Managerial Incentives and the International Organization of Production

Gene M. Grossman

Department of Economics, Princeton University,
Princeton, NJ 08544

Elhanan Helpman[*]

Department of Economics, Harvard University,
Cambridge, MA 02138,
Tel Aviv University, and CIAR

Revised: March 2003

Abstract

We develop a model in which the heterogeneous firms in an industry choose their modes of organization and the location of their subsidiaries or suppliers. We assume that the principals of a firm are constrained in the nature of the contracts they can write with suppliers or employees. Our main result concerns the sorting of firms with different productivity levels into different organizational forms. We use the model to examine the implications of falling trade costs for the relevant prevalence of outsourcing and foreign direct investment.

JEL Classification: L22, F23, D23

Keywords: outsourcing, direct foreign investment, theory of the firm, intra-firm trade

---

[*]Corresponding author: Tel.: +972-3-640-8618; fax: +972-3-642-8074.
E-mail address: grossman@princeton.edu (G.M. Grossman), helpman@post.tau.ac.il (E. Helpman).

1   Introduction

The term "globalization" applies most aptly to a description of the modern manufacturing processes in many industries. Rather than specializing in the production of different goods from start to finish, countries increasingly contribute bits of value added to goods that end up being quite multinational in their origin. The process of vertical specialization lies behind the rapid growth in international trade of intermediate inputs, components, and specialized producer services, which has far outpaced in recent years the growth of world trade in final goods.[1]

Vertical specialization takes two primary forms. Firms may procure specialized components or services from arms-length providers under contractual arrangements, or they may undertake the various production and assembly activities within the boundaries of a single firm by engaging in foreign direct investment (FDI). Borga and Zeile (2001) and Hanson et al. (2001, 2002) document and analyze the substantial rise in intra-firm trade in intermediate inputs that has taken place within multinational corporations. Contractual dealings are more difficult to isolate in the trade data, but the business press is replete with stories about foreign outsourcing.

In this paper, we develop a model that can be used to study the underlying causes of the growth of vertical specialization in trade and especially the form that such trade takes in different industries. We model the endogenous choice of organizational form by principals who are unable to monitor all of the actions undertaken by their agents. We consider an industry with many firms distinguished by their potential productivity. Each principal that enters the industry acquires the technology to produce a differentiated consumer good. But production requires the cooperation of a skilled agent who has the know-how to produce an essential component or service. The principal may hire the agent to manage a "parts division" or else contract with an entrepreneur to serve as independent supplier. The principal also faces the choice of whether to engage the agent as manager or supplier in its home country or to seek to import the intermediate inputs from a subsidiary or supplier located in a foreign land.

Our model incorporates several important trade-offs that a firm faces in its choice of location and organizational form. First, a principal who operates a vertically-integrated

---

[1] A burgeoning empirical literature documents the rapid growth of world trade in intermediate inputs and the increasing extent of vertical specialization. See, for example, Campa and Goldberg (1997), Feenstra (1998), Hummels et al. (1998), Hummels et al. (2001), Feenstra and Hanson (2002), Borga and Zeile (2001), Yeats (2001), and Hanson et al. (2001, 2002).

firm may be better able to monitor her partner than one who deals at arms length. We capture this notion by assuming that a principal can observe a manager's efforts on some fraction of tasks, but she cannot monitor the efforts of an independent contractor at all. Moreover, the ability to monitor may vary with proximity. We assume that a principal is able to observe a manager's efforts in a larger fraction of tasks when the manager's division is located near to the firm's headquarters as compared to when it is located across national borders.

Second, the contracts that the principal can use to motivate her agent may differ in the alternative organizational forms. We do not attempt to derive the restrictions on contracting from first principals, but rather we imbue the alternative contractual relationships with realistic differences. In particular, we assume that a principal cannot ask an employee to post a bond which will be forfeited in case his efforts to serve the principal fail. Nor can the principal ask an employee to front the costs of inputs that will be put at risk in the production-sharing relationship. Rather, the principal structures a contract for the manager that pays him a non-negative wage provided that he performs satisfactorily on tasks that the principal can monitor and a bonus that he receives if the project succeeds. In an outsourcing relationship, the principal similarly cannot fine an entrepreneur for failing to deliver acceptable components (or, at least, the size of any bond that can be posted by a supplier is limited). The supply contract specifies a payment by the principal that will be paid no matter how the project turns out (for example, to defray the entrepreneur's expense for investing in the project and to compensate his efforts), and an amount that will be paid in return for delivery of acceptable components. An important difference between the organizational forms arises from the assumption that the principal bears the cost of labor, capital, and material inputs in an integrated firm, whereas the entrepreneur pays these costs at least initially when he operates a legally distinct entity.

Our model bears a familial relationship to previous research on the organization of the firm and optimal design of contracts for managers. This is a large literature, so we mention only two of many related papers. Holmström and Milgrom (1991) have modeled the choice of organizational form in a setting in which an agent must perform multiple tasks for the principal, some of which can be better observed than others. But their emphasis is on externalities in contract design; that is, on how the incentives provided for one task reflect the difficulty of measuring performance on others. They apply their reasoning to asset ownership, and show that "high-powered incentives" should be

more common when the agent owns the productive asset (outsourcing relationship) than when the principal owns the asset (employment relationship). In our model too the optimal contract for a potential supplier often provides higher-powered incentives than the optimal contract for a manager, but this has more to do with the restrictions we place on payments from the agent to the principal and on our assumptions about who initially bears the cost of labor, capital, and material inputs.

Like us, Horn et al. (1995) study the design of optimal incentive contracts for managers in a world of international trade. However, they do not consider the choice between vertical integration and arms-length dealing. Rather, they focus on whether international trade, by increasing the degree of competition in product markets, brings welfare gains that can be associated with increased effort by the manager and improved internal efficiency of the firm.[2]

The remainder of our paper is organized as follows. In the next section, we consider the choice of organizational form by a principal with a given potential productivity. The principal can manufacture a fixed quantity of final goods if she can obtain the necessary intermediate inputs. The inputs must be produced by a skilled partner, who may manage a division of the principal's firm or head an independent supplier. The ability of the partner to deliver suitable inputs is not assured, but depends on the partner's efforts in a variety of tasks. If the principal hires the agent as an employee, she can monitor the agent's efforts in a fraction of these tasks. If the agent is hired as an independent contractor, no monitoring is possible. In either case, the principal designs an optimal contract subject to the constraints described above, and offers it to an agent with given outside opportunities on a take-it-or-leave-it basis.

Our main result in Section 2 — which allows for no choice of location — concerns the relationship between the principal's potential revenues and her preferred organizational form. We show that outsourcing is preferred by principals who have very high or reasonably low potential revenues, whereas vertical integration may offer the greatest expected profits to a principal whose potential revenues fall in an intermediate range. In Section 3, where we introduce the locational dimension of the principal's decision problem, we find that among firms that opt for vertical integration, those that elect

---

[2]Several authors, beginning with Ethier (1986), have considered contracting problems to be at the core of the choice between foreign direct investment and arms-length trade. See, for example, Horstmann and Markusen (1987), Ethier and Markusen (1996), and Markusen (2001). While these papers share with this one a focus on imperfect contracting as a reason for vertical integration, they consider different sources of contractual incompleteness and, in particular, do not examine the difficulties that arise from imperfect monitoring of the efforts of managers and suppliers.

to keep their parts division close to their headquarters have higher potential revenues than those that engage in FDI.

We close the model in Section 4 by specifying demand for the group of competing outputs and by allowing for endogenous entry into the industry at a given cost. Principals who pay the entry fee draw a productivity level from a known distribution of potential productivities. In equilibrium, each entrant has expected operating profits equal to the fixed entry cost. The distribution of productivities and the endogenous choice of organizational form together determine the market shares of suppliers, of foreign affiliates of multinational corporations and of integrated producers in the North.

In the succeeding sections, we study the determinants of these market shares. In Section 5, we show how improvements in the ability to monitor distant managers result in an increased market share for multinational corporations, and declines in the market shares of components produced by independent suppliers and by vertically integrated producers in the home country. Section 6 analyzes the effects of falling trade costs. There we show that trade liberalization or improvements in transportation may boost the prevalence of outsourcing or of FDI, depending on whether the industry is one in which most outsourcing is undertaken by firms that are highly productive or by firms that are the least productive among those active in the industry. A concluding section contains a summary of our results.

2 Organization of the Firm

We develop a theory of the firm based on the alternative means that a principal has to address the problems caused by imperfect observability of an agent's actions. The principal owns the technology for producing a particular good but needs the cooperation of a skilled partner in order to manufacture it. The partner is needed to oversee production of an essential component. The principal can hire a partner to work as a division manager, in which case she will provide the primary inputs that are needed to produce the components and structure a suitable incentive contract for her employee. Alternatively, she can turn to an arms-length supplier of components; such supplier firms are led by "entrepreneurs" who have skills similar to those of the managers. For now, we ignore issues to do with the location of a potential parts division or supplier; later we shall allow for a choice between a local manager and one who operates in a foreign subsidiary, and between domestic and foreign outsourcing. We focus on the

4

decisions of a single principal until Section 4, where we embed the individual's choice in a model of industry equilibrium.

We assume that the principal can only operate a firm of a given (maximum) size.[3] If the principal succeeds in acquiring suitable components either from a subsidiary or from an external supplier, her output will be θ, where θ indexes the potential productivity of her firm. If she fails to acquire suitable components, output is zero. Output generates revenue $R(\theta)$, with $R(0) = 0$ and $\lim_{\theta \to \infty} R(\theta) = \infty$. For now, we suppress the potential interactions with other firms in the industry.

ה

Figure 1:

The production of components requires "effort" on the part of the principal's partner — be he a division manager or an entrepreneur heading a supplier firm — in a variety of tasks. Let $e(j)$ be the effort exerted by the manager or entrepreneur on task j. Then, with probability $\int_0^1 h[e(j)]dj$, the attempt to manufacture the requisite components "succeeds," and the resulting components can be used by the principal to produce the final good. But with probability $1 - \int_0^1 h[e(j)]dj$ the project fails, and the plans to manufacture the final good must be aborted.[4] In Grossman and Helpman (2002b) we analyze the choice between outsourcing and in-house production for a general function $h(\cdot)$ that is increasing, concave, and that reaches a maximum at some finite level of effort E. Here we focus instead on the special case that arises when $h(\cdot)$ is piecewise linear, as depicted in Figure 1. This case shares all of the important features of the more general analysis, and it provides for a more intuitive exposition.

In the figure, we have introduced a number of key parameters. Specifically, $h_0 = h(0)$ is the baseline contribution to success of a task on which the minimum effort is exerted; e_1 is the effort level at which the marginal product of effort falls from $(h_1 - h_0)/e_1$ to $(1 - h_1)/(1 - e_1)$; and $h_1 = h(e_1)$. Notice that $h(\cdot)$ reaches a maximum of $h(1) = 1$ at $e = 1$. These normalizations are inessential to our analysis. We do need concavity, however, which requires:

Assumption 1 $(h_1 - h_0)/e_1 > (1 - h_1)/(1 - e_1)$.

---

[3]In the appendix of our working paper, Grossman and Helpman (2002b), we extend the model to allow for variable scales of production.

[4]This is an extreme assumption that is used to simplify the algebra. The flavor of the analysis would be preserved if the manager's efforts were to determine the productivity of the plant that manufactures components.

If the principal were to hire a skilled individual to serve as a division manager, she would be able to monitor the manager's effort on a fraction $\delta \leq 1$ of the tasks involved in producing components. This ability to monitor an employee's efforts may give an advantage to in-house production relative to outsourcing, where the partner's activities take place in a legally and perhaps geographically distinct concern. For simplicity, we assume that the principal cannot monitor *any* of the tasks undertaken by an entrepreneur in a separate supply firm; the analysis would be quite similar if a principal could observe the entrepreneur performing some positive fraction of tasks that is smaller than what is possible for an employee.

The production of components requires primary inputs. Some of these may be fixed costs, independent of the scale of component production. Others may be variable costs. However, with a fixed scale of operation for the final producer, the number of components that can be processed is given, and the fixed and variable costs for the parts manufacturer need not be distinguished.[5] We denote the cost of the inputs needed to produce the requisite quantity of components by c. These costs are paid initially by the principal in the case of in-house production and by the entrepreneur in a supplier relationship.

The manager or entrepreneur bears a private cost of effort of $e(j)$ on task j. The total utility cost of effort is $\int_0^1 e(j)dj$. Since marginal returns to effort on a single task are non-increasing and all tasks contribute similarly to the success of the venture, the optimal supply contract induces an equal effort from the entrepreneur on all tasks; call it e_o. Similarly, an optimal employment contract for a division manager generates the same level of effort e_m on all monitorable tasks, and the same level of effort e_n on non-monitorable tasks. All agents are risk neutral, and income and effort are separable in the entrepreneur's or manager's utility function. Therefore, an entrepreneur achieves expected utility of $I_o - e_o$, where I_o is the expected profits net of input costs that accrue to the supplier under an outsourcing contract. A manager enjoys an expected utility of $I_m - \delta e_m - (1 - \delta)e_n$, where I_m is the expected income that accrues to the manager under an employment contract. Skilled individuals have an outside option to achieve utility $\bar{s}$ elsewhere in the economy. Thus, any outsourcing or employment contract must provide an entrepreneur or manager with at least this level of welfare.

We assume that the principal tenders a take-it-or-leave-it offer to the pool of skilled

---

[5]In the appendix of Grossman and Helpman (2002b), where we allow for variable production of final goods, we distinguish the fixed costs from the variable costs.

individuals, subject to some contracting constraints. First, if the principal hires a skilled individual as a manager, the total compensation paid to the manager cannot be negative in any state of nature. In other words, a manager cannot be asked to post a bond as a condition of employment.[6] Second, if the principal seeks a supplier in an outsourcing relationship, the supply contract cannot require a net payment from the supplier to the principal in any state of nature. Again, the entrepreneur cannot be asked to post a bond that is forfeited in case the attempt to produce components fails. We do not try to justify these restrictions on contracting from first principles, but rather take them to approximate realistic institutional and legal constraints that exist in many modern economies. At most, the principal can design a contract that entails a zero payment to the manager or entrepreneur in case of poor performance or an unlucky outcome; penalties or fines are not allowed.[7]

With these restrictions on the feasible contracts, we see a second difference between in-house production and outsourcing. When components are manufactured in a wholly-owned subsidiary, the principal pays the costs of the primary inputs. Then, if the project fails, the principal stands to lose this investment. In contrast, in an outsourcing relationship it is the entrepreneur who fronts the cost of the inputs, unless the principal chooses to include a fixed payment for this purpose in the contract offer. This means that an entrepreneur may have more at stake than a manager and it opens the possibility that higher-powered incentives can be offered under this arrangement. Also, in case the principal finds it optimal to design a contract that leaves (expected) rents to her skilled partner, the ability to shift input costs to a supplier but not to a manager may affect the relative attractiveness to the principal of the alternative organizational forms. When suppliers and managers cannot be asked to post bonds, the fact that the supplier pays the up-front cost of the inputs into parts production tends to favor outsourcing relative to in-house production from the perspective of the principal. This cuts against the advantage of in-house production that stems from the opportunity it affords the principal to monitor some of the manager's actions.

We proceed now to derive the (constrained) optimal contracts under each organizational form. An outsourcing contract is characterized by an amount s that the

---

[6]See Katz (1986) for a discussion of the theoretical and practical difficulties that inhibit the use of performance bonds for employees.

[7]We do not actually need the limit on the smallest payment in case of an unsuccessful project to be zero; a small enough finite negative number would suffice to yield qualitatively similar results. Note too that the exogenous restrictions on the negative payments would not be needed if we were to assume that managers and entrepreneurs are risk averse.

principal promises to pay the entrepreneur whether or not the project succeeds and an amount p_o that the principal will pay in case the supplier is able to provide the components that are needed for production of the final good. The payment of s allows for the possibility of (endogenous) cost sharing, while our restriction constrains the contract to have $s \geq 0$.

An employment contract is characterized by a level of effort e_m that the manager is expected to exert on all tasks that can be monitored by the principal, a wage payment w that the manager will receive irrespective of the outcome of the project provided that he has exerted at least the indicated level of effort on the monitorable tasks, and a bonus b that he will receive in case the project succeeds. Our restriction requires $w \geq 0$.

Figure 2:

In deriving the optimal offers, we will make use of the optimal responses of an entrepreneur or manager to the incentives that are provided in his contract. An entrepreneur chooses e_o to maximize his expected utility, $s + p_o h(e_o) - e_o$. A manager chooses $\tilde{e}_m$ and e_n to maximize $w + b[\delta h(\tilde{e}_m) + (1-\delta)h(e_n)] - \delta\tilde{e}_m - (1-\delta)e_n$, subject to the constraint that the level of effort on the monitorable tasks $\tilde{e}_m$ must be at least as great as the effort e_m specified in the contract. The reader may verify later that the manager has no incentive to exert "extra" effort on the monitorable tasks. So we set $\tilde{e}_m = e_m$ in this problem. Then note that the conditional payment p_o that must be provided to an entrepreneur to induce an effort e on all tasks is the same as the bonus b that is needed to induce that same level of effort from a manager on the monitorable tasks. We will use the function $\phi(e)$ to denote the *smallest* size of the incentive needed to induce an effort level e on an unobservable task; then $p_o = \phi(e_o)$ and $b = \phi(e_n)$. For a general productivity function $h(e)$, $\phi(e)$ is just the reciprocal of the marginal productivity of effort; i.e., $\phi(e) = 1/h'(e)$.

Figure 2 depicts $\phi(e)$ for the piecewise-linear function $h(e)$ shown in Figure 1. The dot at the origin indicates that $\phi(0) = 0$; i.e., no contingent payment is required to induce the minimum level of effort. The principal must, however, promise a strictly positive bonus to a manager or delivery payment to an entrepreneur in order to induce the entrepreneur or manager to exert above-minimum effort. With a constant marginal productivity of effort for all e between 0 and e_1, a contingent payment of $e_1/(h_1 - h_0)$

8

will leave the skilled partner indifferent between effort levels in this range. Thus, $\phi(e) = e_1/(h_1 - h_0)$ for all positive levels of effort up to and including e_1. To induce still higher levels of effort, the contingent payment must be larger, because the marginal product of effort falls at e_1. A contingent payment of $(1 - e_1)/(1 - h_1)$ suffices to make the skilled partner indifferent among all effort levels greater than e_1 up to and including the maximum effort per task of $e = 1$; and these effort levels are strictly preferred by the partner to efforts levels below e_1. We can now regard the principal as if she were choosing the effort exerted by an entrepreneur or a manager on the unobservable tasks subject to the constraint that these choices must be incentive compatible.

Since the principal must offer the same contingent payment for any effort level between 0 and e_1, but the probability of success is strictly increasing for effort levels in this range, she will prefer to induce $e = e_1$ to any positive level of effort smaller than e_1. Similarly, she will prefer to induce effort of $e = 1$ to any level of effort strictly between e_1 and 1. It follows that the contingent payments p_o and b in any optimal outsourcing or managerial contract will be either 0, $e_1/(h_1 - h_0)$, or $(1 - e_1)/(1 - h_1)$ and will induce the skilled partner to exert an effort of either 0, e_1, or 1, respectively.

2.1 Operating Profits under Outsourcing

Suppose the principal chooses to outsource the production of components. Such a principal must choose s and e_o to maximize

$$\Pi_o = h(e_o)R(\theta) - s - h(e_o)\phi(e_o) , \tag{1}$$

the difference between her expected revenues and expected total payments to her parts supplier. The principal's choices are constrained by the requirements that $s \geq 0$ and

$$s + h(e_o)\phi(e_o) - c - e_o \geq \bar{s} . \tag{2}$$

The latter is a *participation constraint*, ensuring that the entrepreneur's expected utility (equal to his expected income less the cost of the primary inputs and the utility cost of his efforts) is no less than the utility he could attain elsewhere in the economy. Of course, the principal could always choose not to engage any supplier or to produce any output, in which case her operating profits invariably will be equal to zero.

For very low levels of potential revenues, the constraint that $s \geq 0$ will not bind.

Suppose for example that $R(\theta)$ is near zero. Then it will not be worthwhile for the principal to induce positive effort from the entrepreneur. With $e_o = 0$ and $\phi(0) = 0$, the participation constraint requires $s \geq c + \bar{s} > 0$, which means that the non-negativity constraint does not bind.

It is possible that the non-negativity constraint will not bind even at higher levels of potential revenues. Suppose, for example, that potential revenues are sufficiently high that the principal wishes to induce maximal effort of $e = 1$ with a contingent payment of $(1 - e_1)/(1 - h_1)$. If $\bar{s} + c > [(1 - e_1)/(1 - h_1)] - 1$, the participation constraint (2) can only be satisfied with $s > 0$. But when $s > 0$, the maximization of (1) subject to (2) achieves a first-best for the principal subject only to a participation constraint. In such circumstances, the comparison of outsourcing with vertical integration is not very interesting, because the former (weakly) dominates as an organizational form. Accordingly, we prefer to restrict out attention to parameters for which the principal does not achieve a first-best with outsourcing for all values of R. This is true when $\bar{s} + c < (h_1 - e_1)/(1 - h_1)$. Then, for high enough values of $R(\theta)$, the principal cannot ignore the non-negativity constraint, and she sets $s = 0$ in the optimal incentive contract for the entrepreneur.

To avoid a taxonomy, we also choose to restrict attention to cases where $\bar{s} + c$ is not too low. When $\bar{s} + c > e_1 h_0/(h_1 - h_0)$, the principal never would choose to induce minimal effort from an entrepreneur (by offering a contingent payment of zero). Whenever the principal prefers $e_o = 0$ to $e_o = e_1$ and the parameter restriction holds, the principal's expected profits under outsourcing are negative.[8] Then she prefers to shut down entirely rather than to enter into a contract with a supplier. Accordingly, we adopt.

Assumption 2 $(h_1 - e_1)/(1 - h_1) > \bar{s} + c > e_1 h_0/(h_1 - h_0)$.

2.2 Operating Profits under Vertical Integration

Now suppose that the principal chooses to manufacture components in-house by hiring a manager to oversee a parts division. The principal must choose a contract for the manager that specifies a wage, an expected level of effort on monitorable tasks, and a bonus for success. Equivalently, we can think of the manager as choosing w, e_m and

---

[8]The reader may verify that profits are higher with $e_0 = 0$ and $p_0 = 0$ than with $e_0 = e_1$ and $p_0 = e_1/(h_1 - h_0)$ when $R(\theta) < e_1/(h_1 - h_0)$. But when $R(\theta)$ is so low and $\bar{s} + c > e_1 h_0/(h_1 - h_0)$, profits $\Pi_o = h_0 R(\theta) - \bar{s} - c$ (with $e_o = 0$) must be negative.

e_n to maximize

$$\Pi_v(R) = [\delta h(e_m) + (1 - \delta)h(e_n)][R - \phi(e_n)] - w - c , \qquad (3)$$

subject to

$$[\delta h(e_m) + (1 - \delta)h(e_n)]\phi(e_n) + w - \delta e_m - (1 - \delta)e_n \geq \bar{s}. \qquad (4)$$

Here, the operating profits are the expected revenues net of the expected bonus payment, the cost of the inputs, and the wage of the manager. The constraint ensures that the welfare of the manager is at least as great as what he could attain by working elsewhere. The principal also is constrained to offer a non-negative wage rate ($w \geq 0$) and to choose non-negative levels of effort for both monitorable and non-monitorable tasks.

2.3 Comparing Organizational Forms

The potential operating profits from outsourcing are depicted by the solid, kinked curve in Figure 3. At very low levels of productivity such that $R(\theta) < R_a = (\bar{s}+c+e_1)/h_1$, the principal prefers to close her shop than to engage a supplier and thereby face negative expected profits. So $\Pi_o = 0$ for $R \leq R_a$. However, when potential revenues are at least as large as R_a, a principal can earn non-negative profits by engaging a supplier and offering an incentive payment of $e_1/(h_1 - h_0)$. Such a contract would induce the entrepreneur to exert an effort of e_1 on all tasks and thereby achieve a probability of success of h_1. The principal would need to make a fixed payment to the entrepreneur of $s = \bar{s} + c + e_1 - e_1 h_1/(h_1 - h_0)$ in order to make the contract acceptable to him.[9] The expected profits from outsourcing when $R = R_a$ and $e = e_1$ are $\Pi_o(R_a) = 0$. No other outsourcing contract does as well.

Figure 3:

For a range of revenue levels above R_a, it remains optimal for a principal who chooses to outsource to write a contract that induces effort e_1 by the entrepreneur. In this range, expected operating profits are given by $\Pi_o(R) = Rh_1 - e_1 - \bar{s} - c$. But when productivity is such that $R(\theta) \geq R_c = [(1 - e_1)/(1 - h_1)^2] - [(\bar{s} + c + e_1)/(1 - h_1)]$, a

---

[9]Note that $s > 0$ in light of the second inequality in Assumption 2.

principal who chooses to outsource prefers to induce the maximal effort level of $e = 1$ from the entrepreneur by offering a contingent payment of $(1 - e_1)/(1 - h_1)$. Here, the non-negativity constraint for payments to the entrepreneur binds. The principal sets $s = 0$ and achieves expected operating profits of $\Pi_o(R) = R - (1 - e_1)/(1 - h_1)$.

If the principal opts instead for in-house production of components, it never is optimal for her to induce the manager to work harder on the tasks that cannot be monitored than on those that are observable. We focus therefore on strategies that involve $e_m \geq e_n$.[10] There now are several possibilities to consider. We can rule out $e_m = e_n = 0$, because with these minimal levels of effort, expected operating profits from in-house production always are negative under Assumptions 1 and 2. Also, by choosing $e_m = e_1$ and $e_n = 0$, the principal faces lower expected profits in an integrated operation than she would by outsourcing, at all revenue levels for which expected profits under this integration strategy are positive.[11] We can therefore disregard this strategy as well.

Another possible strategy has $e_m = e_n = e_1$. This requires a bonus offer of $b = e_1/(h_1 - h_0)$ and may or may not require a positive wage payment in order that the contract satisfies the participation constraint. When it does require a positive wage payment, the principal attains expected operating profits of $Rh_1 - e_1 - \bar{s} - c$. But these are exactly the same expected profits as she can earn by outsourcing for all revenue levels between R_a and R_c. We prefer to concentrate on circumstances in which the principal strictly prefers outsourcing to in-house production for a range of low productivity levels, because we have shown in Grossman and Helpman (2002b) that such a range typically exists when $h(e)$ has a more general form. Accordingly, we make a third parameter restriction, namely

Assumption 3 $\bar{s} < e_1 h_0/(h_1 - h_0)$.

Under Assumption 3, the participation constraint is satisfied when $e_m = e_n = e_1$ and $b = e_1/(h_1 - h_0)$, even with $w = 0$. Since the principal cannot impose a negative wage, she would be forced to leave rents to her manager were she to operate a parts division with $e_m = e_n = e_1$. By doing so, her expected operating profits are $Rh_1 - (h_1 e_1)/(h_1 - h_0) - c$, which is less than what she could earn by outsourcing.

---

[10] We show in Grossman and Helpman (2002b) that $e_m \geq e_n$ for general $h(e)$ functions.

[11] With $e_m = e_1$ and $e_n = 0$, the principal sets $b = 0$ and $w = \bar{s} + \delta e_1$. Operating profits are $\Pi_v(R) = R[\delta h_1 + (1 - \delta)h_0] - \delta e_1 - \bar{s} - c$, which is positive when $R > (\bar{s} + c + \delta e_1)/[\delta h_1 + (1 - \delta)h_0]$. But for such values of R, the principal can earn profits of $Rh_1 - e_1 - \bar{s} - c$ from outsourcing by inducing effort of $e = e_1$. Assumption 2 ensures that these profits are higher.

The principal also prefers outsourcing to in-house production when the optimal strategy under vertical integration would be to set $e_m = e_n = 1$. To induce the maximal level of effort on tasks that cannot be monitored, the principal must offer a bonus payment of $b = (1 - e_1)/(1 - h_1)$. But then the non-negativity constraint on the wage binds, and the principal must leave rents to the manager. The resulting profits are $R - [(1 - e_1)/(1 - h_1)] - c$, which is less than what the principal could earn by outsourcing and inducing an effort level of $e = 1$ on all tasks.

So far we have not identified any productivity levels and associated optimal strategies for which the principal prefers to engage a manager than to find an external supplier. But in-house production can be attractive to the principal when the optimal contract under vertical integration entails $e_m = 1$, $e_n = e_1$, and $b = e_1/(h_1 - h_0)$, as it will for an intermediate range of productivity levels. In Grossman and Helpman (2002b) we have shown that this finding — of a possible choice of in-house production by the principal at intermediate productivity levels, but not when productivity is very high or very low — is a general feature of the model and not one that relies on the piecewise-linear form of $h(e)$.

When $e_m = 1$ and $e_n = e_1$, the non-negativity constraint on the manager's wage may or may not bind. In either case, in-house production may yield higher operating profits to the principal than outsourcing for a range of values of R. Again we wish to avoid a taxonomy, so we make our final parameter restriction, which is

Assumption 4 $\bar{s} > [\delta + (1 - \delta)h_1] \left(\frac{e_1}{h_1 - h_0}\right) - \delta - (1 - \delta)e_1.$

Under Assumption 4, when $e_m = 1$ and $e_n = e_1$ a principal who hires a manager must pay him a positive wage $w > 0$. Then her expected operating profits are given by $\Pi_v(R) = R[\delta + (1 - \delta)h_1] - [\delta + (1 - \delta)e_1] - \bar{s} - c$. These profits are depicted by the dotted line in Figure 3.[12]

The figure depicts a case where the principal prefers outsourcing to in-house production for revenue levels between R_a and R_b, and also for revenue levels in excess of

---

[12] An alternative strategy for in-house production is for the principal to require $e_m = 1$, pay a wage $w = \bar{s} + \delta$, and offer no bonus. Facing such a contract, the manager would make no effort on tasks that are not monitored and would be indifferent between participating in the contract and not. The alternative strategy yields expected profits for the principal of $R[\delta + (1 - \delta)h_0] - \delta - \bar{s} - c$, which — for all $R \geq R_a$ — is strictly less than the profits of $R[\delta + (1 - \delta)h_1] - [\delta + (1 - \delta)e_1] - \bar{s} - c$ that she earns by pursuing the strategy described in the text. However, when Assumption 4 is violated, the strategy of accepting $e_n = 0$ may be the best option available to the principal for some revenue levels. Then the figure would look somewhat different from what we have drawn in Figure 4, but the main conclusion about the choice of organizational form for firms with different productivity levels would remain the same.

R_d. For intermediate revenue levels between R_b and R_d, vertical integration is her preferred mode of organization. The existence of such a range is ensured by Assumption 4.[13] It is easy to calculate that $R_b = (1 - e_1)/(1 - h_1)$, which is to the right of R_a and independent of δ. Also, an increase in δ causes the dotted line to rotate in a counterclockwise direction around point Q, which moves R_d to the right. It follows that the range of productivity levels for which the principal prefers to be vertically integrated expands as the fraction of tasks that she can monitor when producing in-house grows.

We have shown (for parameter values that satisfy Assumptions 1-4) that a principal with sufficiently high or rather low productivity prefers to buy components from an independent parts supplier, whereas one with an intermediate productivity level prefers to produce those components in-house. The advantage of outsourcing at high levels of productivity comes from the opportunity it affords a principal to reduce the rents that must be granted to the agent in a situation in which it is optimal for her to tolerate some rent sharing in order to induce maximal effort. Outsourcing cuts into the agent's rents, because the cost of the inputs used to produce the parts can be shifted to the supplier. For low levels of productivity, outsourcing is advantageous for a different reason. Here, the efficient level of effort is relatively low and the agent captures no rents. But for any level of effort, the contingent payment under outsourcing that leaves the entrepreneur without rents is larger than the bonus payment to a manager that similarly drives him to his reservation level of utility. This is because a contingent payment p_o must compensate an entrepreneur for the cost of the inputs as well as the disutility of his effort, while a bonus payment b repays only the effort. It follows that a principal can induce greater effort from an entrepreneur than from a manager for a given expected outlay. Put differently, an entrepreneur who bears the cost of the inputs has more at stake in a project than a manager who does not. When the principal brings the former to utility level $\bar{s}$, the resulting incentives have higher power than those that would bring a manager to the same level of expected utility.

The possible advantage of vertical integration for an intermediate range of productivity levels stems from the opportunity it affords the principal to monitor some of the

---

[13]Note that — in view of Assumption 3 — Assumption 4 is violated when δ is close to zero. On the other hand, Assumption 4 is satisfied when δ is close to one (because $e_1/(h_1 - h_0) < 1$ as a result of the concavity of $h(\cdot)$). So Assumption 4 essentially states that the fraction of monitorable tasks must be sufficiently large. We have shown in Grossman and Helpman (2002b) that a large enough δ also is required in the general case to ensure the existence of a range of revenue levels for which in-house production dominates outsourcing.

manager's actions. On tasks that can be monitored, the principal can induce a high level of effort without having to leave rents to the manager. She can do so simply by paying a wage that compensates the manager for his effort on these tasks, and demanding that the effort be made. If enough tasks can be monitored, the principal can achieve as high a probability of success with integration as with outsourcing without having to share rents.

3   Choice of Location

In this section, we allow for production of components in two regions distinguished by costs. The two regions, North and South, have associated variables represented by subscripts N and S, respectively. We assume that input costs are lower in the South ($c_S < c_N$) and that the outside options for those with the skills needed to head a production unit are no better there ($\bar{s}_S \leq \bar{s}_N$).[14] If the principal elects to outsource the production of components, she may choose a supplier in the North or in the South. If she opts instead to create a subsidiary to produce components, such a plant must be managed by a local employee. We assume that the principal is better able to monitor a manager's efforts when the production unit is located near the firm's headquarters in the North than when it is located in the more distant South (i.e., $\delta_N > \delta_S$). But even a remote subsidiary affords better opportunities for monitoring a skilled partner than is possible when components are produced by an arms-length supplier (i.e., $\delta_S > 0$).

Note that our model gives no advantage to outsourcing in the North to compensate for the higher costs there. We might, for example, have allowed the principal to monitor some (small) fraction of an entrepreneur's actions, with greater opportunities for this in the North than in the South. Then outsourcing in the North might have become viable for some productivity levels. Or we might have allowed for differences in the "thickness" of the markets for components or in the completeness of enforceable contracts in the alternative legal environments. The role that these latter differences play in the location of outsourcing activity was the focus of Grossman and Helpman (2002a). Here we prefer to keep matters simple, and so we accept that outsourcing in the North is a dominated option in this setting.[15]

However, foreign direct investment (FDI) and in-house production of parts in the

---

[14]Actually, our qualitative results require only that $c_S + \bar{s}_S < c_N + \bar{s}_N$.

[15]More accurately, outsourcing in the North is a strictly dominated option when revenue levels are low and it is only weakly dominated when revenues are high; see further discussion of this point below.

North both may be viable options for some parameter values and some productivity levels. The former affords an opportunity for monitoring some of the manager's efforts without giving up the advantage of low costs. The latter provides still better opportunities for monitoring, but costs are higher. Our next task is to compare the three viable options — outsourcing in the South, home production, and FDI — for different levels of productivity. We can readily show that vertical integration in location j can only compete with outsourcing when the principal would choose $e_{mj} = 1$ for the manager's effort on tasks that can be monitored in location j and $e_{nj} = e_1$ for tasks that cannot be monitored in location j. So we limit our attention to these strategies.

First note that for low levels of R at which foreign outsourcing is only marginally profitable, outsourcing is the preferred mode of organization for the principal. The comparison between foreign outsourcing and FDI is the same as we conducted in Section 2.3; the former dominates at low levels of productivity and revenues, because the principal can structure higher-powered incentives for an entrepreneur who must front the cost of inputs than she can for a manager who must be paid a non-negative wage no matter what the outcome of the project. The principal in turn prefers FDI to home in-house production when Assumption 4 is satisfied for (at least) all revenue levels $R \leq R_b = (1 - e_1)/(1 - h_1)$. To see this point, recall that an increase in δ rotates the line representing profits under vertical integration around the point Q in Figure 3. This means that the fraction of monitorable tasks has no bearing on operating profits under vertical integration when $R = R_b$. But the lower costs in the South give FDI a clear advantage over in-house production in the North at $R = R_b$; in fact, $\Pi_{vS}(R_b) - \Pi_{vN}(R_b) = (\bar{s}_N + c_N) - (\bar{s}_S + c_S) > 0$.

⊔⊓

Figure 4:

Figure 4 shows the maximal expected operating profits for a firm with potential revenues R. Implicitly, this figure reveals the optimal choice of organization and location for each value of R. For $R < R_a = (\bar{s}_S + c_S + e_1)/h_1$, expected operating profits must be negative for each of the three organizational forms, and so exit is the best available strategy. For $R = R_a$, outsourcing in the South achieves zero expected profits. Next comes a range of revenue levels between R_a and R_b for which outsourcing is profitable and the best of the three alternatives. As we just described, outsourcing offers low costs to the principal and affords the opportunity for her to structure high-powered

16

incentives for the entrepreneur without sharing rents. When $R = R_b$, the expected profits for FDI match those for outsourcing and exceed those available to a firm that produces its own components in the North. In the range between R_b and R_v, the principal chooses FDI over outsourcing, because she values the ability to monitor the manager on a fraction δ_S of the tasks. But in this range, the benefit of lower costs still outweighs the cost of less monitoring in the comparison between FDI and in-house production at home.

The figure shows $R_v = R_b + [(\bar{s}_N + c_N) - (\bar{s}_S + c_S)]/[(\delta_N - \delta_S)(1 - h_1)]$ to be the revenue level at which FDI and in-house production in the North yield equal expected profits. At $R = R_v$, the cost savings that favor FDI are matched by the benefits from closer monitoring of the manager's efforts. Since potential revenues are reasonably high in the range above R_v, the principal places great importance on achieving a high probability that the project will succeed. She is willing to pay more for inputs and to compensate the manager more handsomely in order to mitigate the damage caused by the manager's moral hazard.

Finally, highly productive principals prefer foreign outsourcing to in-house production in the North and to FDI. Specifically, when θ is such that $R(\theta) > R_k$, where

$$R_k = \frac{\frac{1-e_1}{1-h_1} - [\delta_N + (1 - \delta_N)e_1 - (\bar{s}_N + c_N)]}{1 - [\delta_N + (1 - \delta_N)h_1]},$$

the principal achieves higher expected operating profits by outsourcing than by opening a local subsidiary.[16] The reason is that when very large revenues are at stake the principal will do whatever is necessary to ensure maximal effort by the entrepreneur or manager and a high probability of success. To induce an effort of $e = 1$ on all tasks, the principal must share rents with the entrepreneur or manager. But the principal foregoes fewer rents with outsourcing than with (any form of) in-house production, because the input bill that is initially paid by the entrepreneur serves as a tax on his take of rents.[17]

---

[16]There is no guarantee that $R_k > R_v$ for all parameter values. If δ_N is not very much larger than δ_S or if $c_N + \bar{s}_N$ is very much greater than $c_S + \bar{s}_S$, then there will be no values of R for which the principal prefers in-house production in the North to both FDI and foreign outsourcing.

[17]Note that, when the principal induces the entrepreneur to exert the maximal level of effort her operating profits are given by $R - (1 - e_1)/(1 - h_1)$, which is independent of the size of the production costs and the size of the entrepreneur's outside option. For this reason the principal could do just as well by outsourcing in the North as in the South when R is large. We prefer to think of outsourcing as taking place in the South nonetheless, based on the analysis in the appendix to Grossman and Helpman (2002b). There we show that a principal who can vary the scale of production strictly

To summarize, we have shown that the firms in an industry have different incentives to open foreign subsidiaries and to engage in foreign outsourcing depending on their productivity levels. Our model suggests that the least and most productive firms will turn to external suppliers for the component needs, while the firms that operate foreign subsidiaries will be less productive than those that manufacture their own components in a plant nearer to their headquarters.

4 Industry Equilibrium

In this section, we embed our model of a firm's choice of organizational mode and location in a setting of industry equilibrium. We assume that principals can enter the industry by bearing a fixed entry cost of f. Those that pay this cost draw a productivity level from a known distribution $G(\theta)$, just as in Melitz (2002) and Helpman, Melitz and Yeaple (2003). The firms then choose their organizational form (including location) and design an optimal purchase or employment contract in the light of their decision. Firms that are successful in acquiring components manufacture differentiated products that compete for consumers' spending.

We assume a world populated by many consumers, each with the utility function $u = y_0 + y^\eta/\eta$, where y_0 is consumption of a homogenous good and y is an index of consumption of the varieties of the differentiated product.[18] The elasticity of demand for the group of differentiated products (with respect to an ideal price index) is $1/(1-\eta)$, where $\eta < 1$. We aggregate consumption of individual varieties using the familiar, CES preference function

$$y = \left[\int y(j)^\alpha\, dj\right]^{1/\alpha}, \ \ 0 < \alpha < 1 \ ,$$

where $y(j)$ is consumption of variety j. With this specification, $1/(1-\alpha)$ is the elasticity of demand for variety j and a higher value of α implies greater substitution across varieties and a higher demand elasticity for each brand. We assume $\alpha > \eta$, so that the varieties substitute more closely for one another than does the group of

---

prefers to outsource in the South rather than in the North at high levels of productivity. Such a principal leaves no rents to the entrepreneur, and so can take advantage of the lower production costs in the South by demanding greater output.

[18]This measure of utility does not include the sebarable component of utility loss associated with the effort that an individual exerts on her job. Note that we could accomodate many industries with differentiated products by assuming $u = y_0 + \sum_l \omega_l y_l^{\eta_l}/\eta_l$; then our analysis here would apply to any industry l.

differentiated products substitute for the numeraire good.[19]

We normalize the measure of consumers to equal one. Then the aggregate world demand for variety j is described implicitly by the first-order condition

$$y^{\eta-\alpha}y(j)^{\alpha-1} = p(j),$$

where $p(j)$ is the price of brand j. A firm's revenue from selling brand j is $y^{\eta-\alpha}y(j)^{\alpha}$.

If a firm with productivity θ is successful in obtaining components, it can produce θ units of output and accrue revenue of

$$R(\theta) = y^{\eta-\alpha}\theta^{\alpha}. \tag{5}$$

Clearly, a firm's potential revenue is the product of two components, a component that depends on aggregate industry characteristics, summarized by y, and an idiosyncratic component, θ. The cumulative distribution function of θ, $G(\theta)$, induces a cumulative distribution on R.

Equation (5) can be used to map each revenue cutoff level in Figure 4 into a productivity cutoff level. For example, since R_v is the revenue level at which a principal is just indifferent between in-house production in the North and FDI, $\theta_v = R_v^{1/\alpha}y^{(\alpha-\eta)/\alpha}$ is the productivity level that generates such indifference. And similarly for the revenue levels R_a, R_b, and R_k, and the corresponding productivity levels θ_a, θ_b, and θ_k. Since $\alpha > \eta$, all of these productivity cutoff levels are increasing in the index of industry consumption.

We can now calculate the expected operating profits facing a principal prior to entry, i.e., before she learns her productivity level. Let $\Pi_{ij}(R)$ denote the operating profits for a firm with revenue R that operates with mode of organization i ($i = o$ or v) and locates parts production in country j ($j = S$ or N). If a principal draws a productivity level below $\theta_a = R_a^{1/\alpha}y^{(\alpha-\eta)/\alpha}$, she will forego the opportunity to engage a supplier or hire a manager and earn zero operating profits. If her productivity level falls between θ_a and θ_b, she will choose to buy components from a supplier in the South (as we know from our earlier discussion) and earn operating profits of $\Pi_{oS}(y^{\eta-\alpha}\theta^{\alpha})$. If productivity falls between θ_b and θ_v, FDI will be the chosen mode of organization, with expected profits of $\Pi_{vS}(y^{\eta-\alpha}\theta^{\alpha})$. In-house production in the North is indicated for

---

[19]The discussion in this section does not require specific functional forms for the utility function. We use these forms for convenience only. An exponential function $y(j)^{\alpha}$ is, however, needed in the next section to derive closed-form solutions for the market shares.

productivity levels between θ_v and θ_k, with resulting expected profits of $\Pi_{vS}\left(y^{\eta-\alpha}\theta^\alpha\right)$. Finally, for high levels of productivity above θ_k, the principal opts for outsourcing in the South and earns expected profits of $\Pi_{oS}\left(y^{\eta-\alpha}\theta^\alpha\right)$. In equilibrium, the expected operating profits for a principal prior to entry match the entry cost, f. We write the free-entry condition as

$$
\int_{R_a^{1/\alpha}y^{(\alpha-\eta)/\alpha}}^{R_b^{1/\alpha}y^{(\alpha-\eta)/\alpha}} \Pi_{oS}\left(y^{\eta-\alpha}\theta^\alpha\right) dG\left(\theta\right) + \int_{R_b^{1/\alpha}y^{(\alpha-\eta)/\alpha}}^{R_v^{1/\alpha}y^{(\alpha-\eta)/\alpha}} \Pi_{vS}\left(y^{\eta-\alpha}\theta^\alpha\right) dG\left(\theta\right)
$$

$$
+ \int_{R_v^{1/\alpha}y^{(\alpha-\eta)/\alpha}}^{R_k^{1/\alpha}y^{(\alpha-\eta)/\alpha}} \Pi_{vN}\left(y^{\eta-\alpha}\theta^\alpha\right) dG\left(\theta\right) + \int_{R_k^{1/\alpha}y^{(\alpha-\eta)/\alpha}}^{\infty} \Pi_{oS}\left(y^{\eta-\alpha}\theta^\alpha\right) dG\left(\theta\right) = f \ .
$$

$$(6)$$

Expected operating profits for a potential entrant are a strictly decreasing function of y, which means that there is a unique index of industry consumption that delivers expected profits equal to the entry cost.

We are interested in how falling trade costs (and other changes in the production environment) affect the relative prevalence of the different modes of organization. For this, we need to define measures of relative prevalence. We could measure this in terms of the numbers of components produced by different sorts of entities, by the output of final goods that embody components produced in different entities, or by the revenues collected by firms of the different types. Fortunately, all of these measures yield similar answers to the questions of interest, so we can focus on just one. We shall measure relative prevalence in the industry by the shares of components that are manufactured by arms-length suppliers, by foreign subsidiaries, and by in-house parts divisions located in the North.

We let X denote the total output of components manufactured by producers of all types and let n denote the number of principals that enter the industry. A fraction $G(\theta_b) - G(\theta_a)$ of the entrants draw productivity levels between θ_a and θ_b, which means that they engage in outsourcing. Of these, a fraction h_1 is successful in acquiring components, because principals with productivity levels in the indicated range induce their outsourcing partners to exert effort of e_1. It follows by the law of large numbers that $n[G(\theta_b) - G(\theta_a)]h_1$ units of components are produced by the supplier firms of principals with productivities in this range. Southern subsidiaries of multinational corporations produce a total of $n[G(\theta_v) - G(\theta_b)][\delta_S + (1 - \delta_S)h_1]$ units of components,

because a fraction $G(\theta_v) - G(\theta_b)$ of principals draw productivity levels that make hiring a Southern manager the optimal strategy, and these managers are induced to exert maximal effort on tasks that can be monitored and intermediate effort of e_1 on tasks that are not observable. By similar reasoning, $n[G(\theta_k) - G(\theta_v)][\delta_N + (1 - \delta_N)h_1]$ is the number of units of components produced in-house by firms with a parts division located in the North. Finally, a fraction $1 - G(\theta_k)$ of entrants draw productivity levels above θ_k. These firms engage in outsourcing and all succeed in acquiring the needed components by inducing their partners to exert maximal effort. The resulting number of components is $n[1 - G(\theta_k)]$. The total output of components is the sum of these numbers, or

$$
\begin{aligned}
X \;=\; & n[G(\theta_b) - G(\theta_a)]h_1 + n[G(\theta_v) - G(\theta_b)][\delta_S + (1 - \delta_S)h_1] \\
& + n[G(\theta_k) - G(\theta_v)][\delta_N + (1 - \delta_N)h_1] + n[1 - G(\theta_k)]. \quad (7)
\end{aligned}
$$

We can now readily compute the shares of components produced under the different modes of organization. Let σ_o represent the share of components produced by outsourcing partners, σ_{vS} represent the share produced by firms that engage in FDI, and σ_{vN} represent the share produced in-house in the North. Then

$$
\sigma_o = \frac{n}{X} \left\{ [G(\theta_b) - G(\theta_a)]h_1 + [1 - G(\theta_k)] \right\} , \quad (8)
$$

$$
\sigma_{vS} = \frac{n}{X} \left\{ [G(\theta_v) - G(\theta_b)][\delta_S + (1 - \delta_S)h_1] \right\} , \quad (9)
$$

and

$$
\sigma_{vN} = \frac{n}{X} \left\{ [G(\theta_k) - G(\theta_v)][\delta_N + (1 - \delta_N)h_1] \right\} . \quad (10)
$$

Notice that the market shares do not depend on the number of principals that enter, because n/X is independent of n.[20]

---

[20] The equilibrium number of entrants n can be solved using the definition of y, the equilibrium output of differentiated products of the various types, and the free-entry condition (6), which determines a unique equilibrium level of y.

5 Improved Monitoring in the South

In this section, we show how the model can be used to investigate the effects of changes in the production environment on the relative prevalence of different modes of organization. We consider an increase in δ_S, which is the fraction of tasks undertaken by a division manager in the South that can be monitored by the principal in the North. Such a gain in monitoring possibilities might result from improvements in communications technology or perhaps from changes in the legal system.

In order to derive specific results, we need to make additional assumptions about the distribution of productivity levels among potential entrants. There is evidence to suggest that this distribution is well approximated by a Pareto distribution, so rather than derive sufficient conditions for $G(\cdot)$ under which certain results may hold, we choose to work with this simple functional form.[21] We henceforth assume that $G(\theta) = 1 - \theta^{-\beta}$, for $\theta \geq 1$ and $\beta > 1$. With this distribution, the minimum productivity level for a potential entrant is one and there is no limit on the maximum productivity level.

As we have noted previously, an increase in δ_S rotates the dotted line in Figure 3 in a counterclockwise direction around point Q. It therefore has no effect on R_b, the level of potential revenue at which outsourcing and FDI yield the same expected profits. It does, however, cause R_v to rise, which means that the principal prefers FDI to in-house production in the North for a wider range of potential revenues. As δ_S increases, there is no change in R_a or R_k, as the ability to monitor managers in the South has no bearing on the profitability of outsourcing in the South or on the relative profitability of outsourcing versus in-house production in the North.

For given y, an improvement in a principal's ability to monitor managers in the South must increase the expected operating profits of potential entry, the term on the left-hand side of (6). This means that additional principals will enter the industry (n rises), causing y to grow as well. However, the induced growth in the number of entrants and the consumption index have no direct effect on the market shares.

The growth in y will also cause an upward shift in all of the cutoff productivity levels, i.e., in θ_a, θ_b, θ_v, and θ_k. This is because, with greater industry competition

---

[21]See, for example, Axtell (2001), who provides evidence that the Pareto distribution fits well the distribution of sales by firm in the United States. Helpman, Melitz and Yeaple (2002) have shown how a Pareto distribution of firm sizes will emerge from a Pareto distribution of productivity levels. They also verify that Axtell's aggregate results hold for all but a handful of the 52 industries in their data set.

as measured by the index of industry output, a given firm must have higher productivity itself to achieve the same level of revenues as before. However, all of the cutoff productivity levels appear similarly in the numerators and denominators of the expressions for market share; each such term is raised to the power of $-\beta$. It follows that the aggregate output of a component type (the numerators in the expressions for market shares) and the aggregate output of all components (the denominators in the expressions for market shares) grow by the same proportion as y increases and thus the aggregate consumption index has no indirect effect on the market shares as well.

We are left with two effects of an increase in the fraction of tasks that can be monitored in the South. An increase in δ_S generates an increase in the probability that any given parts division in the South will be successful in producing components, because managers of Southern subsidiaries devote more effort to tasks that are monitored than to those that are not. And an increase in δ_S causes a wider range of principals to locate their parts divisions in the South, as R_v rises. The former effect tends to increase X/n, the average number of components produced per entrant, while the latter effect tends to reduce it (since parts divisions in the South succeed less often than those in the North). On net, however, the former effect dominates with a Pareto distribution of productivity levels, and so the average output of components per entrant rises.[22]

Now we are ready to discuss the shifts in the market shares. Using our findings on the effects of a rise in δ_S on the cutoff points R_v and on average output X/n, it is straightforward to show that the relative prevalence of outsourcing in the South and of in-house production in the North fall, and that the latter falls by a greater percentage (see Grossman and Helpman 2002b). Since σ_o and σ_{vN} both fall, it must be that σ_{vS} rises, i.e., the output of components by Southern subsidiaries grows by more than the aggregate output of components. Thus, an increased ability to monitor managers in the South makes in-house production in the South a more attractive option relative to both of the alternative modes of organization.

Our results in this section are broadly consistent with some recent empirical evidence on the determinants of the form of foreign investment and the extent of vertical specialization in multinational firms. For example, Lin and Png (2002) examine the form of FDI undertaken by 148 Taiwanese firms that made investments in China be-

---

[22]To justify this claim, we totally differentiate $X/\left[ny^{\beta(\eta-\alpha)/\alpha}\right]$ with respect to δ_S, taking account of the fact that $R_v = R_b + \left(\bar{s}_N + c_N - \bar{s}_S - c_S\right)/\left(\delta_N - \delta_S\right)\left(1 - h_1\right)$. The sign of this derivative is the same as the sign of $q^{-\beta/\alpha} - 1 - (1 - q)\beta/\alpha$, where $q = R_b/R_v < 1$. Since $q^{-\beta/\alpha}$ is convex in q, this derivative must be positive.

tween 1987 and 1991. They consider the firms' decision whether to structure the FDI as a joint venture or as a wholly-owned subsidiary, taking distance from Hong Kong as a proxy for the principal's ability to monitor the local agent. A joint venture is more like outsourcing in our model, inasmuch as the local entrepreneur has a greater stake in the project than does the local manager of a wholly-owned subsidiary. Lin and Png find that joint ventures are more likely to be chosen as distance from Hong Kong increases, which is in keeping with our finding that the range of (low-productivity) firms that choose outsourcing over FDI expands as δ_S declines. In related work, Hanson et al. (2002) examine the determinants of the extent of foreign affiliate processing of inputs imported from a parent U.S. firm. After controlling for trade barriers and transportation costs between the parent and subsidiary, they find that the extent of such vertical specialization is decreasing in the distance between parent and subsidiary, is greater for subsidiaries located in English speaking countries than those that are not, and is higher in Mexico and Canada then would be predicted based on distance (and the other variables) alone. All of these findings can be viewed as consistent with our prediction that an increase in δ_S (proxied by distance, common language, and adjacency) raises the relative profitability of FDI compared to other modes of organization.[23]

6   Falling Trade Costs

In this section, we study how falling trade costs affect the international organization of production. Both trade liberalization and declining transportation costs have contributed in recent years to the globalization of economic activity. We are interested in whether and under what circumstances a decline in trade costs will favor one mode of organization over another.

To examine this issue, we first must extend our model to include trade costs. We assume that transporting a component from South to North entails a per unit cost of τ. This cost may reflect a shipping charge, or it may result from an import tariff imposed by the government of the North. The same cost applies regardless of whether the component is traded within the firm (as when the part is produced in a foreign subsidiary) or at arms-length (as when a firm arranges for delivery of parts from a

---

[23]Hanson et al. (2002) also find that the cost share of intermediate inputs imported by a foreign affiliate from its U.S. parent for further processing is decreasing in the wage of unskilled labor in the host country. This too is consistent with the comparative static properties of our model, i.e., that the market share of FDI falls with c_S.

24

Southern supplier). In case of foreign outsourcing, the contract term p_o now refers to the amount that is paid to the foreign entrepreneur in the event that he delivers suitable components to the principal's assembly plant in the North.[24]

The presence of trade costs modifies the relationship between expected operating profits and potential revenues. Consider first a principal with low productivity for whom foreign outsourcing will be the preferred mode of organization if she chooses to operate at all.[25] Such a principal designs a contract under which the expected utility of the foreign entrepreneur is just equal to $\bar{s}_S$, the utility he could achieve by pursuing his outside option. This means that the incidence of the trade costs falls on the principal. Since we have normalized the number of components she can process to equal one, τ is the total trade cost she will bear in the event that the foreign entrepreneur is able to deliver the components. Thus, expected operating profits from outsourcing are $\Pi_{oS}(R) = (R-\tau)h_1 - e_1 - (\bar{s}_S + c_S)$ for $R \leq R_b$. The minimum potential revenues necessary for such a principal to enter into any contract with a supplier are $R_a = \tau + (e_1 + \bar{s}_S + c_S)/h_1$.

At R_b, the principal is indifferent between outsourcing and engaging a manager to head a subsidiary in the South. In the event of FDI, the principal bears any trade costs that arise. Expected operating profits in this mode of organization and with the optimal employment contract for the manager are $\Pi_{vS}(R) = (R-\tau)[\delta_S + (1-\delta_S)h_1] - [\delta_S + (1-\delta_S)]e_1 - (\bar{s}_S + c_S)$, for R between R_b and R_v. From $\Pi_{oS}(R_b) = \Pi_{vS}(R_b)$ we find that $R_b = \tau + (1-e_1)/(1-h_1)$ in the presence of trade costs.

Costs that arise from importing components do not affect the potential profitability of in-house production in the North for a given R. Thus $\Pi_{vN}(R) = R[\delta_N + (1-\delta_N)h_1] - [\delta_N + (1-\delta_N)]e_1 - (\bar{s}_N + c_N)$ for R between R_v and R_k, as before. We calculate the new value of potential revenues that makes the principal indifferent between in-house production in the South and in the North, and find

$$R_v = \frac{1-e_1}{1-h_1} + \frac{(\bar{s}_N + c_N) - (\bar{s}_S + c_S) - \tau[\delta_S + (1-\delta_S)h_1]}{(\delta_N - \delta_S)(1-h_1)}.$$

Finally, when potential revenues are sufficiently great, a principal will find it optimal to outsource the production of components. Moreover, she will opt to provide sufficient

---

[24] Alternatively, the principal might choose to make the account payable upon delivery of components in the foreign country. It can be shown, however, that at all productivity levels, a principal prefers (at least weakly) to set a c.i.f. price rather than a f.o.b. price.

[25] We assume that the trade costs are not so large as to make domestic outsourcing a more attractive option than foreign outsourcing.

incentives for the foreign entrepreneur to exert maximal effort, thereby creating rents for the entrepreneur. The trade costs, like the cost of the inputs used to produce the components, are borne initially by the foreign entrepreneur, since the contingent payment p_o is made only when the goods are delivered to the principal in the North. When the entrepreneur capture rents, the incidence of the trade costs falls on him. This means that the principal achieves the same profits from outsourcing as when trade costs are zero; namely $\Pi_{oS}(R) = R - (1 - e_1)/(1 - h_1)$ when $R \geq R_k$. The potential revenues R_k at which in-house production in the North and foreign outsourcing yield equal expected profits also is the same as before.

Figure 5:

We consider now the effects of a fall in trade costs τ. In Figure 5, the dotted line indicates that when τ declines, operating profits rise for firms that outsource in the South and for those that undertake FDI. The point at which principals elect to engage a supplier rather than exit the market shifts from R_a to R'_a, because principals are able to make positive operating profits for a wider range of revenue levels when trade costs are lower. The level of potential revenues at which principals are indifferent between FDI and outsourcing also falls (from R_b to R'_b). This reflects the fact that a fall in trade costs boosts $\Pi_{vS}(\cdot)$ by more than it does $\Pi_{oS}(\cdot)$ at a given level of potential revenues, inasmuch as a multinational achieves a higher probability of successfully producing components than does an arms-length supplier; thus, the expected cost savings is greater for the former than the latter.[26] The fall in trade costs makes the South a relatively more attractive location for in-house production of components, so R_v rises. Note again that R_k is unaffected.

It follows readily from this discussion that the market share of components produced by vertically integrated firms in the North falls.[27] It is hardly surprising that a reduction in trade costs increases the market share of imported components. But it

---

[26]The probability of success for a multinational firm operating in the South is $\delta_S + (1 - \delta_S)h_1$, which exceeds h_1, the probability of success for an independent supplier.

[27]Using (7), (10), $G(\theta) = 1 - \theta^{-\beta}$ and $\theta_i = R_i^{1/\alpha} y^{(\alpha-\eta)/\alpha}$ for $i = a, b, v,$ and k, we can write the market share of components produced in-house in the North as

$$\sigma_{vN} = \frac{\left(R_k^{-\beta/\alpha} - R_v^{-\beta/\alpha}\right)[\delta_N + (1 - \delta_N)h_1]}{\Omega}$$

remains to be seen whether this reflects an expansion in the share of components purchased from supplier firms, the share produced by foreign subsidiaries of multinational corporations, or both. We show in Grossman and Helpman (2002b) that the answer to this question depends sensitively on the characteristics of the firms that engage in outsourcing, and in particular on whether these are predominantly low-productivity firms that outsource so as to generate higher-powered incentives for their partners or predominantly high-productivity firms that outsource to conserve on the rents that they pay to their partners. Trade liberalization tends to favor multinational activity in industries in which outsourcing is conducted predominantly by high productivity firms that are seeking to minimize the rents they must share with the head of a parts division in order to induce high levels of effort. In contrast, trade liberalization spurs arms-length trade with suppliers when most outsourcing is undertaken by relatively low productivity firms that are seeking to boost the power of the incentives they can provide to their component producers. In the former case, the fall in trade costs does little to increase the profitability of outsourcing, but firms that engage in FDI realize an immediate cost savings. In the latter case, the principal benefits under either mode of organization involving trade, but outsourcing receives a greater boost because it can expand at the extensive margin.[28]

7   Conclusions

We have developed a model in which the heterogeneous firms in an industry choose their modes of organization and the location of their subsidiaries or suppliers. We assume that the principals of a firm are institutionally or legally constrained in the nature of the contracts they can write with suppliers and employees. In particularly, a supplier

where

$$
\begin{aligned}
\Omega \;=\; & \left(R_a^{-\beta/\alpha} - R_b^{-\beta/\alpha} \right) h_1 + \left(R_b^{-\beta/\alpha} - R_v^{-\beta/\alpha} \right) [\delta_S + (1-\delta_S)h_1] \\
& + \left(R_k^{-\beta/\alpha} - R_v^{-\beta/\alpha} \right) [\delta_N + (1-\delta_N)h_1] + R_k^{-\beta/\alpha} \; .
\end{aligned}
$$

Then, the fact that $dR_a/d\tau = dR_b/d\tau > 0$, $dR_v/d\tau < 0$, and $dR_k/d\tau = 0$ implies $d\sigma_{vN}/d\tau > 0$.

[28]The last result depends importantly on the distribution of productivity. Note from Figure 5 that falling trade costs reduce the revenue level at which a low productivity firm is just indifferent between exiting and outsourcing, as well as the revenue level at which a firm is just indifferent between outsourcing in the South and engaging in FDI. In fact, both cutoff points decline by the same amount. With a Pareto distribution, however, the lower revenues have more weight, and therefore the market share of outsourcing firms in the low productivity range rises.

cannot be asked to post a (large) bond that will be forfeited in the event that the firm's efforts to produce suitable components prove unsuccessful. Similarly, a manager cannot be asked to pay a fine if his division performs poorly, nor can he be asked to pay personally for the inputs used by his division. In such an institutional setting, the contracts that principals can write with their suppliers and division managers may not induce efficient levels of effort on all relevant tasks.

We identified two reasons why a principal may benefit from engaging an external supplier to manufacture components in a setting like this. First, the principal can confront an agent with higher-powered incentives when the agent has more at stake. A supplier can be made to front the cost of the inputs needed to manufacture components and so can be given a greater stake in the project than a manager. Second, when a principal finds it desirable to induce a very high level of effort from her agent, the cost to the principal of providing the necessary incentives is less for an outside supplier than for an employee. Again, the input costs play a key role in this. We showed that principals must leave rents to their agents when they induce the highest level of effort. But the rents are smaller for an entrepreneur than for a manager, because the principal can pass along input costs to the former but not the latter.

Against the benefits of outsourcing, there is an advantage to in-house production that stems from the greater opportunity it affords the principal to monitor the actions of her agent. We assume that the ability to monitor an agent declines with distance; a vertically integrated firm is able to observe a division manager's actions on more tasks when the division is located near the headquarters than when it is located in a different country. Thus, FDI suffers the disadvantage of lesser monitoring compared to in-house production near the headquarters, but the possible advantage of lower costs.

Our main result concerns the sorting of firms in an industry into different organizational forms. The least productive firms that are active in equilibrium choose to subcontract the production of components to suppliers in the South. For these firms, the ability to offer higher-powered incentives with outsourcing weighs most heavily. Firms with intermediate levels of productivity opt for vertical integration, with the less productive of these undertaking foreign investment in the South and the more productive operating a parts division in the North. FDI does not appeal as much to the more productive firms, because the ability to monitor a manager's efforts becomes more valuable as potential revenues rise. Finally, outsourcing is the preferred option for the most productive firms in an industry, because the principals of these firms who

want to induce a high level of effort are able to pass along input costs to a supplier but not to a manager.

We used our model to examine the implications of falling trade costs for the relative prevalence of the different organizational modes. An important observation is that trade liberalization may promote mostly FDI or mostly outsourcing, depending on the characteristics of an industry. In particular, the market share of imports from suppliers will expand as trade costs fall if most of the outsourcing is undertaken by low productivity firms in which the principals are motivated by a desire to give their agents a greater stake in the venture. But the market share of imports from suppliers will contract as trade costs fall if most outsourcing is undertaken by high productivity firms in which the principals are motivated by a desire to minimize the rents captured by their agents. The equilibrium sorting of firms by productivity level plays an important part in delivering these conclusions.

References

Axtell, R.L., 2001. Zipf distribution of U.S. firm sizes. Science 293, 1818–1820.

Borga, M., Zeile, W.J., 2002. Foreign outsourcing and the intrafirm trade of U.S. multinational corporations. U.S. Bureau of Economic Analysis, manuscript.

Campa, J., Goldberg, L., 1997. The evolving external orientation of manufacturing: A profile of four countries. Federal Reserve Bank of New York Economic Policy Review 4, 53–81.

Ethier, W.J., 1986. The multinational firm. Quarterly Journal of Economics 101, 805–833.

Ethier, W.J., Markusen, J.R., 1996. Multinational firms, technology diffusion, and trade. Journal of International Economics 41, 1–28.

Feenstra, R.C., 1998. Integration of trade and disintegration of production in the global economy. Journal of Economic Perspectives 12, 31–50.

Feenstra, R.C., Hanson, G.H., 2003. Global production sharing and rising inequality: A survey of trade and wages. In Harrigan, J., Choi, E.K., (Eds.), Handbook of International Trade. Basil Blackwell Publishers, London (forthcoming).

Grossman, G.M., Helpman, E., 2002a. Outsourcing in the global economy. NBER Working Paper No. 8728.

Grossman, G.M., Helpman, E., 2002b. Managerial incentives and the international organization of production. NBER Working Paper No. 9403.

Hanson, G.H., Mataloni, R.J., Slaughter, M.J., 2001. Expansion strategies of U.S. multinational corporations. Brookings Trade Forum 2001, 245–294.

Hanson, G.H., Mataloni, R.J., Slaughter, M.J., 2002. Vertical specialization in multinational firms. University of California at San Diego, manuscript.

Helpman, E., Melitz, M., Yeaple, S., 2003. Exports versus FDI. NBER Working Paper No. 9439.

Holmström, B., Milgrom, P. 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law, Economics and Organization 7, 24–52.

Horn, H., Lang, H., Lundgren, S., 1995. Managerial effort, incentives, X-inefficiency and international trade. European Economic Review 39, 117–138.

Horstmann, I.J., Markusen, J.R., 1987. Licensing versus direct investment: A model of internalization by the multinational enterprise. Canadian Journal of Economics 20, 464–481.

Hummels, D., Rapoport, D., Yi,K-M., 1998. Vertical specialization and the changing nature of world trade. Federal Reserve Bank of New York Economic Policy Review 5, 79–99.

Hummels, D., Ishii, J., Yi,K-M., 2001. The nature and growth of vertical specialization in world trade. Journal of International Economics 54, 75–96.

Katz, L.F., 1986. Efficiency wage theories: A partial evaluation. NBER Macroeconomics Annual 1, 235–275.

Lin, C-C.S., Png, I., 2002., Monitoring costs and the mode of international investment. National University of Singapore, manuscript.

Markusen, J.R., 2001. Contracts, intellectual property rights, and multinational investment in developing countries. Journal of International Economics 53, 189–204.

Melitz, M., 2002. The impact of trade on intra-industry reallocations and aggregate industry productivity. NBER Working Paper No. 8881. Forthcoming in Econometrica.

Yeats, A.J., 2001. Just how big is global production sharing? In Arndt, S.W., Kierzkowski, H., (Eds.), Fragmentation: New Production Patterns in the World Economy. Oxford University Press, Oxford.