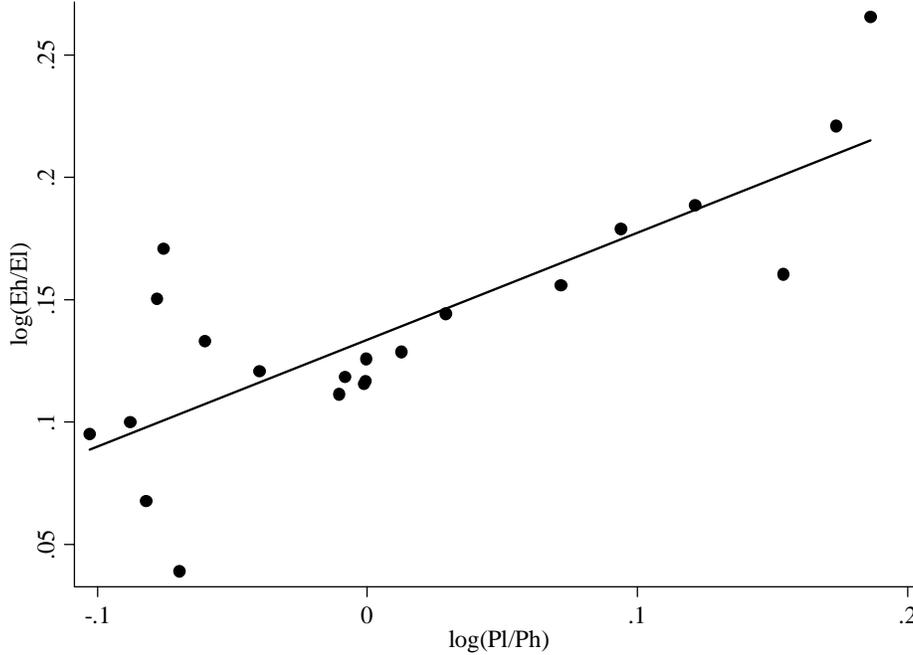


Figure 2. Elasticity of substitution between low and high skill-intensive goods. Source: Epifani and Gancia (2002)

Finally, by predicting that the aggregate relative demand for skilled workers is increasing with size, our model provides an explanation for a robust empirical finding by Antweiler and Trefler (2002), namely, 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 et al. (1994), Autor et al. (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 US.¹¹

¹⁰Freeman (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. Lastly, Krusell et al. (2000) and Katz and Murphy (1992) report estimates for the US economy of 1.67 and 1.41, respectively.

¹¹When 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 sub-sectors, 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, $S(h)_i = \frac{w_h H_i}{w_l L_i}$, is proportional to the relative output of the two sub-sectors (Y_{hi}/Y_{li}). Hence, under the assumptions discussed earlier ($\epsilon > 1$ and a higher markup in sub-sector 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.

3.1 THE PRO-COMPETITIVE EFFECT OF SCALE EXPANSION: EVIDENCE FROM U.S. INDUSTRIES

In this section, we provide evidence on the mechanism underlying the skill bias of scale according to our theory. In particular, 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 pro-competitive effect which reduces markups; c) the pro-competitive 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 low and high skill-intensive goods is also greater than one. 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:

$$MK_{it} = \beta_0 Y_{it} + \beta_1 (H/L)_{it} + \beta_2 ((H/L)_{it} \cdot Y_{it}) + \eta_i + d_t + X_{it} + \varepsilon_{it}, \quad (21)$$

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 pro-competitive effect of a scale expansion suggests that the expected sign of β_0 is negative; the expected sign of β_1 is instead positive, since our theory implies that markups are higher in the skill-intensive industries. Finally, the interaction term, $(H/L)_{it} \cdot Y_{it}$, captures the assumption that the pro-competitive effect of scale expansion is stronger in the skill-intensive industries. The expected sign of β_2 is therefore negative.

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 *a 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 industry dimension is 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 (2001) and widely used in the empirical literature on the pro-competitive effect of trade liberalization in the developing world (e.g., Roberts and 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 about 450 US manufacturing industries at the 4-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 US wage inequality.¹² Here, we show a novel way of exploiting information in this dataset to uncover a potentially relevant mechanism underlying the evolution of wage inequality in the US.

Price-cost margins are computed as the value of shipments less the cost of labor, materials and energy, divided by the value of shipments.¹³ 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 non-production workers, and therefore our measure of skill-intensity is the ratio of non-production 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.

We also add controls to the specification of equation (21) in order to isolate the pro-competitive effect of scale expansion from other relevant sources of variation in price-cost margins. In particular, we need to control for variation in industry profitability, since its rise would bring about a rise in both the price-cost margin and industry size. Therefore, without controlling for it, the pro-competitive effect of industry size expansion would be underestimated. A rough way to cope with this problem is to control for variation in industry profitability 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 size and price-cost margins, we also estimate equation (21) by instrumental variables. Interestingly, both procedures lead to similar results.

We also control for capital-intensity as, *ceteris paribus*, more capital-intensive industries require higher price-cost margins to cover the cost of capital. 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 US imports by 4-digit SIC industry (1972-basis) for the the period from 1958 to 1994 comes from the NBER Trade Database by Feenstra. Our model also suggests that the pro-competitive effect of foreign competition is stronger in the skill-intensive industries. To

¹²See, in particular, Berman et al. (1994), Autor et al. (1998), and Feenstra and Hanson (1999).

¹³This measure is not perfectly consistent with our model, since 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 this measure is considered a good proxy for firms' market power in the empirical literature (e.g., Robers and Tybout, 1996; Tybout, 2001) . At any rate, by slightly changing our model to assume that the fixed cost is in terms of physical capital only, this proxy would be perfectly consistent with our model, while leaving its main implications unchanged.

capture this effect, we also include the interaction term $(H/L)_{it} \cdot (M/PY)_{it}$, 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 (21) by using the fixed-effects within estimator. We always include time dummies to avoid spurious results due to correlation of our covariates with a time trend. 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 highly significant; 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, since 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 industry profitability, the coefficient β_0 would be upward biased and the pro-competitive 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 impact of industry size on price-cost margins is restored and is highly significant. Note, also, that the coefficient of TFP is highly significant and very large in magnitude, and that the coefficients of the other explicatives are also significant and have the expected sign.

Capital-intensive industries generally require higher price-cost margins to cover the fixed cost of capital. Therefore, in column (3) we also control for capital-intensity ($(K/N)_{it}$, where N stands for employment), whose coefficient has the expected sign and is marginally significant at the 5-percent level. Our main variables still have the expected sign and are significant at the 1-percent 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 10-percent and 1-percent levels, respectively), suggesting that import competition reduces markups and that this effect is stronger in the skill-intensive industries.

Since the within estimator uses only temporal variation to estimate coefficients, in column (5) we complement our analysis by rerunning our previous specification using the random-effects estimator. Note that the coefficients estimated by random-effects are broadly consistent with those estimated by fixed-effects; more precisely, the size and significance of the coefficients of the two variables related to foreign competition is slightly increased, the size and significance of our theory-based regressors is slightly reduced, while the coefficient of capital-intensity turns negative, although insignificant. However, Hausman's specification test strongly suggests that

treating unobservable industry heterogeneity as random may lead to misspecification¹⁴.

Although the fixed-effects results reported in Table 1 represent an interesting test of our theory, they leave some important methodological issues unsolved. In particular, while fixed-effects regressions remedy to 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-hand side variables, and in particular between the price-cost margins and industry size. Therefore, we rerun various specifications of (21) by using instrumental variables. Table 2 reports the results of the fixed-effects-instrumental-variables estimation. In all specifications, we instrument all right-hand 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, 6th and 7th lags turn out to be appropriate instruments for all endogenous covariates in all specifications.¹⁵ As shown by the p -value of the Sargan test in the bottom line of the first 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. Therefore, in the second 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 indeed strong.¹⁶

Since 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 (21) without controls. Note that the coefficients of all variables suggested by our theory have the expected sign and are significant at the 1-percent level. Moreover, columns (2) to (4) show that adding controls to our baseline regression steadily increases the size and significance of the coefficients of our covariates of interest. It is remarkable, in particular, that even without controlling for the TFP (columns (1)-(3)), the coefficient of industry size, which captures the pro-competitive effect of scale expansion, is always negative and highly significant, just as in columns (2)-(4) of Table 1, where we do control for the TFP. This suggests that controlling for the TFP in a *non-IV* regression washes out much of the simultaneity bias due

¹⁴The p -value of the null-hypothesis of appropriateness of the random-effects estimator is lower than 0.001.

¹⁵An exception is the TFP variable in equation (4), which is instrumented by using its 9th, 10th and 11th lag.

¹⁶Staiger and Stock (1997) have in fact shown that Two Stage Least Squares estimates are unreliable when the first stage F -statistic is less than ten.

to mutual interaction between price-cost margins and industry size.

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

3.2 RELATED LITERATURE

A few recent papers have identified alternative and more specific channels through which larger markets may be associated to a higher demand for skill. Neary (2002) shows that in the presence of oligopolistic markets, increased competition encourages strategic over-investment 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-Knarvik (2001), firms can choose between two technologies: one exhibits high skill-intensive fixed costs and low unskill-intensive marginal costs, while the other exhibits 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 general limit of these models where fixed costs are skill-intensive is that they tend to imply counterfactually that markups should rise with skill premia. Dinopoulos and Segerstrom (1999) argues 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 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 (about 2% of GDP in the US) and its stability through time. Finally, in Epifani and Gancia (2002), we show that in the presence of an elasticity of substitution in consumption greater than one 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 the 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.¹⁷ In this respect, an important implica-

¹⁷See, among others, Autor et al.(1998) for empirical evidence and Aghion (2002) for theoretical perspectives.

tion 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, since 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 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 which, although very simple, is surprisingly more general than the existing ones, since 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 a microfoundation for why skill-intensive sectors become more productive as an economy grows.

4 CONCLUSIONS

We have shown that, under plausible and fairly general assumptions about market structure, preferences and technology, scale is non-neutral 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 direct evidence on the mechanism underlying the skill bias of scale according our theory. In particular, using the NBER Productivity Database, we have shown that the evidence on US industries strongly 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.

REFERENCES

- [1] Acemoglu, Daron (2003). "Patterns of Skill Premia," *Review of Economic Studies* 70, 199-230.
- [2] Acemoglu, Daron (2002). "Directed Technical Change," *Review of Economic Studies* 69, 781-809.
- [3] Aghion, Philippe (2002). "Schumpeterian Growth Theory and the Dynamics of Income Inequality," *Econometrica* 70, 855-882.
- [4] Antweiler, Werner and Trefler Daniel (2002). "Increasing Returns and All That: A View from Trade," *American Economic Review* 92, 93-119.
- [5] Autor, David H., Lawrence F. Katz and Alan B. Krueger (1998). "Computing Inequality: Have Computers Changed the Labor Market?," *Quarterly Journal of Economics* 113, 1169-1213.
- [6] Baumol, Wilian J., John C. Panzar, and Robert D. Willig (1982). *Contestable Markets and the Theory of Industry Structure*, San Diego, Harcourt Brace Jovanovich.
- [7] Berman, Eli, John Bound and Zvi Griliches (1994). "Changes in the Demand for Skilled Labor within U.S. Manufacturing Industries: Evidence from the Annual Survey of Manufacturing ," *Quarterly Journal of Economics* 109, 367-398.
- [8] Denny, Michael, and Melvyn Fuss (1983). "The Effect of Factor Prices and Technological Change on the Occupational Demand for Labor: Evidence from Canadian Telecommunications," *Journal of Human Resources* 18, 161-176.
- [9] Dixit, Avinash and Joseph Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity," *American Economic Review* 67, 297-308.
- [10] Dinopoulos, Elias and Paul Segerstrom (1999). "A Schumpeterian Model of Protection and Relative Wages," *American Economic Review* 89, 450-472.
- [11] Ekholm, Karolina and Karen Helene Midelfart-Knarvik (2001). "Relative Wages and Trade-Induced Changes in Technology," *CEPR Discussion Paper* No. 2677.

- [12] Epifani, Paolo and Gino A. Gancia (2002). "The Skill Bias of World Trade," *Institute for International Economic Studies Seminar Paper* No. 707.
- [13] Feenstra, Robert C. and Gordon H. Hanson (1999). "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979-1990," *Quarterly Journal of Economics* 114, 907-940.
- [14] Feenstra, Robert C. and Gordon H. Hanson (1996). "Foreign Investment, Outsourcing and Relative Wages," in R.C. Feenstra, G.M. Grossman and D.A. Irwin eds., *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*, MIT Press, 89-127.
- [15] Freeman, Richard B. (1986). "Demand for Education," Chapter 6 in Orley Ashenfelter and Richard Layard (editors) *Handbook of Labor Economics*, North Holland, Vol I, 357-386.
- [16] Gali', Jordi (1995). "Product Diversity, Endogenous Markups and Development Traps," *Journal of Monetary Economics* 36, 39-64.
- [17] Hamermesh, Daniel and James Grant (1979). "Econometric Studies of Labor-Labor Substitution and Their Implications for Policy," *Journal of Human Resources* 14, 518-542.
- [18] Hanson, Gordon H. and Ann Harrison (1999). "Trade Technology and Wage Inequality: Evidence from Mexico" *Industrial and Labor Relations Review* 52, 271-288.
- [19] Harrison, Ann (1994). "Productivity, Imperfect Competition and Trade Reform," *Journal of International Economics* 36, 53-73.
- [20] Katz, Lawrence and Kevin Murphy (1992). "Changes in Relative Wages: Supply and Demand Factors," *Quarterly Journal of Economics* 107, 35-78.
- [21] Helpman, Elhanan (1984). "Increasing Returns, Imperfect Markets and Trade Theory," in R.W. Jones and P. B. Kenen (eds.), *Handbook of International Economics*, Vol.1, North Holland, Amsterdam.
- [22] Helpman, Elhanan and Paul Krugman (1985). *Market Structure and Foreign Trade*, MIT Press, Cambridge MA.
- [23] Krishna, Pravin and Devashish Mitra (1998) "Trade Liberalization, Market Discipline and Productivity Growth: New Evidence from India" *Journal of Development Economics* 56, 447-462.
- [24] Krugman, Paul (1979). "Increasing Returns, Monopolistic Competition and International Trade," *Journal of International Economics* 9, 469-479.

- [25] Krusell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull and Giovanni L. Violante (2000). “Capital-skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica* 68, 1029 -1053.
- [26] Lancaster, Kelvin (1979). *Variety, Equity and Efficiency*, New York, Columbia University.
- [27] Levinsohn, James A. (1993). “Testing the Imports-as-Market-Discipline Hypothesis,” *Journal of International Economics* 35, 1-22.
- [28] Neary, Peter (2001). “Foreign Competition and Wage Inequality,” *Review of International Economics*, 10:4, 680-693
- [29] Paul, Catherine J. Morrison and Donald S. Siegel (1999). “Scale Economies and Industry Agglomeration Externalities: A Dynamic Cost Function Approach”, *American Economic Review* 89, 272-290.
- [30] Robbins, Donald (1996). “HOS Hits Facts: Facts Win; Evidence on Trade and Wages in the Developing World,” *HIID Discussion Paper 557*, Harvard University.
- [31] Roberts, Mark J. and James R. Tybout (1996). “*Industrial Evolution in Developing Countries*,” Oxford University Press, Oxford.
- [32] Rotemberg, Julio and Michael Woodford (1999). “The Cyclical Behavior of Prices and Costs,” *Handbook of Macroeconomics*, Vol 1B, Amsterdam North-Holland.
- [33] Salop, Steven C. (1979). “Monopolistic Competition with Outside Goods,” *Bell Journal* 10, 141-156.
- [34] Slaughter, Matthew (2000). “What Are the Results of Product-Price Studies and What Can We Learn from Their Differences?,” in R.C. Feenstra eds., *The Impact of International Trade on Wages*, 129-165, Chicago, University of Chicago Press for the NBER.
- [35] Staiger, Douglas and James H. Stock (1997). “Instrumental Variables Regression with Weak Instruments,” *Econometrica* 65, 557-586.
- [36] Tybout, James R. (2001). “Plant-and Firm-Level Evidence on ‘New’ Trade Theories,” *NBER Working Paper W 8418*.
- [37] Tybout, James R. and M. Daniel Westbrook (1995). “Trade Liberalization and the Dimensions of Efficiency Change in Mexican Manufacturing Industries,” *Journal of International Economics* 39, 53-78.

- [38] Tybout, James R., Jaime De Melo and Vittorio Corbo (1991). "The Effects of Trade Reforms on Scale and Technical Efficiency: New Evidence from Chile," *Journal of International Economics* 31, 231-250.
- [39] Wood, Adrian (1998). "Globalisation and the Rise in Labour Market Inequalities," *The Economic Journal* 108, 1463-1482.

Table 1. Pro-competitive effect of scale expansion (Fixed-Effects)Dependent variable: price-cost margin (MK_{it})

Indep. variables	(1)	(2)	(3)	(4)	(5)
	FE	FE	FE	FE	RE
Y_{it}	.020*** (.006)	-.064*** (.007)	-.061*** (.007)	-.055*** (.008)	-.041*** (.007)
$(H/L)_{it}$.142*** (.034)	.126*** (.033)	.133*** (.033)	.118*** (.038)	.098*** (.037)
$Y_{it} \cdot (H/L)_{it}$	-.012*** (.004)	-.011** (.005)	-.012*** (.004)	-.015*** (.005)	-.008* (.005)
TFP_{it}		.324*** (.014)	.326*** (.014)	.328*** (.015)	.302*** (.015)
$(K/N)_{it}$.016** (.008)	.009 (.008)	-.008 (.007)
$(M/PY)_{it}$				-.007* (.004)	-.010** (.004)
$(H/L)_{it} \cdot (M/PY)_{it}$				-.011*** (.002)	-.013*** (.003)
Time dummies	Yes	Yes	Yes	Yes	Yes
Observations	17398	17398	17398	15775	15775
Groups	448	448	448	433	433
R-squared	.13	.16	.16	.15	.15

Notes: all variables in logs. Standard errors in parentheses. ***, **, * = significant at the 1, 5 and 10-percent levels, respectively. Estimation is by Fixed-Effects (within) in (1)-(4) and by Random-Effects in (5). Coefficients of time dummies not reported. Data Source: NBER Productivity Database (by Bartelsman and Gray) and NBER Trade Database (by Feenstra).

Table 2. Pro-competitive effect of scale expansion (IV)
 Dependent variable: price-cost margin (MK_{it})

Indep. variables	(1)	(2)	(3)	(4)
Y_{it}	-.032*** (.010)	-.033*** (.010)	-.041*** (.012)	-.059*** (.018)
$(H/L)_{it}$.421*** (.086)	.433*** (.086)	.552*** (.107)	.646*** (.140)
$Y_{it} \cdot (H/L)_{it}$	-.025*** (.009)	-.026*** (.010)	-.041*** (.011)	-.049*** (.014)
$(K/N)_{it}$.0008 (.0153)	.002 (.017)	.033 (.022)
$(M/PY)_{it}$			-.004 (.013)	-.010 (.017)
$(H/L)_{it} \cdot (M/PY)_{it}$			-.002 (.007)	.007 (.009)
TFP_{it}				.048 (.051)
Time dummies	Yes	Yes	Yes	Yes
<i>P</i> -value Sargan test	.999	.502	.238	.180
<i>F</i>-statistics of excluded instruments in first stage regressions				
Y_{it}	2126	1598	861	497
$(H/L)_{it}$	251	190	96	54
$Y_{it} \cdot (H/L)_{it}$	422	319	159	94
$(K/N)_{it}$		722	398	225
$(M/PY)_{it}$			199	103
$(H/L)_{it} \cdot (M/PY)_{it}$			281	167
TFP_{it}				181
Observations	14267	14267	12712	11047
Groups	448	448	433	433
R-squared	.08	.08	.06	.06

Notes: all variables in logs. Standard errors in parentheses. ***, **, * = significant at the 1, 5 and 10-percent levels, respectively. Coefficients of time dummies not reported. Estimation is by Fixed-Effects (within) Instrumental-Variables. All RHS variables are treated as endogenous using their lagged values as instruments (see the main text). 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 Source: NBER Productivity Database (by Bartelsman and Gray) and NBER Trade Database (by Feenstra).