DOI: 10.1111/jems.12564

Revised: 11 October 2023



Two-sided productivity heterogeneity, firm boundaries, and assortative matching

Kanişka Dam¹ 💿 | Daniel Ripperger-Suhler² | Konstantinos Serfes³ 💿

¹Center for Research and Teaching in Economics (CIDE), Mexico City, Mexico

²U.S. Bureau of Economic Analysis, Suitland, Maryland, USA

³School of Economics, LeBow College of Business, Drexel University, Philadelphia, Pennsylvania, USA

Correspondence

Konstantinos Serfes, School of Economics, LeBow College of Business, Drexel University, 3220 Market St, Philadelphia, PA 19104, USA. Email: ks346@drexel.edu

Abstract

We consider a market where each firm is created by the combination of two complementary assets that are heterogeneous in their productivity. After assets match endogenously, their owners choose between two ownership structures: centralized organization (integration) and arm's length organization (nonintegration). Our main focus is on the interplay between productivity heterogeneity and firm boundary decisions. When firms choose between distinct ownership structures, the standard single-crossing condition that guarantees positive assortative matching may fail to hold. We provide a novel condition-the congruent marginal contributions property-which guarantees monotone matching with respect to asset productivity. Furthermore, we provide conditions under which integration at the bottom of the productivity ladder is the market equilibrium; an organizational pattern that has been largely unexplored by the theoretical and empirical literature. We investigate the effect of model primitives on the equilibrium distribution of output. Moreover, our model offers interesting testable implications regarding firm boundary decisions.

1 | INTRODUCTION

Firm productivity heterogeneity is pervasive in almost every industry and is a crucial determinant of industry performance and welfare (e.g., Melitz, 2003). Syverson (2011) identifies a number of factors within and outside organizations that contribute to productivity differences across businesses. Differences in management practices (Bloom & Van Reenen, 2010), productivity levels of inputs (Abowd et al., 2005) and firm boundary decisions, that is, nonintegration versus integration (e.g., Forbes & Lederman, 2011), among others, are the main *internal drivers* of firm productivity differences. Moreover, the composition of input characteristics (e.g., worker skill, managerial talent) is important for firm revenue, and hence, competition for vertically heterogeneous production units induces endogenous sorting (e.g., Garicano, 2000; Teulings, 1995).

Empirical research in this area has benefited tremendously from the recent availability of detailed firm-level data. Given that firm boundary decisions affect the productivity of the enterprises, and endogenous sorting amplifies productivity differences, understanding how heterogeneous firms make decisions and interact with each other in these markets is crucial for assessing industry performance. However, the extant theoretical literature has been largely silent regarding how these aforementioned factors interact to determine firm heterogeneity. We contribute to the literature in

A significant portion of this research was conducted when Daniel Ripperger-Suhler was a Ph.D. candidate at Drexel University.

15309134, 0, Downloaded from https://onlnetibrary.wiley.com/doi/10.1111/jems.12564 by Centro De Investigacion Y Docencia Economicas, A.C., Wiley Online Library on [14/1/2023], See the Terms and Conditions (https://onlinelibrary.wiley.com/enus

and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

-WILEY-A Journal of Economics & Management Strategy

2

this direction. We propose a tractable model to analyze how endogenous sorting of heterogeneous production units and firm boundary decisions interact with each other in determining the heterogeneity among firms in any given industry.

Guided by recent empirical evidence, and building on the framework of Legros and Newman (2013), we posit a parsimonious model where firms are formed by the combination of two complementary assets: A and B. Within each asset side of the market, there is a continuum of producers (units), heterogeneous with respect to the productivity levels. One A producer is matched one-to-one with one B producer to form a "firm" or "enterprise." Given the two-sided heterogeneity of the market, in equilibrium, there is endogenous sorting of productivity types of A and B units, and hence, firm heterogeneity is endogenous.

Within each enterprise the unit managers decide either to work as separate units (decentralized or arm's length organization) or to integrate (centralized organization). In a nonintegrated firm, asset managers retain the decision rights, whereas in an integrated organization the decision rights are conferred to an outsider, called the *headquarter*. Managerial actions are not verifiable, entail private costs, and coordination of managerial actions enhances the likelihood of high revenue. Nonintegration places higher weights on private managerial costs resulting in poor coordination. On the other hand, integration improves coordination between the units, as revenue maximization becomes the objective of the integrated entity; however, private costs are ignored under this organizational mode.¹ Thus, each ownership structure entails its own inefficiency. This description of the firm is consistent with recent empirical evidence (e.g., Atalay et al., 2014), where each organization is a collection of complementary assets (inputs) and the main motive of integration is to facilitate a more efficient transfer of these inputs within the organization, that is, better coordination (we review the empirical literature below).

Given that each side of the market is vertically heterogeneous with respect to productivity, a natural question is whether the productivity types are matched following a positive assortative matching (PAM) pattern. Legros and Newman (2007) show that the *generalized increasing differences* (GID) condition of the bargaining frontier is necessary and sufficient for PAM when utility is imperfectly transferable. However, when each firm chooses between two distinct ownership structures, GID may fail to hold. This is due to the bargaining frontier of a firm being the maximum of two frontiers, each one associated with a given ownership structure.

We provide a novel sufficient condition, which we term the *congruent marginal contributions* property, that guarantees GID, and hence, PAM. Under both nonintegration and integration, a higher productivity type of each unit increases the surplus. However, whether the equilibrium matching is positive assortative depends upon the organizational form to which a higher type contributes more. The congruent marginal contributions property asserts that both *A* and *B* units gain by switching to the *same* ownership structure when there is organizational restructuring. This property along with GID under each organizational mode induce GID of the combined bargaining frontier. Consequently, the equilibrium of the supplier market exhibits PAM irrespective of the equilibrium choices of ownership structures.

Discrete choice is a common feature in many organizations after a match is formed. There are many examples in contexts other than our model—a matched firm-worker pair or two cofounders may need to choose one of many different projects, a matched couple may need to decide on whether or not to have children or who (if anybody) participates in the labor market, or in an entrepreneur-investor pair, the entrepreneur with two independent projects may require to decide between separate or joint financing (conglomerate) of the projects.² Our paper is the first to derive the sufficient conditions under which assortative matching is preserved in the presence of such discrete choices.

After establishing the existence of assortative matching, we then examine how the choice of ownership structures interacts with PAM to contribute to firm heterogeneity. We concentrate on monotone equilibria wherein either (a) low productivity firms stay separate and high productivity firms integrate, or (b) low productivity firms integrate and high productivity ones choose nonintegration. We call the former type of monotone equilibrium $\langle N, I \rangle$, and the latter type, $\langle I, N \rangle$.

The emergence of different types of monotone equilibria depend crucially on the relative bargaining power of the two units in each firm, which in turn affects whether the shares of the surplus between the two units is balanced or unbalanced. In the former case, coordination is achieved even without integration, while in the latter case integration is necessary for better coordination. As is typical in a model of endogenous sorting, the bargaining power of each unit in a given match is a function of the bargaining power distribution in all the matches located below this specific match on the productivity ladder. How the surplus is allocated in all these matches (firms) depends on the initial condition (reservation utility, i.e., the utility that accrues to an asset if not employed in the industry under consideration) and how the shares (utilities) evolve as we move up the productivity ladder, which is a function of the types on each side of the market.

ournal of Economics & -WILF

3

Our analysis offers theoretical support for an organizational pattern where, despite the fact that there is assortative matching with respect to productivity, integration is the preferred choice only at the bottom of the productivity ladder, $\langle I, N \rangle$. Empirical research so far (e.g., Atalay et al., 2014; Corcos et al., 2013) has focused on the opposite pattern, integration at the top, $\langle N, I \rangle$, but with no rigorous market equilibrium theoretical model in the background. Our results offer a novel testable implication for future empirical research (integration at the bottom), but can also be used to justify existing findings (integration at the top).

Our model also sheds light on output heterogeneity and "inequality" in the market. Holding the ownership structure fixed, PAM of productivity types implies that as we move up the productivity ladder expected firm output increases. On the other hand, holding firm productivity fixed, the expected output of an integrated firm is higher than that of a nonintegrated firm (due to better coordination under integration). Therefore, PAM and equilibrium organizational modes *together* contribute to firm heterogeneity, which we measure by different moments of the equilibrium distribution of the expected output—namely, the variance and the skewness (as in Kehrig, 2015). Because of the imperfect transferability of firm surplus, our model cannot be solved analytically to obtain closed-form solutions. To confirm the significance of the equilibrium choices of ownership structures in determining the output distributions, in a numerical exercise, we show that the variance and skewness of the equilibrium expected output are higher under $\langle N, I \rangle$ than under $\langle I, N \rangle$. Integration at the top of the productivity ladder, as in the first type of equilibrium, coupled with the fact that integrated firms produce more output distribution (and consequently, firm revenue). On the other hand, integration at the bottom, as in the second type of equilibrium, is a countervailing force that compresses the equilibrium distribution of output.

We further offer important testable implications pertaining to the roles of the reservation utility and relative productivity of the supplier units in determining firm boundary decisions as well as firm heterogeneity. By relating reservation utility to the specificity of assets in a given industry, our results imply that while intermediate asset specificity induces integration at the top, extreme asset specificity may lead to nonintegration at the top. Moreover, the former implies greater inequality in terms of firm productivity. The first set of implications is related to the previous research on the relation between asset specificity and the incidence of vertical integration (e.g., Williamson, 1981). Additionally, recent empirical measures of asset specificity (e.g., Kermani & Ma, 2021) allow us to identify industries wherein firms are more likely to be vertically integrated.

In sum, the main contribution of our paper is twofold. First, we provide a simple sufficient condition for positive sorting in an environment where matched pairs face binary choices, which is useful in many settings beyond the firm boundary decision as in our model. The second important contribution is that firm heterogeneity and organizational form are endogenous and lead to empirically relevant equilibrium implications.

Our paper is related to those of Legros and Newman (2013) and Dam and Serfes (2021) who contribute to the recent literature on Organizational Industrial Organization (OIO) which is concerned with how market structure affects firm boundary decisions. In Legros and Newman (2013), there is no productivity heterogeneity, while in Dam and Serfes (2021) there is only one-sided heterogeneity. Thus, in both papers the issue of assortative matching does not arise. Furthermore, both the aforementioned papers analyze the structure of the product market (e.g., competition) as an *external* driver of enterprise heterogeneity in the sense that an exogenous change in market price affects firms' decision about whether to integrate or stay as separate production units. In contrast, the present paper focuses on the *internal* drivers of productivity heterogeneity, in particular, how differences in managerial practices translate into differences in firm productivity. Our results conform to recent empirical evidence on sorting of productivity in vertically or horizontally related markets.

1.1 | Related empirical literature

Recent empirical literature presents evidence in support of the main ingredients of our theoretical model—namely, enhanced coordination through firm boundary decisions, and endogenous sorting. The airline industry example in footnote 1 offers clear evidence of the trade-off between integration and nonintegration, where regional airlines often operate as "subcontractors" for major network carriers to connect smaller cities to the major airline hubs. Forbes and Lederman (2011) show that airlines that own their regional affiliates (integration) experience shorter delays and fewer cancellations than those who outsource services (nonintegration). As setting out the decision rights within the organization clearly improves coordination, an integrated firm copes better with unforeseen adversities. However, this

4

enhanced coordination comes at an expense—mainly the higher transaction costs of integration (e.g., higher wages for the employees to compensate for changes in the their daily routine introduced by integration resulting in higher private costs).³ Grossman and Helpman (2002), Hart and Hölmstrom (2010), and Legros and Newman (2013) posit tractable models of firm boundary decisions, on which we build, where such trade-offs naturally emerge. In the context of international trade, Antràs and Helpman (2004) find that low-productivity firms outsource (nonintegration) whereas high-productivity ones insource (integration).

When heterogeneous production units match with each other in vertically or horizontally related markets, evidence on positive sorting is also widespread. Dragusanu (2014) analyzes matched importer–exporter data to show that more capable Indian manufacturing suppliers (exporters) match with more capable US buyers (importers) following a PAM pattern with capability being proxied by firm size. Benguria (2021) finds similar evidence from United States–Colombia trade relations—more productive Colombian distributors import from more productive US suppliers. Here, the firm productivity is measured by the residual revenue. Both Dragusanu (2014) and Benguria (2021) consider models of search and matching where firms invest in costly search for trading partners. In contrast, Sugita et al. (2021) consider a classical matching model like ours where firms can seek alternative trading partners (partner switching). They also find evidence of PAM, that is, more productive Mexican suppliers are matched with more productive US distributors. Atalay et al. (2014) analyze the role of productivity differences in the context of firm boundary decisions. Vertically integrated firms are not only larger and more productive on average, but also the authors present evidence of positive sorting of productivity and size of both upstream and downstream firms. In the context of horizontal mergers, Braguinsky et al. (2015) find evidence of PAM among heterogeneous firms with respect to productivity in the Japanese cotton spinning industry.

The rest of the paper is organized as follows. We present the model in Section 2. In Section 3, we focus on an arbitrary firm and we analyze its choice between integration and nonintegration. We analyze the equilibrium of the supplier market in Section 4. Then, in Section 5, we discuss the empirical relevance of our results. We conclude in Section 6.

2 | A MATCHING MODEL OF FIRM BOUNDARY DECISIONS

2.1 | Technology and matching

Consider a supplier market that consists of two sides, where on each side there is a continuum of production units of measure 1. Producers are vertically differentiated with respect to productivity. In particular, let $J_A = [0, 1]$ be the set of "*A* units" on the one side of the market and $J_B = [0, 1]$, the set of "*B* units" on the other side. Each unit $i \in J_A$ is assigned a type or "productivity" $a = a(i) \in A$ and each $j \in J_B$ has an assigned type $b = b(j) \in B$ where the type spaces $A = [\underline{a}, \overline{a}]$ and $B = [\underline{b}, \overline{b}]$ are sub-intervals of \mathbb{R}_{++} . Let G(a) be the fraction of *A* units with productivity lower than *a*, that is, G(a) is the (cumulative) distribution function of *a* with the associated density function g(a) > 0 for all $a \in A$. Likewise, let F(b) be the distribution function of *b* with the associated density function f(b) > 0 for all $b \in B$.

Production of a homogeneous consumer good requires one *A* and one *B* unit who are matched one-to-one to form a "firm" or "enterprise". All decisions and payoffs of each enterprise depend only on the types of the two participating units, and hence, (a, b) denotes a typical firm. A matching is a one-to-one mapping $\mu : B \to A$ which assigns to each $b \in B$ a type $a = \mu(b) \in A$. The firms may include lateral or vertical relationships. The stochastic output or revenue of a firm (a, b) is given by⁴:

$$\tilde{y}(a,b) = \begin{cases} z(a,b), & \text{with probability } \pi(e_A,e_B) \equiv 1 - (e_B - e_A)^2, \\ 0, & \text{otherwise.} \end{cases}$$

We assume that $z(a, b) \ge 1$ is twice continuously differentiable and increasing for all (a, b).⁵ Each unit must make a noncontractible production decision: $e_A \in [0, 1]$ by an A unit and $e_B \in [0, 1]$ by a B unit. These decisions can be made by the manager overseeing the assets or by someone else. The manager of each unit is risk neutral and incurs private cost for the managerial action. If managers coordinate, that is, $e_A = e_B$, then a firm can reach its full potential in that success is obtained with probability 1. However, coordination is costly for the units. The private cost of an A unit is $C_A(e_A) = e_A^2$ and that of a B unit is $C_B(e_B) = (1 - e_B)^2$. The functional forms of cost functions suggest that there is a disagreement about the direction of the decisions—what is easy for one is difficult for the other. Also, managers, who have zero cash endowments, are protected by limited liability, that is, their state-contingent incomes must always be nonnegative.⁶

2.2 | Organizational structures and contracts

The organizational structure is contractible. We assume two different options, each of which implies a different allocation of decision rights. First, the production units can remain separate (the *nonintegration* regime, denoted by N). In this case, managers have full control over their decisions. Second, the two units can integrate, a regime denoted by I, into a single entity by selling their assets to a third party, called the headquarter (HQ), which gives HQ full control over managerial decisions, e_A and e_B , assuming that the third party is endowed with enough cash to acquire the firm.⁷ The headquarter is motivated entirely by revenue and incurs no costs from the managerial decisions. These costs are still borne by the managers. As argued by Hart and Hölmstrom (2010), integration results in an organization where less weight is placed on private costs than under nonintegration. This, however, is offset by the fact that total revenue is maximized rather than individual unit profits.

The revenue of each firm is publicly verifiable, and hence, ex ante contractible. We assume that each *A* unit has all the bargaining power in an arbitrary firm (a, b) and makes take-it-or-leave-it offers to the *B* unit.⁸ A contract $(s, d) \in [0, 1] \times \{N, I\}$ specifies a revenue share *s* for the *B* unit and an organizational mode *d*.

Consider an arbitrary enterprise (a, b). If the units stay separate, then a revenue-sharing contract is simply a share *s* of the total revenue that accrues to the *B* unit. As we assume limited liability, the units get nothing in the case of failure. When the two units integrate, *HQ* buys the assets of *A* and *B* units at predetermined prices in exchange for a share contract $\mathbf{s} = (s_A, s_B, s_{HQ}) \in \mathbb{R}^3_+$ with $s_A + s_B + s_{HQ} = 1$. We assume that the *HQ* has no market power.

2.3 | Timing of events

The economy lasts for two dates, t = 1, 2. At date 1, one *A* producer and one *B* producer match one-to-one to form a firm (a, b) and each *A* unit makes a take-it-or-leave-it contract offer (s, d) to each *B* unit. At date 2, the manager of each unit chooses e_A and e_B . We solve the model by backward induction.

2.4 | The equilibrium allocations

An equilibrium of the supplier market consists of a set of firms formed through feasible contracts, that is, organizational structures and corresponding revenue shares, for each firm. Recall that there are two possible organizational modes for each enterprise—integration (*I*) and nonintegration (*N*). In general, choice of organization depends on the revenue share that accrues to each member of a firm and the firm output. An allocation for the market $\langle \mu, v, u \rangle$ specifies a one-to-one matching rule $\mu : B \to A$, and payoff functions $v : A \to \mathbb{R}_+$ and $u : B \to \mathbb{R}_+$ for the *A* and *B* units, respectively. Further, let $\phi(a, b, u)$ denote the *bargaining* or *Pareto frontier* of firm (*a*, *b*), which is the maximum payoff that a type *aA* unit can achieve given that the *B* unit of type *b* consumes *u*.

Definition 1 (Equilibrium). Given the type distributions G(a) and F(b), an allocation $\langle \mu, \nu, u \rangle$ constitutes an equilibrium of the economy if they satisfy the following conditions:

- (a) Feasibility: The payoff allocations (u, v) are feasible, that is, $v \le \phi(a, b, u)$ and $0 \le u \le \bar{u}(a, b)$ where $\phi(a, b, \bar{u}(a, b)) = 0$ for all firms (a, b);
- (b) Optimization: Each A unit of a given type chooses optimally a B unit to form an enterprise, that is, given u for b ∈ B, each a ∈ A solves

$$v = \max_{b} \phi(a, b, u). \tag{1}$$

(c) Measure consistency: The equilibrium matching function satisfies the following "measure consistency" condition. For any subinterval $[i_0, i_1] \subseteq J_A$, let $i_k = G(a_k)$ for k = 0, 1, that is, a_k is the productivity of the A

unit at the i_k th quantile. Similarly, for any subinterval $[j_0, j_1] \subseteq J_B$, let $j_h = F(b_h)$ for h = 0, 1. If $[a_0, a_1] = \mu([b_0, b_1])$, then it must be the case that

$$j_1 - j_0 = F(b_1) - F(b_0) = G(a_1) - G(a_0) = i_1 - i_0.$$
(2)

Definition 1(b) asserts that each A unit chooses her partner optimally which is equivalent to the notion of "stability" in an assignment model. Part (c) of the above definition simply says that one cannot match say two-third of the A units to one-third of the B units because the matching is constrained to be one-to-one.

We build on Legros and Newman's (2013) model wherein units on each side of the supplier market are homogeneous. Consequently, all firms produce the same output, which is taken as exogenous to the firms. Although our analysis of the choice of ownership structure in Section 3 is the same as in Legros and Newman (2013), the market equilibrium is quite different because of the two-sided heterogeneity of the supplier units. In particular, due to endogenous sorting of *A* and *B* units, firm output z(a, b) is endogenous in any equilibrium allocation.

3 | ORGANIZATIONAL CHOICE OF AN ARBITRARY ENTERPRISE

We first analyze the optimal contract for an arbitrary firm (a, b) under each organizational structure separately, and then we determine how each firm (a, b) chooses between the two structures.

3.1 | Nonintegration

Under this organizational mode, the shares accruing to the units affect both the size and the distribution of surplus between them. In other words, utility is imperfectly transferable (ITU) between the unit managers. Writing $z \equiv z(a, b)$, the optimal contract for a nonintegrated firm (a, b) solves the following maximization problem:

$$\max_{s} \quad V_A \equiv \pi (e_A, e_B)(1 - s)z - e_A^2, \tag{3}$$

subject to
$$U_B \equiv \pi(e_A, e_B)s_Z - (1 - e_B)^2 = u,$$
 (4)

$$e_A = \operatorname*{argmax}_{e_A} \{ \pi(e, e_B)(1 - s)z - e^2 \}, \tag{5}$$

$$e_B = \operatorname*{argmax}_{e} \{ \pi(e_A, e) sz - (1 - e)^2 \},$$
(6)

where *u* is the outside option of the *B* unit. We assume that $u \ge u_0$, where $u_0 > 0$ is the reservation utility of all *B* producers, that is, the utility any *B* unit obtains if it does not participate in any firm. Constraint (4) is the *participation constraint* of the *B* unit, whereas constraints (5) and (6) are the *incentive compatibility constraints* of the *A* firm and the *B* firm, respectively. Solving for e_A and e_B from (5) and (6), we obtain

$$e_A(s) = \frac{(1-s)z}{1+z}$$
 and $1 - e_B(s) = \frac{sz}{1+z}$

As the share of the *B* unit, *s*, increases, e_A decreases and $1 - e_B$ increases. Substituting $e_A(s)$ and $e_B(s)$ into (4), and solving we get the equilibrium share, s(z, u). Evaluating the objective function at s(z, u), $e_A(s(z, u))$ and $e_B(s(z, u))$, we obtain the maximum payoff that accrues to the *A* unit given that the *B* unit consumes *u*:

$$\phi^{N}(a,b,u) \equiv \Phi^{N}(z,u) = u - z^{2} + \frac{z}{1+z}\sqrt{z^{2}(2+z)^{2} - 4(1+z)^{2}u} \quad \text{for} \quad 0 \le u \le \frac{z^{2}}{1+z}.$$
(7)

The function $\phi^N(a, b, u)$ is the bargaining frontier of firm (a, b) under nonintegration. The bargaining frontier is strictly increasing in *a* and *b*, and strictly decreasing in *u*. Note that *u* must lie between 0, which corresponds to s(z, u) = 0, and $z^2/(1 + z)$, the level corresponding to s(z, u) = 1. The frontier is symmetric with respect to the 45⁰ line, on which $\phi^N(a, b, u) = u$ and s(z, u) = 1/2. The total firm surplus is maximized when the shares of the two nonintegrated units are equal, while it is minimized at s(z, u) = 0 or s(z, u) = 1. Equal or more broadly, "balanced" shares provide strong incentives for the managers to coordinate their actions. Finally, higher firm revenue *z* also induces better coordination as $e_B - e_A$ decreases with *z*. Any nonintegrated firm with production *z* has an expected output that is given by:

$$q^{N}(z) = \frac{\pi(e_{A}, e_{B})}{1 - \frac{1}{(1+z)^{2}}} \cdot z = z \left(1 - \frac{1}{(1+z)^{2}} \right).$$
(8)

urnal of Economics &

Note that $q^{N}(z)$ is lower than z, that is, a nonintegrated firm does not reach its full potential.

3.2 | Integration

When the units integrate, the firm is acquired by HQ who is conferred with the decision making rights. Motivated entirely by incomes, HQ will choose e_A and e_B to maximize

$\pi(e_A, e_B)s_{HQ} \cdot z,$

its expected share of revenue as long as $s_{HQ} > 0$. This induces $e_A = e_B$, and hence, $\pi(e_A, e_B) = 1$ for each integrated firm. *HQ* breaks even in the absence of market power. The private costs of managerial actions are still borne by the individual unit managers. The aggregate managerial cost, $e_A^2 + (1 - e_B)^2$ is minimized when $e_A = e_B = 1/2$. Thus, the bargaining frontier under integration is given by

$$\phi^{I}(a, b, u) \equiv \Phi^{I}(z, u) = z - \frac{1}{2} - u \quad \text{for } 0 \le u \le z - \frac{1}{2}.$$
(9)

The above function is linear in u, that is, utility is fully transferable (TU) between the two units because neither the action taken by HQ nor the costs borne by the managers depend on the revenue shares. The frontier is strictly increasing in a and b, strictly decreasing in u (with slope -1) and symmetric with respect to the 45⁰ line. The expected output produced by an arbitrary integrated enterprise (a, b) is given by

$$q^{I}(z) = \underbrace{\pi(e_{A}, e_{B})}_{1} \cdot z = z.$$
(10)

Although utility is fully transferable between the A and B producers, this form of organization is in general inefficient as HQ, having a stake in the firm revenue, places too little weight on private managerial costs while maximizing expected revenue.

3.3 | Choice of organization

We now analyze the optimal choice of organizational structure by a given firm (a, b). The difference between the two organizational structures stems from the difference in the nature of inefficiency, that is, $\phi^N(a, b, u)$ represents the surplus of an ITU model, whereas $\phi^I(a, b, u)$ is associated with fully transferable utility. Nonintegration is conducive to poor coordination as this organization places too much weight on the private managerial costs. Integration, on the other hand, places too much importance on coordination, and tends to ignore private costs. The optimal choice of organization thus depends on firm revenue, z(a, b) as well as the way the aggregate surplus is shared between the two units, that is, the utility allocation (u, v). At any level of utility u accruing to the B firm, an arbitrary enterprise (a, b) WILEY-A Journal of Economics &

would choose *N* over *I* if and only if $\phi^N(a, b, u) > \phi^I(a, b, u)$ or equivalently, $\Phi^N(z, u) > \Phi^I(z, u)$. The following lemma characterizes the optimal choice of ownership structures.

Lemma 1. For an arbitrary firm (a, b), let $z \equiv z(a, b)$. There are threshold values of utility of the B unit, $u_L(z)$ and $u_H(z)$ that are given by

$$u_L(z) = \frac{(z-1)(1+2z)}{4(1+z)}$$
 and $u_H(z) = \frac{2z^2+3z-1}{4(1+z)}$,

where $0 \le u_L(z) < u_H(z) \le \frac{z^2}{1+z}$ such that the firm (a) chooses to stay separate if $u_L(z) < u < u_H(z)$; (b) chooses to integrate if either $u < u_L(z)$ or $u > u_H(z)$; (c) is indifferent between N and I if $u = u_L(z)$, $u_H(z)$. Thus, the bargaining frontier of an arbitrary firm (a, b) with output $z \equiv z(a, b)$ is given by

$$\phi(a, b, u) = \max\{\phi^N(a, b, u), \phi^I(a, b, u)\}$$

The proof of the above and many subsequent results are in Appendix A. The choice of organization by the unit managers depends on the firm revenue *z* as well as how the revenue is shared, that is, 1 - s and *s* between the two units. In a nonintegrated enterprise, both the level and share of revenue play crucial roles in coordinating managerial actions. An integrated firm, on the other hand, achieves full coordination because $e_A = 1 - e_B = 1/2$ irrespective of the revenue sharing rule; however, such ownership structure, driven by revenue motives, ignores private costs of managerial actions. There is no clear dominance of one ownership structure over the other. At a more balanced utility allocation, coordination among the two units can be easily achieved without being integrated. Thus, for intermediate values of *u*, that is, $u_L < u < u_H$, the units prefer to stay separate. On the other hand, for the extreme values of *u*, either high or low, integration is preferred because the shares are tilted in favor of one of the two units, and the incentives for revenue maximization are strong.

The bargaining frontier, $\Phi(z, u) = \max\{\Phi^N(z, u), \Phi^I(z, u)\}$ for a given *z* is depicted in Figure 1. Note that the bargaining frontier associated with *I*, $\Phi^I(z, u)$ is linear in *u* for each *z* meaning that firm surplus under integration is constant with respect to *u* for each *z*. On the other hand, the frontier that corresponds to nonintegration, $\Phi^N(z, u)$ is strictly concave in *u* for each *z*, which means that the aggregate surplus is not independent of how it is shared (imperfect transferability). Because both the frontiers are symmetric (around the 45°-line), if they intersect, they must intersect exactly twice at $(u_L, \Phi(z, u_L))$ and $(u_H, \Phi(z, u_H))$.

Remark 1. We have assumed that managerial actions are noncontractible. If, instead, actions were contractible, a firm would implement the first-best actions which are given by



FIGURE 1 The bargaining frontier $\Phi(z, u)$ is the upper envelope of $\Phi^N(z, u)$, the concave function and $\Phi^I(z, u)$, the linear function. The minimum surplus under *N* is given by $A_N = B_N = \frac{z^2}{1+z}$. On the other hand, $A_I = B_I = z - \frac{1}{2}$ is the fixed surplus under *I*. A firm with output *z* is indifferent between nonintegration and integration at the two kinks which correspond to $u_L(z)$ and $u_H(z)$. [Color figure can be viewed at wileyonlinelibrary.com]

$$(e_A^*, e_B^*) = \underset{\{e_A, e_B\}}{\operatorname{argmax}} \left\{ \pi(e_A, e_B)z - e_A^2 - (1 - e_B)^2 \right\} = \left(\frac{z}{1 + 2z}, \frac{1 + z}{1 + 2z}\right).$$

The corresponding first-best surplus is given by $\frac{2z^2}{1+2z}$ which is strictly higher than $z - \frac{1}{2}$, the surplus under integration. Thus, in the absence of the managerial incentive problem, a firm would always choose nonintegration over integration.¹⁰ Nevertheless, noncontractible managerial effort is a more realistic assumption given that the two units under nonintegration are independent.

4 | THE MARKET EQUILIBRIUM

In what follows, we analyze the equilibrium of the supplier market with a continuum of producer types.

4.1 | The indifference curves

We first derive the "indifference curve" of a given A unit with productivity a. The indifference curve of any a is the combinations of b and u which yield the same level of utility v to a, that is,

$$\phi(a, b, u) = v$$

Now fix an *a*. Because $\phi(a, b, u)$ is decreasing in *u* for all (a, b), its inverse with respect to *u* is well-defined, which is given by

$$u = \psi(b, a, v) = \max\{\psi^{N}(b, a, v), \psi^{I}(b, a, v)\},$$
(11)

which represents the indifference curve of any *A* supplier with type *a* at a constant level of utility *v*. The function $\psi^d(b, a, v)$ denotes the indifference curve of *a* at level *v* under organizational structure d = N, I. We shall now derive the indifference map of any given *a* from our primitives, the bargaining frontier of a firm (a, b).

For any given *a*, an increase in *b* implies an increase in the firm revenue, and hence, an increase in firm surplus. As a result, the bargaining frontier $\phi(a, b, u)$ shifts out as *b* or equivalently, *z* increases. Such expansions are depicted in the left panel of Figure 2. Note that the shifts of the linear part of the frontier are parallel, but those of the nonlinear part are not as the slope of $\phi^N(a, b, u)$ depends on *z*.¹¹

Three bargaining frontiers are drawn in Figure 2. The lowest-frontier, denoted by $\phi(a, \underline{b}, u)$ is associated with a firm formed by matching *a* and <u>*b*</u>, the lowest-productivity *B* unit. The two other frontiers correspond to (a, \hat{b}) and (a, b') where $\underline{b} < \hat{b} < b'$. Each firm is indifferent between nonintegration and integration at the two kinks of the associated bargaining frontier. The curves labeled IL^- and IL^+ are the *revenue expansion paths* which are drawn by joining the indifference points $u_L(z(a, b))$ and $u_H(z(a, b))$ by varying *b*. The revenue expansion paths divide the *u*-*v* space into three distinct regions. In the regions labeled *I*, any firm (a, b) chooses integration because the share of surplus is unbalanced in favor of either *a* or *b*. By contrast, in the region labeled *N*, nonintegration is the dominant choice. Clearly, on IL^- and IL^+ , any firm is indifferent between *N* and $I.^{12}$

The indifference map of any a is derived as follows:

• Fix the utility level of *a* at v_0 . At this level of constant utility, for any firm (a, b) with $b \ge \underline{b}$, integration is the dominant choice because any utility allocation (u, v_0) lies below the revenue expansion path IL^- , that is, in the region labeled *I*. The corresponding (b, u) combinations represent the indifference curve which is given by

$$\psi(b, a, v_0) = \psi^I(b, a, v_0)$$

in the right panel of the figure. Clearly, higher *b* implies higher *u* for the given *A* unit utility level v_0 , that is, the indifference curve is upward-sloping.



FIGURE 2 The left panel depicts the bargaining frontiers of a given *A* of type *a* by varying her partner's type *b*. The right panel depicts the corresponding indifference map of *a* in her matched partner's type-utility space. [Color figure can be viewed at wileyonlinelibrary.com]

Fix the utility level of *a* at *v*₁ such that *v*₁ > *v*₀. Note that the horizontal line at *v*₁ starts at the bargaining frontier φ(*a*, *b*, *u*) and intersects the revenue expansion path *IL*⁻ only once. Because there is a continuum of *b*, the indifference map of any *A* unit is dense, and hence, there is a unique *b* such that φ(*a*, *b*, *u*) passes through the intersection of *v*₁ and *IL*⁻. In other words, firm (*a*, *b*) is indifferent between nonintegration and integration. Moreover, for any *b* < *b*, all utility allocations (*u*, *v*₁) lie above *IL*⁻, that is, any firm (*a*, *b*) such that *b* < *b* chooses to stay separate. By contrast, for any *b* > *b*, the allocations (*u*, *v*₁) lie below *IL*⁻, and hence, any firm (*a*, *b*) such that *b* > *b* chooses to integrate. The corresponding indifference curve is depicted in the right panel of Figure 2, which is given by

$$\psi(b, a, v_1) = \begin{cases} \psi^N(b, a, v_1), & \text{if } b < \hat{b}, \\ \psi^I(b, a, v_1), & \text{if } b \ge \hat{b}. \end{cases}$$

Note that this indifference curve has a unique kink at $b = \hat{b}$. This is to say that *a* is indifferent between *N* and *I* at $(b, u) = (\hat{b}, \hat{u})$.

• Next, let the constant utility level of *a* be v_2 (as depicted in the left panel of Figure 2) with $v_2 > v_1$. The horizontal line at v_2 intersects both the revenue expansion paths, IL^+ and IL^- . In this case, the ownership structure is nonmonotonic in *b*. Firm (*a*, *b*) chooses to integrate if the value of *b* is either low or high. By contrast, firm (*a*, *b*) chooses to stay separate for intermediate values of *b*. The corresponding indifference curve thus has two kinks, \hat{b}_1 and \hat{b}_2 . The indifference curve is given by

$$\psi(b, a, v_2) = \begin{cases} \psi^N(b, a, v_2), & \text{if } b < \hat{b}_1, \\ \psi^I(b, a, v_2), & \text{if } \hat{b}_1 \le b \le \hat{b}_2, \\ \psi^N(b, a, v_2), & \text{if } b > \hat{b}_2. \end{cases}$$

• Finally, consider the constant utility level $v_3 > v_2$. The horizontal line at v_3 intersects IL^+ only once, but does not intersect IL^+ . Note that, at this constant utility level v_2 , firm (a, \underline{b}) is not feasible anymore because $v_3 > \phi(a, \underline{b}, 0)$. In this case, firm (a, b) chooses to integrate (stay separate) if values of *b* are low (high). The corresponding indifference curve, which has a single kink, is given by.

urnal of Economics & anagement Strategy

$$\psi(b, a, v_3) = \begin{cases} \psi^I(b, a, v_3), & \text{if } b < \hat{b}, \\ \psi^N(b, a, v_3), & \text{if } b \ge \hat{b}. \end{cases}$$

Note that this indifference curve starts at a $b > \underline{b}$ because v_3 is not feasible for the firm (a, \underline{b}) , and it has only one kink (below which integration is the dominant choice, and above which nonintegration is strictly preferred).

Note that (in the right panel of Figure 2), because at a given u, higher b implies greater utility for an A unit with productivity a, "higher" indifference curves lie to the southeast.

4.2 | The congruent marginal contributions and PAM

4.2.1 | Generalized increasing differences

Our objective is to analyze the nature of the equilibrium matching function, $a = \mu(b)$. In particular, we shall show that matching is positive assortative, that is, $\mu(b)$ is an increasing function. Legros and Newman (2007) assert that if the bargaining frontier, $\phi(a, b, u)$, satisfies the *generalized increasing differences* (GID) condition, that is, for any a'' > a', b'' > b' and u'' > u', we have

$$\phi(a', b'', u'') = \phi(a', b', u') \quad \Rightarrow \quad \phi(a'', b'', u'') \ge \phi(a'', b', u'), \tag{12}$$

then the equilibrium matching is *positive assortative* (PAM).¹³ We first provide a simple characterization of GID in terms of *supermodularity* of the indifference curves of A units.

Proposition 1. The bargaining frontier $\phi(a, b, u)$ satisfies GID if and only if the indifference curve $\psi(b, a, v)$ is supermodular in (b, a), that is, for any a'' > a' and b'' > b',

$$\psi(b', a', v) + \psi(b'', a'', v) \ge \psi(b', a'', v) + \psi(b'', a', v)$$
(13)

for each constant utility level v of the A producers.

Intuitively, GID asserts that if two type-payoff combinations (b', u') and (b'', u'') with (b'', u'') > (b', u') are on the same indifference curve of a lower type a', that is, $\phi(a', b'', u'') = \phi(a', b', u')$, then for any higher type a'', (b'', u'') must lie on a higher indifference curve than the one on which (b', u') lies, that is, $\phi(a'', b'', u'') \ge \phi(a'', b', u')$. On the other hand, supermodularity of $\psi(b, a, v)$ in (b, a) means that the indifference curve of an A producer of higher productivity is everywhere steeper than that of her lower type counterpart. Thus, GID is a single-crossing condition.¹⁴ The intuition is simple. In the left panel of Figure 3, $\psi(b, a'', v)$ is everywhere steeper than $\psi(b, a', v)$ for each v, that is, condition (13) holds for these two arbitrarily chosen types a' and a'' with a'' > a'.¹⁵ The points (b', u') and (b'', u'') lie on the same indifference curve of a', but (b'', u'') lies on the indifference curve of a'', which is higher than the one passing through (b', u'), which imply GID. On the other hand, if supermodularity, that is, condition (13), is violated the indifference curve of a'' is flatter than that of the lower type a' at some (b', u'), then one can find points like (b'', u'') > (b', u') such that type a' is indifferent between (b', u') and (b'', u''), but (b'', u'') yields lower utility to a'' than (b', u') does (because (b'', u'') lies on a lower indifference curve), which contradicts GID.¹⁶

4.2.2 | Congruence of marginal contributions

Proposition 1 asserts that an economy, characterized by imperfectly transferable utility, to satisfy GID and hence exhibit PAM, the indifference curves of *A* producers, $\psi(b, a, v)$ are to be supermodular in (b, a) for each *v*. However, in our model, the indifference curves are given by max{ $\psi^N(b, a, v), \psi^I(b, a, v)$ } with both $\psi^N(b, a, v)$ and $\psi^I(b, a, v)$ being supermodular (see Theorem 1). It is well-known that the maximum of two supermodular functions is *not necessarily* supermodular. We therefore, require additional conditions on $\psi(b, a, v)$ to guarantee that the indifference curves of *A*



FIGURE 3 The left panel depicts that the indifference curves are supermodular in (b, a), that is, $\psi(b, a'', v)$ is everywhere steeper than $\psi(b, a', v)$ for a'' > a' and for each v, and hence, they cross at most once. On the other hand, the right panel depicts a possible violation of supermodularity, which implies that the GID must also be violated. [Color figure can be viewed at wileyonlinelibrary.com]

units are supermodular, or equivalently, GID of the bargaining frontiers. We first define an additional sufficient condition, termed as the *congruent marginal contribution property* (CMC) in terms any two real-valued functions of (b, a).

Definition 2 (Congruent marginal contributions). Let x(b, a) and y(b, a) be real-valued functions. These functions are said to satisfy the "congruent marginal contributions" property if

either
$$y(b, a'') - x(b, a'') \ge y(b, a') - x(b, a')$$
 for all b and for any $a' < a'', (CMC_a^+)$
 $y(b'', a) - x(b'', a) \ge y(b', a) - x(b', a)$ for all a and for any $b' < b'', (CMC_b^+)$

or
$$y(b, a'') - x(b, a'') \le y(b, a') - x(b, a)$$
 for all b and for any $a' < a'', (CMC_a^-)$
 $y(b'', a) - x(b'', a) \le y(b', a) - x(b', a)$ for all a and for any $b' < b'', (CMC_b^-)$

Think of x(b, a) and y(b, a) as describing any two distinct production technologies or any two distinct organizations which a firm may choose between. The congruent marginal contributions property asserts that if a more productive *B* producer gains (loses) by switching from one organization to the other (say, switching from *x* to *y*), so does a more productive *A* producer. In other words, the marginal gains of *B* and *A* units due to technology switching or organizational restructuring are of the same sign. We first prove the following important result.

Proposition 2. Let x(b, a) and y(b, a) be continuous, increasing and supermodular in (b, a), and satisfy the congruent marginal contributions property. Then, $w(b, a) = \max\{x(b, a), y(b, a)\}$ is supermodular in (b, a).

The intuition behind the above result is best understood if we consider the differentiable version of the congruent marginal contributions property (Definition 2):

$$sign[y_a(b, a) - x_a(b, a)] = sign[y_b(b, a) - x_b(b, a)],$$
(14)

that is, the marginal gains (losses) from switching technology or organization point in the same direction for both A and B producers. In other words, there is no "disagreement" between A and B in switching to the same technology.¹⁷ Because each "technology" is supermodular, a more productive A unit has comparative advantage in producing

urnal of Economics &

surplus along with a more productive B unit. Moreover, the congruent marginal contributions property asserts that both higher a and higher b gain more by switching to the same technology or organization. Thus, supermodularity of each technology along with the congruence of marginal contributions make the combined technology (i.e., the maximum function) supermodular. Proposition 2 is not only useful in proving GID of the bargaining frontiers in the present context, but can also be helpful in establishing single-crossing in environments where matched units faces binary choices.

4.2.3 | PAM

In the following theorem, our main result of this section, we show that, under certain restrictions on the production function, z(a, b), the bargaining frontier of each firm satisfies the conditions of Proposition 2, and hence has GID. Consequently, the equilibrium exhibits PAM.

Theorem 1. Let firm revenue z(a, b) be strictly increasing and log-supermodular in (a, b). Then, the bargaining frontier of each firm (a, b), $\phi(a, b, u) = \max\{\phi^N(a, b, u), \phi^N(a, b, u)\}$, satisfies GID. Consequently, in the equilibrium of the supplier market, more productive A producers match with more productive B producers to form enterprises following a PAM pattern.

Log-supermodularity of z(a, b), which is stronger than supermodularity, is needed to guarantee that the bargaining frontier of each nonintegrated firm, $\phi^N(a, b, u)$, satisfies GID. On the other hand, the bargaining frontier of each integrated firm, $\phi^I(a, b, u)$, satisfies GID even if z(a, b) is supermodular, because the frontier is linear in u (perfectly transferable utility).¹⁸

Next, we show that the indifference curves of any *a* under nonintegration and integration, $\psi^N(b, a, v)$ and $\psi^I(b, a, v)$ satisfy the congruent marginal contributions property for each *v*.¹⁹ Therefore, $\psi(b, a, v)$, is supermodular in (b, a) for each *v*, which follows from Proposition 2. Thus, the bargaining frontier $\phi(a, b, u)$ satisfies GID (cf. condition (12)).

The above theorem can be understood in terms of the diagram in the right panel of Figure 3. Let $\hat{b}(a)$ be the *b* that solves $\psi^N(\hat{b}(a), a, v) = \psi^I(\hat{b}(a), a, v)$ for each *v*, and assume that $\hat{b}(a)$ is unique, that is, the graph of $\psi(b, a, v)$ has a single kink for a given *v*. Clearly, a sufficient condition for $\psi(b, a, v)$ to satisfy supermodularity is that the kink of the indifference curve moves to the left as *a* increases, that is, $d\hat{b}/da \le 0.20$ This condition holds in our model because

$$\frac{d\hat{b}}{da} = -\frac{\psi_a^N(\hat{b}(a), a, v) - \psi_a^I(\hat{b}(a), a, v)}{\psi_b^N(\hat{b}(a), a, v) - \psi_b^I(\hat{b}(a), a, v)} = -\frac{z_a(a, \hat{b})}{z_b(a, \hat{b})} < 0.$$

If $\hat{b}(a)$ decreases with *a* (as in the left panel of Figure 3), then for each v, $\psi(b, a', v)$ and $\psi(b, a'', v)$ for two distinct a' and a'' can cross each other only once. The generalized increasing difference property for our economy easily follows because the *A* and *B* units possess the same production technology, z(a, b), both in nonintegrated and integrated firms. Hence, the marginal contribution of each unit toward the firm revenue does not alter from one organization to the other.

Because the bargaining frontier, $\phi(a, b, u) = \max\{\phi^N(a, b, u), \phi^I(a, b, u)\}$ satisfies GID or equivalently, the indifference curve of the *A* producers, $\psi(b, a, v) = \max\{\psi^N(b, a, v), \psi^I(b, a, v)\}$ is supermodular in (b, a) for each *v*, the fact that the equilibrium matching is PAM follows from Legros and Newman (2007). The intuition is simple. If an *A* producer with lower productivity, *a'* is indifferent between two type-payoff combinations, (b', u') and (b'', u'') with (b', u') < (b'', u''), that is, $\phi(a', b', u') = \phi(a', b'', u'') = v$, then GID implies that, for an *A* unit with higher productivity, *a''*, we have $\phi(a'', b', u') = v \le v' = \phi(a', b'', u'')$, that is, (b'', u'') lies on a higher indifference curve of *a''*. In other words, *a''* is willing to pay more than *a'* to lure *b''*. Consequently, *a'* is matched with *b'* and *a''* is matched with *b''*, following a PAM pattern.

In fact, both types of organization treat the two units symmetrically (unlike the example of footnote 17). Theorem 1 can be generalized to a situation where the output of an integrated firm is a monotonic transformation of that of a nonintegrated firm, that is, $z^{I}(a, b) = h(z^{N}(a, b))$ where $h : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function. Clearly, a

monotonic transformation preserves the ranking of the two units in terms of their marginal contributions toward the firm revenue.²¹

We have demonstrated that the equilibrium of the supplier market exhibits PAM, and firm heterogeneity comes to be endogenous. We turn next to the equilibrium choice of organizations. We shall show that even though equilibrium matching is PAM with respect to unit productivity and more productive firms produce higher expected outputs (all else equal), there can be an equilibrium where expected output is nonmonotonic in firm productivity due to organizational restructuring in equilibrium.

4.3 | Equilibrium ownership structures

14

-WILEY-

We now analyze the choice of organization of each equilibrium enterprise ($\mu(b)$, b). To this end, we shall require two equilibrium objects—namely, the *equilibrium indifference loci* and the *equilibrium utility function* of the *B* producers (the equilibrium utility of the *A* producers is then derived from the bargaining frontier). Because the equilibrium matching is PAM irrespective of the choice of ownership structures, each equilibrium enterprise ($\mu(b)$, b) can be uniquely identified by b, the productivity of its *B* unit.

4.3.1 | The equilibrium indifference loci

The equilibrium indifference loci are given by evaluating the threshold utilities in Lemma 1 along the equilibrium assignment, so that:

$$U_L(b) \equiv u_L(z(\mu(b), b))$$
 and $U_H(b) \equiv u_H(z(\mu(b), b)).$

When the types of the *B* producers vary continuously on $[\underline{b}, \overline{b}]$, both $U_L(b)$ and $U_H(b)$ are continuous functions of *b*, which we term the *equilibrium indifference loci* (or indifference loci, for brevity). Clearly, the characteristics of the indifference locus depend on the output of each firm in equilibrium, $z(\mu(b), b)$, which is an equilibrium object because it involves the matching function, $\mu(b)$. However, it can be easily pinned down once the type distributions, G(a) and F(b), are known. From the measure consistency condition and PAM, it follows that $G(\mu(b)) = F(b)$, and hence, $\mu'(b) = f(b)/g(\mu(b))$. From Lemma 1, it follows that $u'_H(z) > u'_L(z) > 0$ for any *z*, and hence, $U'_H(b) > U'_L(b)$ for all *b*. Note also that $u_H(z) - u_L(z)$ is strictly increasing in *z*, which implies that the two loci never meet each other. The indifference loci, $U_L(b)$ and $U_H(b)$ are drawn in Figure 4.

4.3.2 | The equilibrium utility function

Next, we analyze the equilibrium utility function of the *B* producers, which is denoted by u(b). Note that, in any equilibrium allocation, u(b) is a strictly increasing function. To see this, consider any two b' and b" with b' < b" and $u(b') \ge u(b")$. Suppose further that $a = \mu(b')$. Because the bargaining frontier $\phi(a, b, u)$ of any *a* is strictly increasing in *b* and strictly decreasing in *u*, we have $\phi(a, b", u(b")) > \phi(a, b', u(b'))$, and hence, *a* is better-off being matched with *b*" rather than *b*'. This argument is reflected in the first order condition of the maximization problem (1) which is given by the following ordinary differential equation (ODE)²²:

$$u'(b) = -\frac{\phi_2^d(a, b, u(b))}{\phi_3^d(a, b, u(b))} \text{ for } a = \mu(b) \text{ and } d \in \{N, I\}.$$
(15)

Because in all firms, the *A* unit makes a take-it-or-leave-it offer to the *B* unit, the lowest-productivity *B* units consume the reservation utility, u_0 , the utility any *B* unit would have consumed if not matched with any *A* producer, which is the initial condition for the above ODE. The equilibrium utility function, u(b) is depicted in Figure 4.



FIGURE 4 Equilibrium choice of ownership structures. (a) Equilibrium $\langle I \rangle$ with $u(b) < U_L(b)$, (b) equilibrium $\langle N \rangle$ with $U_L(b) < u$ (b) $< U_H(b)$, (c) equilibrium $\langle N, I \rangle$ with $u(b^*) = U_H(b^*)$, and (d) equilibrium $\langle I, N \rangle$ with $u(b^{**}) = U_L(b^{**})$. [Color figure can be viewed at wileyonlinelibrary.com]

4.3.3 | Equilibrium organizations

Although our economy exhibits PAM of productivity irrespective of the choice of ownership structure, there are potentially many equilibria in the choice of organizations. However, we analyze four types of equilibria. Two of them exhibit only a single organizational mode, that is, all the equilibrium enterprises either choose nonintegration, denoted by $\langle N \rangle$, or integration, denoted by $\langle I \rangle$. The other two are monotone equilibria—there is a unique indifferent firm with *B* unit productivity *b** so that either (a) all equilibrium enterprises with $b < b^*$ choose to stay separate, and all firms with $b > b^*$ integrate, or (b) all firms with $b < b^*$ integrate, and the rest with $b > b^*$ stay separate. These monotone equilibria are denoted $\langle N, I \rangle$ and $\langle I, N \rangle$, respectively. Although it is possible to have nonmonotone equilibria with multiple indifferent firms, the driving forces behind the equilibria are no different than the four cases we consider. The aforementioned equilibria are depicted in Figure 4.

Broadly speaking, the equilibrium organizational choice is determined by the relative bargaining power of A and B suppliers. When the bargaining power is unbalanced and one side receives a large portion of the surplus, integration is preferred by both suppliers. When bargaining power is more balanced, nonintegration is preferred by both suppliers. The relative bargaining power in a matching model is determined by three main forces: (i) the reservation utility for both suppliers, (ii) the relative marginal contribution to the match surplus, and (iii) the relative scarcity of suppliers of certain types. The first is especially important for the least productive pair as it determines how they split their match surplus. Thus, when the reservation utility of B suppliers, u_0 , is very large or very small, the bargaining power in the least productive firm is unbalanced leading to integration at the bottom of the market as seen in Figure 4a,d. For intermediate values of u_0 , the bargaining power is more balanced, and nonintegration prevails at the bottom as in Figure 4b,c.

Not only does the first force dictate the organizational choice at the bottom of the market, it influences the organizational choice for all matched pairs throughout the market. However, the second and third forces are the main drivers behind the equilibrium organizational structure of the entire market. The relative marginal contribution is

15

simply the ratio of marginal productivities, z_b/z_a , while the relative scarcity is captured by the slope of the matching function, $f(b)/g(\mu(b))$. We combine these two forces into a single term that we refer to as the *measure-adjusted* Marginal Productivity Ratio (MPR) of b relative to a, which is defined as:

$$\xi(a, b) \equiv \frac{z_b(a, b)/f(b)}{z_a(a, b)/g(a)}$$

This term quantifies the bargaining power of *B* suppliers relative to *A* suppliers when splitting the marginal match surplus, whereas the reciprocal quantifies the bargaining power for *A* suppliers relative to *B* suppliers. Thus, we drop the "*b* relative to *a*", and simply refer to the term as the *measure-adjusted MPR* for exposition. When the bargaining power is unbalanced for a given pair of suppliers, $\xi(a, b)$ is either very large or very small (i.e., the *B* suppliers have either very high or very low bargaining power). This corresponds to a utility function that is either very steep or very flat, respectively. When the measure-adjusted MPR is close to one, the bargaining power of both sides of the market is balanced. While the balance of the bargaining power for the least productive firm, represented by u_0 , determines the ownership structure at the bottom, the organizational choice for any other matched pair of suppliers is determined by the relative bargaining power of all the less productive pairs in the market. The clearest example of this fact can be seen in Figure 4d, where the bargaining power for the least productive enterprise is very unbalanced, which leads them to integrate. Additionally, the measure-adjusted MPR at the bottom diverges from one, but eventually comes close to one. In other words, the bargaining power at the bottom is very unbalanced in favor of the *B* units, but it becomes more balanced as one goes up in the productivity ladder. The resulting equilibrium organizational structure is $\langle I, N \rangle$.

We now introduce Theorem 2, which gives sufficient conditions in terms of the reservation utility and the measureadjusted MPR. Each subcase of the theorem corresponds to the associated panel in Figure 4.

Theorem 2. Let $\underline{z} = z(\underline{a}, \underline{b})$. There exist bounds on $\xi(a, b)$: $l_I(z(a, b))$, $h_I(z(a, b))$, $l_N(z(a, b), u)$ and $h_N(z(a, b), u)$ with $l_I(z(a, b)) < h_I(z(a, b))$ and $l_N(z(a, b), u) < h_N(z(a, b), u)$, such that

- (A) if $u_0 < u_L(\underline{z})$ and $\xi(a, b) < l_I(z(a, b))$, or $u_0 > u_H(\underline{z})$ and $\xi(a, b) > h_I(z(a, b))$ for all (a, b), all equilibrium enterprises, $(\mu(b), b)$ choose to integrate, that is, the equilibrium ownership structure is $\langle I \rangle$;
- (B) if $u_L(\underline{z}) \le u_0 \le u_H(\underline{z})$ and $l_N(z(a, b), u) \le \xi(a, b) \le h_N(z(a, b), u)$ for all (a, b, u), all equilibrium enterprises, $(\mu(b), b)$ choose to stay separate, that is, the equilibrium ownership structure is $\langle N \rangle$;
- (C) if $u_L(\underline{z}) \le u_0 \le u_H(\underline{z})$, and either $\xi(a, b) \le \min\{l_N(z(a, b), u), l_I(z(a, b)\}\}$ or $\xi(a, b) \ge \max\{h_N(z(a, b), u), h_I(z(a, b))\}\$ for all (a, b, u), there exists a unique $b^* \in (\underline{b}, \overline{b}]$ such that the equilibrium enterprises, $(\mu(b), b)$ chooses nonintegration (integration) according as $b < (>) b^*$, that is, the equilibrium ownership structure is $\langle N, I \rangle$;
- (D) if either $u_0 < u_L(\underline{z})$ and $\max\{l_N(z(a, b), u), l_I(z(a, b))\} < \xi(a, b) < h_N(z(a, b), u)$, or $u_0 > u_H(\underline{z})$ and $l_N(z(a, b), u) < \xi(a, b) < \min\{h_N(z(a, b), u), h_I(z(a, b))\}$ for all (a, b, u), there exists a unique $b^{**} \in (\underline{b}, \overline{b}]$ such that the equilibrium enterprises, $(\mu(b), b)$ chooses integration (nonintegration) according as $b < (>) b^{**}$, that is, the equilibrium ownership structure is $\langle I, N \rangle$.

The conditions for $\langle I \rangle$ are straightforward. They require that the slope of the utility function is greater (less) than the slope of the upper (lower) locus for each enterprise, implying no intersection of the two, and consequently no change in organizational structure. The conditions for $\langle N \rangle$ are similarly simple except that the slope of the utility function must be between the slopes of both loci for all enterprises to prevent a change in organizational structure. In other words, the bargaining power of the two sides of the supplier market is more or less balanced in all firms formed in equilibrium via endogenous matching of types.

Although they look complicated, the sufficient conditions for the monotone equilibria, $\langle N, I \rangle$ and $\langle I, N \rangle$, have a simple interpretation. For part (C), $u_L(\underline{z}) \leq u_0 \leq u_H(\underline{z})$ means that the bargaining power of *A* and *B* units in the least productive firm is balanced and hence, this firm chooses to stay separate. However, $\xi(a, b)$ low or high for all (a, b) implies that the bargaining power becomes more and more unbalanced as one goes up the productivity ladder. As a result, there is nonintegration at the bottom and integration at the top. The intuition behind the other monotone equilibrium is similar, that is, the bargaining power is more unbalanced at the bottom and more balanced at the top, which implies the $\langle I, N \rangle$ equilibrium structure.

At this juncture, it is worth addressing the assumption that production of all firms is sufficiently large, that is, $z(a, b) \ge 1$ for all (a, b). Suppose we assume instead that z(a, b) < 1 for low values of (a, b). Clearly, for low-productivity firms, including $(\underline{a}, \underline{b})$, nonintegration is the dominant choice. This assumption would change $\langle I \rangle$ to $\langle N, I \rangle$, and $\langle I, N \rangle$ to $\langle N, I, N \rangle$ (a nonmonotone equilibrium). Atalay et al. (2014) show that more productive firms are vertically integrated, which part (C) of the above theorem predicts. On the other hand, in part (D), Theorem 2 also predicts that more productive firms may stay separate while the less productive ones integrate, which can be implied by either $\langle I, N \rangle$ or $\langle N, I, N \rangle$ equilibrium. However, the monotonic structure is easier to understand, which with the increasing availability of firm-level data can and should be tested empirically. The implication of our assumption is that $\langle I, N \rangle$ will only occur in a real world setting if all the firms are sufficiently productive, that is, $z(a, b) \ge 1$.

5 | DISTRIBUTIONAL CONSEQUENCES AND EMPIRICAL RELEVANCE

In this section, we begin by discussing the impact equilibrium ownership structures have on the resulting distribution of expected output. Then we present a few testable implications of our model.

5.1 | Equilibrium output distribution

In our model, the expected output of enterprises is heterogeneous not only because they are formed by matching endogenously heterogeneous productive units, but also because firms with distinct levels of productivity may end up choosing distinct ownership structures (e.g., Syverson, 2011, classifies both productivity heterogeneity and firm boundary decisions as internal drivers of firm heterogeneity). In relation to the literature on international trade with heterogeneous firms (e.g., Melitz, 2003), we see expected output in our model as a good measure of productivity or firm size, and we view the results here as having important implications for future empirical research. The impact of the choice of ownership structures on the distribution of expected output can be seen by comparing its variance and skewness in the four equilibria depicted in Figure 4.

As the results for general type distributions cannot be derived analytically due to the nature of the ODE in (15), we carry out a numerical exercise using the Cobb–Douglas production function, $z(a, b) = a^{\alpha}b^{\beta}$, and truncated Pareto type distributions defined on a common positive support $[\underline{\theta}, \overline{\theta}]^{23}$:

$$G(a) = \frac{1 - (\underline{\theta}/a)^{\gamma_a}}{1 - (\underline{\theta}/\overline{\theta})^{\gamma_a}} \quad \text{and} \quad F(b) = \frac{1 - (\underline{\theta}/b)^{\gamma_b}}{1 - (\underline{\theta}/\overline{\theta})^{\gamma_b}}.$$

We set the following parameter values: $\alpha = 0.25$ and $\beta = 0.5$ for the production function; and $\gamma_a = \gamma_b = 12$, $\theta = 15$, $\bar{\theta} = 20$ for the type distributions. The implication of the distributional assumptions is that within a matched pair, the suppliers have the same productivity types. As a result, the measure-adjusted MPR evaluated along the equilibrium path is simply $\beta/\alpha = 2$ because the matching function is given by $\mu(b) = b$ when type distributions are identical.²⁴ Thus, the *B* suppliers in this numerical example have more bargaining power. In the numerical exercise, we vary u_0 on [2.6, 4.1] so that any change in the expected output distribution comes from a change in the reservation payoff of the *B* producers, u_0 . Note that we have fixed the second and third factors by choosing a constant measure-adjusted MPR.

The effect of organizational choice on the equilibrium distribution of expected output is depicted in Figure 5a. Unsurprisingly, the average output is the highest when all enterprises integrate and the lowest when all are nonintegrated. Moreover, extreme values of u_0 , high or low, make integration more likely. Turning to variance and skewness, the immediately noticeable feature of both distributional statistics is that they respond in a similar manner to changes in u_0 , which is because the distribution is right-skewed. The first comparison to consider is across the integration and nonintegration equilibria. As revealed from (8) to (10), the main characteristics of the expected output functions, $q^N(z)$ and $q^I(z)$ are that, for all z, both $q^N(z)$ and $q^I(z)$ are strictly increasing functions with $q^N(z) < q^I(z)$, and $q^N(z)$ is steeper than $q^I(z)$. The distributional implication of these features is that the variance of expected output in the $\langle N \rangle$ equilibrium is larger than in the $\langle I \rangle$ equilibrium for a large class of type distributions (a notable exception is the exponential distribution). The most important aspect of Figure 5 is the comparison of the two types of monotone equilibria, $\langle I, N \rangle$ and $\langle N, I \rangle$. Let $z^* \equiv z(\mu(b^*), b^*)$ be the output of the equilibrium enterprise, $(\mu(b^*), b^*)$, that is indifferent between nonintegration and integration. Because $q^N(z^*) < q^I(z^*)$, under the $\langle N, I \rangle$ equilibrium (i.e., any

17

urnal of Economics &



FIGURE 5 Distributional impact of equilibrium ownership structures. (a) The mean of the expected output, (b) the variance, and (c) the skewness, all as a function of u_0 . [Color figure can be viewed at wileyonlinelibrary.com]

equilibrium firm with output *z* chooses *N* if and only if $z < z^*$), there is an "upward jump" in the equilibrium expected output (as a function of *z*) at *z**. In contrast, at $z = z^*$, there is a "downward jump" in the equilibrium expected output under the $\langle I, N \rangle$ equilibrium (i.e., any equilibrium firm with output *z* chooses *I* if and only if $z < z^*$). In other words, the $\langle I, N \rangle$ equilibrium "compresses" the equilibrium output distribution relative to the $\langle N, I \rangle$ equilibrium. Thus, the effect of the upward and downward jumps in the equilibrium output under these two equilibria is reflected in their respective distributional statistics. When there is a downward jump in the equilibrium expected output under $\langle I, N \rangle$, we see both statistics shrink. On the other hand, the upward jump under $\langle N, I \rangle$ leads to an increase in both statistics. The shape of these statistics with respect to u_0 is attributed to the direction of change in the productivity of the indifferent enterprise under the two monotone equilibria.

To conclude, high skewness, as in the $\langle N, I \rangle$ equilibrium, informs us that, relatively speaking, some enterprises produce very high output as a consequence of both PAM between complementary assets and integration at the top of the productivity ladder. Skewness in the $\langle I, N \rangle$ equilibrium, on the other hand, is lower due to the countervailing force of integration at the bottom of the productivity ladder.

5.2 | Empirical relevance

We present several testable implications related to the significance of the reservation utility of *B* producers, u_0 , and the measure-adjusted MPR, $\xi(a, b)$, in shaping firm boundary decisions as well as firm heterogeneity. Both scarcity of *A* supplier assets (e.g., if the market for within industry purchasers of capital, i.e., *A* suppliers is thin) and relatively low productivity of the *B* units contribute to low values of $\xi(a, b)$. Consequently, this results in lower bargaining power for the *B* units and a flatter equilibrium utility function, u(b), regardless of the specific values of u_0 .²⁵ Although an empiricist can estimate the measure-adjusted MPR, $\xi(a, b)$ from the production functions and type distributions, finding an adequate empirical measure of u_0 is not a straight forward task. One possible interpretation is that the reservation utility, u_0 , reflects the *specificity of assets* of the *B* units. The more specific the *B* producer asset is to the production process in a given industry, the less valuable it is in a different market. Consequently, u_0 will be relatively small, because *A* producers do not need to worry as much about inter-market competition for *B* units.²⁶ Once, the empirical measures of u_0 and $\xi(a, b)$ are settled, our results imply:

Implication 1. In industries wherein assets of B producers are of high (low) specificity, and the measure-adjusted MPR is low (high), low productivity firms integrate and high productivity firms stay separate. By contrast, in industries with intermediate specificity of assets of B producers and the measure-adjusted MPR is close to one, low productivity firms stay separate and high productivity firms integrate.

The above implication finds resonance in Williamson (1981) who asserts that the decision to vertically integrate is positively related to asset specificity in a given industry. However, the motive for vertical or lateral integration in our model is different. In Williamson's (1981) work, when firms are required to invest in assets with high specificity, they are more exposed to holdup problems by suppliers. This increased transaction cost can be better mitigated by vertical integration. In contrast, our model posits that high degree of specificity of assets of *B* suppliers in a given industry would induce firms to integrate (due to the effect it has on the ability of units to coordinate). Moreover, differences in asset specificity help trace differences across industries. Kermani and Ma (2021) find that the liquidation recovery rate of inventory is very high (i.e., low asset specificity) in the auto dealers industry as opposed to the market for restaurants where this rate is very low. Therefore, our result would predict a lesser degree of vertical integration in industries with low levels asset specificity, for example, auto dealers.

Our second set of implications are with regard to the distribution of equilibrium expected output:

Implication 2. In industries wherein assets of *B* producers are of high (low) specificity, and the measure-adjusted *MPR* is low (high), are more likely to have low output variance and skewness. On the other hand, in industries with intermediate specificity of assets of *B* producers, and a measure-adjusted MPR close to one, are more likely to have high output variance and skewness.

This implication is particularly helpful when the equilibrium organizational structure is not observed in the industry-level data, since we can infer the likely equilibrium using variations in the moments of the output distribution over time and across industries. The relationship between the share of integrated enterprises, and the variance or skewness of expected output provides sufficient variation to distinguish between the $\langle I, N \rangle$ and $\langle N, I \rangle$ equilibria.

6 | CONCLUDING REMARKS

In this paper, we present a very general and empirically relevant market description that encompasses many industries. We have examined the interplay between asset productivity heterogeneity and firm boundaries (integration vs. nonintegration) in determining firm productivity heterogeneity. Under reasonable conditions, matching is PAM with respect to asset productivity. Because each enterprise chooses between two forms of organization, the generalized increasing differences condition of Legros and Newman (2007), and hence, PAM may fail to hold. We provide a novel condition—the *congruent marginal contributions* property—which suffices GID in an environment wherein utility is imperfectly transferable, and guarantees PAM. Our analysis also offers the theoretical foundation for a novel organizational pattern—low productivity firms integrate, while high productivity ones stay separate—which can be tested empirically. Under another set of conditions, there also exists an equilibrium with the opposite organizational pattern: integration at the top of the productivity ladder. Which pattern emerges in equilibrium crucially depends on how the surplus within a firm is shared between the asset owners, which in turn is a function of asset specificity and the marginal productivity relative to the scarcity of the assets.

The recent merger guidelines from the US Department of Justice explicitly recognize that mergers involving complementary inputs, where neither input is upstream nor downstream, may generate a "range of potentially cognizable efficiencies that benefit competition and consumers."²⁷ Our analysis provides a framework for evaluating these types of mergers and their impacts on market efficiency and innovation.

ACKNOWLEDGMENTS

We are grateful to Daniel Spulber (the editor), a coeditor and two anonymous referees for insightful suggestions that helped improve the paper. We also thank Hector Chade, David Pérez-Castrillo, Edona Reshidi (discussant at IIOC 2022), Tridib Sharma, Dongsoo Shin, seminar participants at Santa Clara University, and conference participants at IIOC 2022 (Boston), the 16th Meeting of the Society for Social Choice and Welfare 2022 (Mexico City), EWET 2022 (Warsaw), ASSET 2022 (Crete) and SAET 2023 (Paris) for helpful comments. The views expressed in this paper are

19

urnal of Economics &

15309134, 0, Dow

from https://onlinelibrary.wiley.

.com/doi/10.1111/jems.12564 by Centro De Investigacion Y Docencia Economicas, A.C., Wiley Online Library on [14/11/2023]. See the Terms

and Conditions

(hup

Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Common

those of the authors and do not necessarily represent the U.S. Bureau of Economic Analysis or the U.S. Department of Commerce.

ORCID

Kanişka Dam 💿 http://orcid.org/0000-0001-7418-623X Konstantinos Serfes 💿 http://orcid.org/0000-0002-0449-6774

NOTES

- ¹ For example, in the airline industry, "major" and "regional" airlines must coordinate to connect smaller cities to the major hubs (e.g., Forbes & Lederman, 2011), while in the healthcare industry, hospitals and skilled nursing facilities (SNFs) must coordinate to guarantee the best health outcome for each patient. Some major airlines are integrated with the regionals, and some hospitals are integrated with the SNFs (e.g., Zhu et al., 2018).
- ² For example, in a model of managerial incentives under endogenous matching, Alonso-Paulí and Pérez-Castrillo (2012) analyze the choice between incentive and Codes of Best Practice contracts. Macho-Stadler et al. (2014) show the robust coexistence of two contracting modes—namely, short- and long-term contracts in a labor market where heterogeneous firms are endogenously matched with heterogeneous workers.
- ³ In the healthcare industry, as of 2015, approximately 20% of all Medicare fee-for-service hospital admissions ended in skilled nursing facility (SNF) stays (see Zhu et al., 2018). In this example, hospitals and SNFs are the two complementary units that should coordinate to guarantee the best health outcome for each patient. Traditionally, hospitals and SNF's received separate payments for the care they provide. To reduce spending and improve quality of care, Medicare recently introduced bundled payment programs that link payments for multiples services related to a single episode of care.
- ⁴ We normalize the price of the single homogeneous good to 1.
- ⁵ We can think of a situation wherein each organization runs a project whose initial outlay is 1, and hence the output in case of success must at least cover the project cost, that is, $z(a, b) \ge 1$ so that all firms are viable. We shall discuss later the implications if $z(a, b) \in (0, 1)$.
- ⁶ We assume quadratic probability of success and effort cost functions for tractability. All the subsequent results hold under very general functional forms such as (a) $\pi(e_A, e_B) = 1 p(e_B e_A)$ with p(0) = 0, p(1) = 1, p', p'' > 0, and (b) $C_A(e_A) = C(e_A)$ and $C_B(e_B) = C(1 e_B)$ with C', C'' > 0.
- ⁷ The two units supply complementary inputs to produce a single, homogeneous good. If we think of enterprises as vertical relationships, one unit, say, *A* may be named the "upstream" producer, and the other, the "downstream" producer. In our model, lateral and vertical relationships are somewhat equivalent because the sole motive for integration is to improve coordination among the units which is achieved by conferring the decision making rights on a third party. We do not consider vertical integration in a more traditional sense where the rights to make decisions belong to the integrated entity, and in which there are the usual efficiency gains such as ameliorating the problem of double marginalization.
- ⁸ In a model with a continuum of types, a particular bargaining protocol is irrelevant, and hence, assuming a take-it-or-leave-it bargaining protocol is innocuous. This is because, due to the continuum assumption, the factor owners do not earn rents over their next best opportunity within the market, as the next competitor type is arbitrarily close. However, in a model with discrete types there would be match-specific rents left to bargain over.
- ⁹ When z < 1, we have $\frac{z^2}{1+z} > z \frac{1}{2}$. In this case, $\Phi^N(z, u)$ always lies above $\Phi^I(z, u)$, that is, nonintegration strictly dominates integration irrespective of the revenue shares for the *A* and *B* units.
- ¹⁰ An alternative form of integration would be wherein the decision right is conferred to one of the two unit managers, say *A*. In this case the *A* manager chooses both e_A and e_B to maximize her expected utility

$$\pi(e_A, e_B)(1-s)z - e_A^2$$

which yields managerial actions $e_A = e_B = 0$ and firm surplus z - 1. Because $\frac{2z^2}{1+2z} > z - 1$, this form of organization is also strictly dominated by nonintegration in the absence of incentive problem.

¹¹ With a higher *b* or equivalently, higher *z*, the nonlinear part of the frontier may become flatter or steeper. However, the nature of shifts is not relevant for this analysis.

¹² Note that IL^- and IL^+ do not cross each other because $u_H(z) - u_L(z) = \frac{z}{1+z} > 0$ for any $z \ge 1$.

- ¹³ In fact, GID is a sufficient condition for PAM, and it is necessary if, for any type distributions, PAM is the equilibrium matching.
- ¹⁴ See Chade et al. (2017) for an excellent discussion on the equivalence between GID and the single-crossing condition.

- ¹⁵ Condition (13) can be written as, which can be written as $\psi(b'', a'', v) \psi(b', a'', v) \ge \psi(b'', a', v) \psi(b', a', v)$, which is the increasing differences condition asserting that the marginal function $\psi(b'', a, v) \psi(b', a, v)$ is increasing in *a* for each *v*.
- ¹⁶ The right panel of Figure 3 depicts one of the many possible ways to violate supermodularity of $\psi(b, a, v)$. The only thing we require is that $\psi(b, a'', v'')$ is flatter relative to $\psi(b, a', v')$ at some (b', u'), that is, (13) does not hold for the types a' and a'' at this point. Then, we can always find a point on $\psi(b, a', v')$, which lies to the northeast of (b', u') at which GID is violated because for a range of values of b to the right of $b', \psi(b, a', v')$ lies above $\psi(b, a'', v'')$. Also, in Figure 3 the indifference curves are drawn with a single kink, but the result of Proposition 1 holds even if an indifference curve has more than one kink.
- ¹⁷ Let $x(b, a) \equiv b^2 a$ and $y(b, a) \equiv ba^2$ be two real-valued supermodular functions for a > 0 and b > 0. Note that $y_a(b, a) x_a(b, a) \ge 0$ and $y_b(b, a) - x_b(b, a) \le 0$, that is, the congruent marginal contributions property fails to hold for any (a, b) such that $1/2 \le a/b \le 2$.
- ¹⁸ Shimer and Smith (2000) considers an economy with prefectly transfereable utility and search frictions to show that the standard supermodularity of the production function may not be sufficient for PAM. They assert that if $z_b(a, b)$ and $z_{ab}(a, b)$ are log-supermodular, then the economy exhibits PAM. The aforementioned conditions are different from the log-supermodularity of z(a, b). For instance, the condition of the first derivative being log-supermodular is equivalent to the fact that z(a, b) becomes more *concave* in *b* as *a* increases.
- ¹⁹ For any given v, let $x(b, a) \equiv \psi^N(b, a, v)$ and $y(b, a) \equiv \psi^I(b, a, v)$.
- 20 In the proof of Proposition 2, we prove the case for more than one kinks.
- ²¹ Such assumption can be rationalized as vertical or lateral integration may often imply a loss of the firm revenue due to the presence of diverse transaction costs (e.g., Grossman & Helpman, 2002) or the unit managers may have to give up a fraction of output produced to *HQ* in the pursuit of a more favorable action.
- ²² Whenever d = I, the ODE assumes a simple form and an analytical solution can be easily obtained. However, for d = N, that is, for values of u(b) in the interval $[U_L(b), U_H(b)]$, the ODE is much more complicated. If $u_0 > 0$, there is a unique solution to the ODE (see Simmons, 2017, theorem B, p. 634). However, a closed-form analytical solution cannot be obtained. In the examples that follow, we solve the ODE numerically. The codes are available upon request.
- ²³ These functional forms are meant to emulate real world situations as documented in previous literature, but the results in this section extend to a more general class of production functions and productivity distributions. Additionally, the truncated Pareto type distributions are chosen to match the productivity distributions used in Melitz and Redding (2015).
- ²⁴ In this example, we have $\gamma_a = \gamma_b = \gamma$, that is, *a* and *b* have the same distribution, and hence, $\xi(a, b) = (\beta/\alpha)(b/a)^{\gamma}$. Thus, $\xi(\mu(b), b) = \beta/\alpha$. Alternatively, if we assume $\alpha = \beta$, that is, *a* and *b* are equally productive, but allow the type distributions to be different in that $\gamma_a \neq \gamma_b$, the measure-adjusted MPR would be:

$$\xi(a,b) = \frac{\gamma_a (1 - (\underline{\theta}/\bar{\theta})^{\gamma_b})}{\gamma_b (1 - (\underline{\theta}/\bar{\theta})^{\gamma_a})} \Big(\frac{\underline{\theta}}{a}\Big)^{\gamma_a} \Big(\frac{\underline{b}}{\underline{\theta}}\Big)^{\gamma_b}.$$

Thus, we have the same general measure-adjusted MPR, which is driven by distributional differences instead of productivity differences. However, we should note that the measure-adjusted MPR is not the same along the equilibrium path, because the matching function is not given by $\mu(b) = b$ as in the former case.

- ²⁵ Ramey and Shapiro (2001) assert that market thinness can contribute to hugely discounted prices for capital even within industry. This difficulty in finding a quality match for a firm's capital can be directly related to our model through the measure-adjusted MPR.
- ²⁶ For example, Kermani and Ma (2021) use the liquidation recovery rate of property, plant, and equipment (PPE), and inventory, among others, as proxies for asset specificity in a given industry.
- ²⁷ See Example 6 in Vertical Merger Guidelines (June, 2020), where the two complementary inputs are batteries and motors in the manufacturing of electric scooters (U.S. Department of Justice and the Federal Trade Commission, 2020).
- ²⁸ The proof of this assertion is cumbersome. The *Mathematica* codes are available upon request.

REFERENCES

Abowd, J., Haltiwanger, J., Jarmin, R., Lane, J., Lengermann, P., McCue, K., McKinney, K., & Sandusky, K. (2005). The relation among human capital, productivity, and market value: Building up from micro evidence. In C. Corrado, J. Haltiwanger, & D. Sichel (Eds.), *Measuring capital in the new economy* (pp. 153–198). University of Chicago Press.

Alonso-Paulí, E., & Pérez-Castrillo, D. (2012). Codes of best practice in competitive markets for managers. *Economic Theory*, 49, 113–141. Antràs, P., & Helpman, E. (2004). Global sourcing. *Journal of the Political Economy*, 112, 552–580.

Atalay, E., Hortaçsu, A., & Syverson, C. (2014). Vertical integration and input flows. The American Economic Review, 104, 1120-1148.

15309134, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jems.12564 by Centro De Investigacion Y Docencia

micas, A.C., Wiley Online Library on [14/11/2023]. See the Terms

and Co

(hup

ons) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons

- Benguria, F. (2021). The matching and sorting of exporting and importing firms: Theory and evidence. *Journal of International Economics*, *131*, 1034300.
- Bloom, N., & Van Reenen, J. (2010). Why do management practices differ across firms and countries? *Journal of Economic Perspectives*, 24, 203–224.
- Braguinsky, S., Ohyama, A., Okazaki, T., & Syverson, C. (2015). Acquisitions, productivity, and profitability: Evidence from the Japanese cotton spinning industry. *American Economic Review*, 105(7), 2086–2119.
- Chade, H., Eeckhout, J., & Smith, L. (2017). Sorting through search and matching models in economics. *Journal of Economic Literature*, 55, 493–544.
- Corcos, G., Irac, D., Mion, G., & Verdier, T. (2013). The determinants of intrafirm trade: Evidence from French firms. *The Review of Economics and Statistics*, 95, 825–838.
- Dam, K., & Serfes, K. (2021). A price theory of vertical and lateral integration under productivity heterogeneity. *The Journal of Industrial Economics*, 69, 751–784.
- Dragusanu, R. (2014). Firm-to-firm matching along the global supply chain. Mimeo, Harvard University.
- Forbes, S., & Lederman, M. (2011). Does vertical integration affect firm performance? Evidence from the airline industry. *RAND Journal of Economics*, 41, 765–790.
- Garicano, L. (2000). Hierarchies and the organization of knowledge in production. Journal of Political Economy, 108, 874-904.
- Grossman, G., & Helpman, E. (2002). Integration versus outsourcing in industry equilibrium. *Quarterly Journal of Economics*, *117*, 85–120. Hart, O., & Hölmstrom, B. (2010). A theory of firm scope. *The Quarterly Journal of Economics*, *125*, 483–513.
- Kehrig, M. (2015). *The cyclical nature of the productivity distribution*. U.S. Census Bureau Center for Economic Studies Paper No. CES-WP-11-15.
- Kermani, A., & Ma, Y. (2021). Asset specificity of non-financial firms. Mimeo, University of California, Berkeley and University of Chicago.
- Legros, P., & Newman, A. (2007). Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities. *Econometrica*, 75, 1073–1102.
- Legros, P., & Newman, A. (2013). A price theory of vertical and lateral integration. The Quarterly Journal of Economics, 128, 727-770.
- Macho-Stadler, I., Pérez-Castrillo, D., & Porteiro, N. (2014). Coexistence of long-term and short-term contracts. Games and Economic Behavior, 86, 145-164.
- Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica, 71, 1695–1725.
- Melitz, M., & Redding, S. (2015). New trade models, new welfare implications. American Economic Review, 105(3), 1105–1146.
- Ramey, V., & Shapiro, M. (2001). Displaced capital: A study of aerospace plant closings. Journal of Political Economy, 109, 958-992.
- Shimer, R., & Smith, L. (2000). Assortative matching and search. Econometrica, 68, 349-369.
- Simmons, G. (2017). Differential equations with applications and historical notes (3rd ed.). CRC Press.
- Sugita, Y., Teshima, K., & Seira, E. (2021). Assortative matching of exporters and importers. The Review of Economics and Statistics. (Forthcoming).
- Syverson, C. (2011). What determines productivity? Journal of Economic Literature, 49, 326-365.
- Teulings, C. (1995). The wage distribution in a model of the assignment of skills to jobs. Journal of Political Economy, 103, 280-315.
- U.S. Department of Justice and the Federal Trade Commission. (2020). Vertical Merger Guidelines. https://www.ftc.gov/system/files/ documents/reports/us-department-justice-federal-trade-commission-vertical-merger-guidelines/vertical_merger_guidelines_6-30-20.pdf
- Williamson, O. (1981). The economics of organization: The transaction cost approach. American Journal of Sociology, 87, 548-577.
- Zhu, J. M., Patel, V., Shea, J. A., Neuman, M. D., & Werner, R. M. (2018). Hospitals using bundled payment report reducing skilled nursing facility use and improving care integration. *Health Affairs*, *37*, 1282–1289.

How to cite this article: Dam, K., Ripperger-Suhler, D., & Serfes, K. (2023). Two-sided productivity heterogeneity, firm boundaries, and assortative matching. *Journal of Economics & Management Strategy*, 1–32. https://doi.org/10.1111/jems.12564

APPENDIX A: PROOFS

Proof of Lemma 1. We first analyze the optimal contract under nonintegration. Under this ownership structure, the optimal contract is obtained by maximizing (3) subject to (4), (5), and (6). Note that V_A and U_B are strictly concave in e_A and e_B , respectively, and hence, (5) and (6) can be replaced by their respective first-order conditions. Setting these first-order conditions equal to zero, we obtain

ournal of Economics & lanagement Strategy

$$e_A(s) = \frac{(1-s)z}{1+z}$$
 and $1 - e_B(s) = \frac{sz}{1+z}$.

Substituting the above into the expressions of V_A and U_B , the maximization problem reduces to

$$\max_{s} \quad V_A(s) \equiv \frac{z^2(1-s)(1+z+s)}{(1+z)^2}$$

subject to
$$U_B(s) \equiv \frac{z^2s(2+z-s)}{(1+z)^2} = u.$$

In general, the constraint is given by $U_B(s) \ge u$. However, it is easy to show that the constraint binds at the optimum. Solving *s* from the binding constraint, we get

$$s(z, u) = 1 + \frac{z}{2} - \frac{\sqrt{z^2(2+z)^2 - 4u(1+z)^2}}{2z}$$

Substituting s = s(z, u) in $V_A(s)$, we obtain the bargaining frontier under nonintegration, that is, $\Phi^N(z, u) \equiv \phi^N(a, b, u)$.

Next, consider the optimal contract under integration. The HQ solves

$$\max_{\{e_A,e_B,s_{HQ}\}} \pi(e_A,e_B)s_{HQ}\cdot z.$$

Given that s_{HQ} , the first-order conditions with respect to e_A and e_B yield $e_A = e_B$, which imply $\pi(e_A, e_B) = 1$. The aggregate managerial cost is given by $e_A^2 + (1 - e_B)^2$ which is minimized at $e_A = e_B = \frac{1}{2}$, and the minimum cost is given by $\frac{1}{2}$. Because *HQ* breaks even, the entire surplus accrues to the managers, which is given by z, and hence the net surplus is given by $z - \frac{1}{2}$. Because under integration the utility is perfectly transferable, we have $v + u = z - \frac{1}{2}$. Thus, the bargaining frontier under integration is given by $\Phi^I(z, u) = z - \frac{1}{2} - u$.

Finally, we analyze the choice of ownership structure of a given firm (a, b) whose output is $z \equiv z(a, b)$. The firm chooses nonintegration over integration if $\Phi^N(z, u) \ge \Phi^I(z, u)$. Because both the frontiers are downwardsloping and symmetric [around the 45°-line], $\Phi^N(z, u)$ is strictly concave in u and $\Phi^I(z, u)$ is linear in u for each z, there are two possibilities—either they do not intersect each other or they intersect exactly twice. First, note that, under nonintegration, the minimum surplus corresponds to s = 0 or s = 1, which is given by

$$\frac{z^2}{1+z}$$

On the other hand, the maximum surplus is associated with equal revenue shares, that is, $s = \frac{1}{2}$, which is given by

$$\left(\frac{3}{2}+z\right)\left(\frac{z}{1+z}\right)^2.$$

Recall also that, under integration, the constant firm surplus is given by $z - \frac{1}{2}$ (i.e., $\Phi^{I}(z, u)$ is linear in u for each z). It is easy to see that

• If $z \in (0, 1)$, then

$$\frac{z^2}{1+z} > z - \frac{1}{2}$$

that is, $\Phi^N(z, u)$ stays strictly above $\Phi^I(z, u)$ for all u, that is, nonintegration strictly dominates integration. However, this does not emerges in the present context, as we have assumed $z \ge 1$.

• On the other hand, for any $z \ge 1$, we have

24 WILEY-

$$\frac{z^2}{1+z} < z - \frac{1}{2} < \left(\frac{3}{2} + z\right) \left(\frac{z}{1+z}\right)^2,$$

that is, the two frontiers intersect each other exactly twice at $(u_L, \Phi(z, u_L))$ and $(u_H, \Phi(z, u_H))$.

So, what remains to prove is that $0 \le u_L(z) < u_H(z) \le \frac{z^2}{1+z}$. Recall that

$$u_L(z) = \frac{(z-1)(1+2z)}{4(1+z)}$$
 and $u_H(z) = \frac{2z^2+3z-1}{4(1+z)}$.

First, note that $z \ge 1$ implies that $u_L(z) \ge 0$. Next,

$$u_H(z)-u_L(z)=\frac{z}{1+z},$$

and hence, $u_H(z) > u_L(z)$ for any $z \ge 1$. Finally,

$\frac{z^2}{1+z} - u_H(z) = \frac{(z-1)(2z-1)}{4(1+z)},$

and hence, $u_H(z) \leq \frac{z^2}{1+z}$ for $z \geq 1$.

Proof of Proposition 1. We first prove the sufficiency of (13). Take any a'' > a', b'' > b' and u'' > u', and let $\phi(a', b', u') = \phi(a', b'', u'') = v$, that is, (b', u') and (b'', u'') are on the same indifference curve of a' that yields v. Thus, $u'' - u' = \psi(b'', a', v) - \psi(b', a', v)$. Now, consider the indifference curve of a'' that yields v to her and passes through (b', u'). Further, consider the point (b'', \hat{u}) on it, that is, $\phi(a'', b', u') = \phi(a'', b'', \hat{u}) = v$ (clearly, $\hat{u} > u'$). Thus, $\hat{u} - u' = \psi(b'', a'', v) - \psi(b', a'', v)$. Condition (13) implies that $\hat{u} \ge u''$. Because $\phi(a, b, u)$ is strictly decreasing in u, we have $v' = \phi(a'', b'', u'') \ge \phi(a'', b'', \hat{u}) = \phi(a'', b', u')$. Thus, $\phi(a', b', u'') = \phi(a', b'', u'')$ implies that $\phi(a'', b'', u'') \ge \phi(a'', b', u')$, which is nothing but GID.

Next, we show the necessity of (13). Given two arbitrary types a' and a'' with a' < a'', suppose supermodularity of $\psi(b, a, v)$, that is, (13) is violated for these two types around b = b', that is, $\psi(b, a'', v)$ crosses $\psi(b, a', v)$ from above at b = b' (as shown in the right panel of Figure 3). Then, there is b'' > b' such that

$$\psi(b'', a', v) - \psi(b'', a'', v) > \psi(b', a', v) - \psi(b', a'', v) = 0 \quad \Rightarrow \quad \psi(b'', a', v) > \psi(b'', a'', v).$$

Let $\hat{u} = \psi(b'', a'', v)$ and $u'' = \psi(b'', a', v)$. Therefore, the above inequality is equivalent to $u'' > \hat{u}$. Also, $\phi(a', b', u') = \phi(a', b'', u'')$ because (b', u') and (b'', u'') lie on the same indifference curve of a'. Likewise, $\phi(a'', b', u') = \phi(a'', b'', \hat{u})$ as (b', u') and (b'', \hat{u}) are on the same indifference curve of a''. Because $\phi(a, b, u)$ is strictly decreasing in u, we have

$$\phi(a'', b'', u'') < \phi(a'', b'', \hat{u}) = \phi(a'', b', u'),$$

that is, GID does not hold for these two given types a' and a'' at b = b''.

Proof of Proposition 2. Let $\hat{b}(a)$ solve x(b, a) = y(b, a) for each a. The congruent marginal contributions conditions above imply that $\hat{b}(a)$ is decreasing in a. To see this, differentiating $x(\hat{b}(a), a) = y(\hat{b}(a), a)$ with respect to a, we obtain

DAM ET AL.

$$\frac{d\hat{b}}{da} = -\frac{y_a(\hat{b}(a), a) - x_a(\hat{b}(a), a)}{y_b(\hat{b}(a), a) - x_b(\hat{b}(a), a)}.$$

Note that (CMC_a^+) and (CMC_b^+) together or (CMC_a^-) and (CMC_b^-) together imply that the numerator and denominator of the above expression are of the same sign, and hence, $d\hat{b}/da \le 0$. Now, we show that the function $w(b, a) = \max\{x(b, a), y(b, a)\}$ is supermodular, that is, for any b'' > b' and a'' > a', we have

$$w(b', a') + w(b'', a'') \ge w(b'', a') + w(b', a'').$$
(A1)

urnal of Economics &

Notice that x(b, a) = y(b, a) may have multiple solutions in *b* for some *a*, that is, the graph of w(b, a) may have more than one kink. For our purpose, we analyze the following two situations—(i) there are two solutions of x(b, a) = y(b, a) in *b* for each $a, \hat{b}_1(a)$ and $\hat{b}_2(a)$ with $\hat{b}_1(a) < \hat{b}_2(a)$ (as shown in the right panel of Figure 4), and there is a unique solution to x(b, a) = y(b, a) in *b* for each a, \hat{b} . The result can be extended to the case with more than two kinks of w(b, a). When there are two solutions to $x(b, a) = y(b, a), \hat{b}_1(a)$ and $\hat{b}_2(a)$, the function w(b, a) can be written as

$$w(b, a) = \begin{cases} x(b, a), & \text{for all } b < \hat{b}_1(a), \\ y(b, a), & \text{for all } \hat{b}_1(a) \le b \le \hat{b}_2(a), \\ x(b, a), & \text{for all } \hat{b}_2(a) < b. \end{cases}$$

Consider any two b' and b'' with b' < b'', and a' and a'' with a' < a''. Because $d\hat{b}/da \le 0$, we have $\hat{b}_1(a'') \le \hat{b}_1(a')$, and $\hat{b}_2(a'') \le \hat{b}_2(a')$. For the interest of length, we assume that $\hat{b}_1(a') < \hat{b}_2(a'')$. This assumption is innocuous as an increase in a from a' to a'' is a local change. Depending on the location of b'', there may be more than one possible cases to consider. For example, if $\hat{b}_1(a') < b'' \le \hat{b}_2(a'')$, there are three corresponding situations—namely, (i) $b' < \hat{b}_1(a'')$, (ii) $\hat{b}_1(a'') < b' \le \hat{b}_1(a')$, and (iii) $\hat{b}_1(a') < b'' \le \hat{b}_2(a'')$. There are 15 possible cases (see Figure A1) which we analyze below.

Case 1: $b' < b'' < \hat{b}_1(a'')$. In this case, w(b, a) = x(b, a) for b = b', b'' and a = a', a''. Therefore, the supermodularity of x(b, a) implies that of w(b, a).

Case 2: $b' < \hat{b}_1(a'') < b'' < \hat{b}_1(a')$. In this case, w(b', a') = x(b', a'), w(b', a'') = x(b', a''), w(b', a'') = x(b', a''), w(b'', a'') = x(b'', a''). Thus, w(b, a) supermodular if and only if

$$x(b', a') + y(b'', a'') \ge x(b'', a') + x(b', a'').$$
(A2)

Because $\hat{b}_1(a'') < b'' < \hat{b}_1(a')$, we have $y(b'', a'') \ge x(b'', a'')$. Therefore,

$$x(b',a') + y(b'',a'') \ge x(b',a') + x(b'',a'') \ge x(b'',a') + x(b',a''),$$

and hence, w(b, a) is supermodular. The last inequality follows from the fact that x(b, a) is a supermodular function.

Case 3: $\hat{b}_1(a'') < b' < b'' < \hat{b}_1(a')$. In this case, w(b', a') = x(b', a'), w(b', a'') = y(b', a''), w(b'', a'') = y(b', a''). Thus, w(b, a) is supermodular if and only if

$$x(b', a') + y(b'', a'') \ge x(b'', a') + y(b', a'').$$
(A3)

Because x(b, a) is supermodular we have

$$x(b', a') + x(b'', a'') \ge x(b'', a') + x(b', a'').$$

On the other hand, (CMC_b^+) for a = a'' implies that

$$y(b'', a'') - x(b'', a'') \ge y(b', a'') - x(b', a'')$$

25



FIGURE A1 Possible values of b' and b'' with b' < b'' when the graph of w(b, a) has two kinks.

Journal of Economics & Management Strategy -WILEY

Case 4: $b' < \hat{b}_1(a'')$ and $\hat{b}_1(a') < b'' < \hat{b}_2(a'')$. In this case, w(b', a') = x(b', a'), w(b', a'') = x(b', a''), w(b', a'') = y(b'', a') and w(b'', a'') = y(b'', a''). Thus, w(b, a) is supermodular if and only if

$$x(b', a') + y(b'', a'') \ge y(b'', a') + x(b', a'').$$
(A4)

Because x(b, a) is supermodular we have

$$x(b', a') + x(b'', a'') \ge x(b'', a') + x(b', a'').$$

On the other hand, (CMC_a^+) for b = b'' implies that

$$y(b'', a'') - x(b'', a'') \ge y(b'', a') - x(b'', a').$$

Adding the last two inequalities, we obtain (A4).

Case 5: $\hat{b}_1(a'') < b' < \hat{b}_1(a') < b'' < \hat{b}_2(a'')$. In this case, w(b', a') = x(b', a'), w(b', a'') = y(b', a''), w(b'', a'') = y(b'', a'') and w(b'', a'') = y(b'', a''). Thus, w(b, a) is supermodular if and only if

$$x(b', a') + y(b'', a'') \ge y(b'', a') + y(b', a'').$$
(A5)

Because $b' < \hat{b}_1(a')$, we have $x(b', a') \ge y(b', a')$. Therefore,

$$x(b', a') + y(b'', a'') \ge y(b', a') + y(b'', a'') \ge y(b'', a') + y(b', a'')$$

and hence, w(b, a) is supermodular. The last inequality follows from the fact that y(b, a) is a supermodular function.

Case 6: $\hat{b}_1(a') < b' < b'' < \hat{b}_2(a'')$. In this case, w(b, a) = y(b, a) for b = b', b'' and a = a', a''. Therefore, the supermodularity of y(b, a) implies that of w(b, a).

Case 7: $b' < \hat{b}_1(a'')$ and $\hat{b}_2(a'') < b'' < \hat{b}_2(a')$. In this case, w(b', a') = x(b', a'), w(b', a'') = x(b', a''), w(b'', a'') = x(b', a''). Thus, w(b, a) is supermodular if and only if

$$x(b', a') + x(b'', a'') \ge y(b'', a') + x(b', a'').$$
(A6)

We first prove the following inequality

$$y(b', a') + x(b'', a'') \ge y(b'', a') + x(b', a'').$$
(A7)

Because y(b, a) is supermodular we have

 $y(b', a') + y(b'', a'') \ge y(b'', a') + y(b', a'').$

On the other hand, (CMC_b^-) for a = a'' implies that

$$x(b'', a'') - y(b'', a'') \ge x(b', a'') - y(b', a'').$$

Adding the last two inequalities, we obtain (A7). Because $b' < \hat{b}_1(a')$, we have $x(b', a') \ge y(b', a')$. Therefore,

$$x(b',a') + x(b'',a'') \ge y(b',a') + x(b'',a'') \ge y(b'',a') + x(b',a''),$$

which proves inequality (A6).

Case 8: $\hat{b}_1(a'') < b' < \hat{b}_1(a')$ and $\hat{b}_2(a'') < b'' < \hat{b}_2(a')$. In this case, w(b', a') = x(b', a'), w(b', a'') = y(b', a''), w(b', a'') = y(b', a''), w(b'', a'') = y(b'', a'') and w(b'', a'') = x(b'', a''). Thus, w(b, a) is supermodular if and only if

 $x(b', a') + x(b'', a'') \ge y(b'', a') + y(b', a'').$

Because $b' < \hat{b}_1(a')$, we have $x(b', a') \ge y(b', a')$, and $x(b'', a'') \ge y(b'', a'')$ as $b'' > \hat{b}_2(a'')$. Therefore,

$$x(b',a') + x(b'',a'') \ge y(b',a') + y(b'',a'') \ge y(b'',a') + y(b',a''),$$

which proves inequality (A8). The last inequality follows from the fact that y(b, a) is a supermodular function.

Case 9: $\hat{b}_1(a') < b' < \hat{b}_2(a'') < b'' < \hat{b}_2(a')$. In this case, w(b', a') = y(b', a'), w(b', a'') = y(b', a''), w(b', a'') = y(b', a''), w(b', a'') = y(b', a''), w(b', a'') = y(b', a''). Thus, w(b, a) is supermodular if and only if

$$y(b', a') + x(b'', a'') \ge y(b'', a') + y(b', a'').$$
 (A9)

Because $b'' > \hat{b}_2(a'')$, we have $x(b'', a'') \ge y(b'', a'')$. Therefore,

$$y(b', a') + x(b'', a'') \ge y(b', a') + y(b'', a'') \ge y(b'', a') + y(b', a''),$$

which proves inequality (A10). The last inequality follows from the fact that y(b, a) is a supermodular function.

Case 10: $\hat{b}_2(a'') < b' < b'' < \hat{b}_2(a')$. In this case, w(b', a') = y(b', a'), w(b', a'') = x(b', a''), w(b', a'') = x(b', a''), w(b'', a'') = x(b'', a''). Thus, w(b, a) is supermodular if and only if

$$y(b', a') + x(b'', a'') \ge y(b'', a') + x(b', a'').$$
(A10)

Because y(b, a) is supermodular we have

$$y(b', a') + y(b'', a'') \ge y(b'', a') + y(b', a'').$$

On the other hand, (CMC_b^-) for a = a'' implies that

$$x(b'', a'') - y(b'', a'') \ge x(b', a'') - y(b', a'').$$

Adding the last two inequalities, we obtain (A10).

Case 11: $b' < \hat{b}_1(a'')$ and $\hat{b}_2(a') < b''$. In this case, w(b, a) = x(b, a) for b = b', b'' and a = a', a''. Therefore, the supermodularity of x(b, a) implies that of w(b, a).

Case 12: $\hat{b}_1(a'') < b' < \hat{b}_1(a')$ and $\hat{b}_2(a') < b''$. In this case, w(b', a') = x(b', a'), w(b', a'') = y(b', a''), w(b'', a'') = x(b'', a'') and w(b'', a'') = x(b'', a''). Thus, w(b, a) is supermodular if and only if

$$x(b', a') + x(b'', a'') \ge x(b'', a') + y(b', a'').$$
(A11)

We first prove the following inequality

$$y(b', a') + x(b'', a'') \ge x(b'', a') + y(b', a'').$$
(A12)

Because y(b, a) is supermodular we have

$$y(b', a') + y(b'', a'') \ge y(b'', a') + y(b', a'').$$

On the other hand, (CMC_a^-) for b = b'' implies that

$$x(b'', a'') - y(b'', a'') \ge x(b'', a') - y(b'', a').$$

Adding the last two inequalities, we obtain (A12). Because $b' < \hat{b}_1(a')$, we have $x(b', a') \ge y(b', a')$. Therefore,

(A8)

DAM ET AL.

 $x(b', a') + x(b'', a'') \ge y(b', a') + x(b'', a'') \ge x(b'', a') + y(b', a''),$

which proves inequality (A11).

Case 13: $\hat{b}_1(a') < b' < \hat{b}_2(a'')$ and $\hat{b}_2(a') < b''$. In this case, w(b', a') = y(b', a'), w(b', a'') = y(b', a''), w(b', a'') = x(b'', a'') and w(b'', a'') = x(b'', a''). Thus, w(b, a) is supermodular if and only if

$$y(b', a') + x(b'', a'') \ge x(b'', a') + y(b', a'').$$
(A13)

Because y(b, a) is supermodular we have

$$y(b', a') + y(b'', a'') \ge y(b'', a') + y(b', a'').$$

On the other hand, (CMC_a^-) for b = b'' implies that

$$x(b'', a'') - y(b'', a'') \ge x(b'', a') - y(b'', a').$$

Adding the last two inequalities, we obtain (A14).

Case 14: $\hat{b}_2(a'') < b' < \hat{b}_2(a') < b''$. In this case, w(b', a') = y(b', a'), w(b', a'') = x(b', a''), w(b', a'') = x(b'', a''). Thus, w(b, a) is supermodular if and only if

$$y(b', a') + x(b'', a'') \ge x(b'', a') + x(b', a'').$$
(A14)

Because $\hat{b}_1(a') < b' < \hat{b}_2(a')$, we have $y(b', a') \ge x(b', a')$. Therefore,

$$y(b', a') + x(b'', a'') \ge x(b', a') + x(b'', a'') \ge x(b'', a') + x(b', a''),$$

which proves inequality (A14). The last inequality follows from the fact that x(b, a) is a supermodular function.

Case 15: $\hat{b}_2(a') < b' < b''$. In this case, w(b, a) = x(b, a) for b = b', b'' and a = a', a''. Therefore, the supermodularity of x(b, a) implies that of w(b, a).

The case with a single kink. When x(b, a) = y(b, a) has unique solution $\hat{b}(a)$ in b, w(b, a) has a single kink. There are two possible sub-cases. First, y(b, a) intersects x(b, a) from below, that is, w(b, a) is given by

$$w(b, a) = \begin{cases} x(b, a), & \text{for all } b \le \hat{b}(a), \\ y(b, a), & \text{for all } b > \hat{b}(a). \end{cases}$$

The proof of the proposition in this case is identical to those in Cases 1-6 above wherein $\hat{b}_1(a)$ is replaced by $\hat{b}(a)$ for a = a', a''. The second sub-case is when x(b, a) intersects y(b, a) from below, that is, w(b, a) is given by

$$w(b, a) = \begin{cases} y(b, a), & \text{for all } b \le \hat{b}(a), \\ x(b, a), & \text{for all } b > \hat{b}(a). \end{cases}$$

The proof of the proposition in this case is identical to those in Cases 6, 9, 10, and 13–15 above wherein $\hat{b}_2(a)$ is replaced by $\hat{b}(a)$ for a = a', a''.

The case with more than two kinks. The above proof can easily be extended to when x(b, a) = y(b, a) has more than two solutions in *b*, that is, w(b, a) has more than two kinks. We only require to determine which one of the two functions, x(b, a) or y(b, a), intersects the other from below at each kink. Then, the steps of the above proof can be repeated to obtain the proposition (Figure A1).

Proof of Theorem 1. Theorem 1 follows from Proposition 2. We require to verify that the bargaining frontier, $\phi(a, b, u)$ indeed satisfies the conditions of Proposition 2.

DAM ET AL.

We first show that both $\psi^N(b, a, v)$ and $\psi^I(b, a, v)$ are supermodular in (b, a), that is, both $\phi^N(a, b, u)$ and $\phi^I(a, b, u)$ satisfy GID. Note first that each frontier is differentiable everywhere, and hence, GID under each organizational mode is equivalent to (see Chade et al., 2017)

$$\frac{\partial}{\partial a} \left(-\frac{\phi_2^d(a, b, u)}{\phi_3^d(a, b, u)} \right) \ge 0 \quad \text{for } d = N, I.$$
(A15)

To prove GID of the bargaining frontier under nonintegration, note that $\phi^N(a, b, u) \equiv \Phi^N(z(a, b), u)$, and hence, $\phi_2^N = \Phi_z^N \cdot z_b > 0$ and $\phi_3^N = \Phi_u^N < 0$. Define

$$\sigma(z, u) \equiv -\frac{\Phi_z^N(z, u)}{\Phi_u^N(z, u)} > 0.$$

Thus,

$$\frac{\partial}{\partial a} \left(-\frac{\phi_2^N}{\phi_3^N} \right) = \frac{\partial}{\partial a} (\sigma(z, u) \cdot z_b) = z_a z_b \left(\sigma_z(z, u) + \frac{\sigma(z, u)}{z} \cdot \frac{z z_{ab}}{z_a z_b} \right).$$
(A16)

Let $\delta \equiv zz_{ab}/z_a z_b$. It can be shown that, under quadratic probability of success, if $\delta \ge 1$, the expression on the right-hand-side of (A16) is positive. Note that $\delta \ge 1$ is equivalent to z(a, b) being a log-supermodular function. On the other hand, $\delta < 1$ may lead to the right-hand-side of (A16) being negative.²⁸ Therefore, log-supermodular z(a, b) is a sufficient condition for GID of $\phi^N(a, b, u)$.

Showing that $\phi^{I}(a, b, u)$ satisfies GID is straightforward. Because $\phi^{I}(a, b, u) \equiv \Phi^{I}(z(a, b), u) = z(a, b) - \frac{1}{2} - u$, we have $-(\phi_{2}^{I}/\phi_{3}^{I}) = z_{b}(a, b)$. Note that z(a, b) is log-supermodular, which implies that z(a, b) is strictly supermodular in (a, b), that is, $z_{ab} > 0$. Consequently, $-(\phi_{2}^{I}/\phi_{3}^{I})$ is increasing in a.

Next, we show that, the bargaining frontiers $\psi^N(b, a, v)$ and $\psi^I(b, a, v)$ satisfy the congruent marginal contributions property. Because both frontiers, $\psi^N(b, a, v)$ and $\psi^I(b, a, v)$ are differentiable in *a* and *b*, we can use the differentiable version of the congruent marginal contribution property, that is, condition (14). By differentiating $\phi^d(a, b, \psi^d(b, a, v)) = v$ for each d = I, *N* with respect to *a* and *b*, respectively, we obtain

$$\psi_{a}^{N}(b, a, v) = -\frac{\phi_{1}^{N}}{\phi_{3}^{N}} = \sigma(z, u)z_{a}, \text{ and } \psi_{b}^{N}(b, a, v) = -\frac{\phi_{2}^{N}}{\phi_{3}^{N}} = \sigma(z, u)z_{b},$$
$$\psi_{a}^{I}(b, a, v) = -\frac{\phi_{1}^{I}}{\phi_{3}^{I}} = z_{a}, \text{ and } \psi_{b}^{I}(b, a, v) = -\frac{\phi_{2}^{I}}{\phi_{3}^{I}} = z_{b}.$$

Therefore,

$$\begin{split} \psi_{a}^{I}(b, a, v) &- \psi_{a}^{N}(b, a, v) = [1 - \sigma(z, u)] z_{a}(a, b), \\ \psi_{a}^{I}(b, a, v) &- \psi_{a}^{N}(b, a, v) = [1 - \sigma(z, u)] z_{b}(a, b). \end{split}$$

Because $z_a, z_b > 0$, from the above it follows that

$$\operatorname{sign}\left[\psi_{a}^{I}(b, a, v) - \psi_{a}^{N}(b, a, v)\right] = \operatorname{sign}\left[\psi_{b}^{I}(b, a, v) - \psi_{b}^{N}(b, a, v)\right]$$

for each v, and hence, the congruent marginal contributions property holds.

Let $x(b, a) \equiv \psi^N(b, a, v)$ and $y(b, a) \equiv \psi^I(b, a, v)$ for any given v. From the above, it follows that the conditions of Proposition 2 hold for our economy, and hence, $\psi(b, a, v) = \max\{\psi^N(b, a, v), \psi^I(b, a, v)\}$

supermodular in (b, a) for each v. Thus, we have that $\phi(a, b, u) = \max{\phi^N(a, b, u), \phi^I(a, b, u)}$ satisfies GID, which follows from Proposition 1. So, it follows from Legros and Newman (2007) that the equilibrium matching is PAM.

Proof of Theorem 2. We first find the slopes of the equilibrium indifference loci $U_L(b) \equiv u_L(z(\mu(b), b))$ and $U_H(b) \equiv u_H(z(\mu(b), b))$ where $u_L(z)$ and $u_H(z)$ are given in Lemma 1. Given that $\mu'(b) = f(b)/g(\mu(b))$, we have for $a = \mu(b)$,

$$U'_{L}(b) = u'_{L}(z(a,b)) \left(z_{a}(a,b) \cdot \frac{f(b)}{g(a)} + z_{b}(a,b) \right),$$
(A17)

$$U'_{H}(b) = u'_{H}(z(a,b)) \left(z_{a}(a,b) \cdot \frac{f(b)}{g(a)} + z_{b}(a,b) \right).$$
(A18)

It is immediate to show that $u'_L(z) < u'_H(z)$, and hence, $U'_L(b) < U'_H(b)$ for all b.

Next, we find the slope of the equilibrium utility function, u(b). Let $u(b) \equiv \max\{u_N(b), u_I(b)\}$. Because we can write $\phi^d(a, b, u) \equiv \Phi^d(z(a, b), u)$ for d = N, I, we have for $a = \mu(b)$ and u = u(b)

$$u'_{N}(b) = \left(-\frac{\Phi_{z}^{N}(z(a,b),u)}{\Phi_{u}^{N}(a,b),u}\right) z_{b}(a,b) \equiv \sigma(z(a,b),u) z_{b}(a,b),$$
(A19)

$$u'_{l}(b) = z_{b}(a, b).$$
 (A20)

Now, we analyze the sufficient conditions for diverse equilibrium ownership structures. For that, we take bargaining frontiers $\phi^d(a, b, u)$ for d = N, I as the fundamentals, and hence, the sufficient conditions will involve functions of a, b, u and u_0 .

1. First, consider the equilibrium ownership structure, $\langle I \rangle$. Given that $\underline{z} \equiv z(\underline{a}, \underline{b})$, either $u_0 < u_L(\underline{z})$ or $u_0 > u_H(\underline{z})$ implies that the lowest-productivity firm, $(\underline{a}, \underline{b})$ chooses to integrate.

Consider first $u_0 < u_L(\underline{z})$. If all firms in equilibrium choose to integrate, we have $u(b) = u_I(b)$ for all $b \in [\underline{b}, \overline{b}]$. If $u'_I(b) \leq U'_L(b)$ for all b, u(b) stays strictly below $U_L(b)$ for all b. Thus, a sufficient condition for $\langle I \rangle$ equilibrium is

$$z_b(a,b) \le u'_L(z(a,b)) \left(z_a(a,b) \cdot \frac{f(b)}{g(a)} + z_b(a,b) \right) \Leftrightarrow \xi(a,b) \le \underbrace{\frac{u'_L(z(a,b))}{1 - u'_L(z(a,b))}}_{l_l(z(a,b))}.$$

The above guarantees that $u'_l(b) \leq U'_L(b)$ for all *b*.

Next, consider $u_0 > u_H(\underline{z})$. Using the similar argument as above, if $u'_I(b) \ge U'_H(b)$ for all *b*, every firm chooses *I* in the equilibrium. Thus, a sufficient condition is given by

$$z_{b}(a,b) \geq u'_{H}(z(a,b)) \left\{ z_{a}(a,b) \cdot \frac{f(b)}{g(a)} + z_{b}(a,b) \right\} \Leftrightarrow \xi(a,b) \geq \underbrace{\frac{u'_{H}(z(a,b))}{1 - u'_{H}(z(a,b))}}_{h_{I}(z(a,b))}.$$

Because $u'_L(z) < u'_H(z)$ for all z, we have $l_I(z) < h_I(z)$.

2. Next, consider the equilibrium structure, $\langle N \rangle$. Note that $u_L(\underline{z}) \leq u_0 \leq u_L(\underline{z})$ imply that the lowestproductivity enterprise, $(\underline{a}, \underline{b})$ chooses to stay separate. Therefore, for an $\langle N \rangle$ equilibrium, we require that $U_L(b) < u(b) = u_N(b) < U_H(b)$ for all *b*, which holds if $U'_L(b) \leq u'_N(b) \leq U'_H(b)$. The sufficient conditions for the last set of inequalities are

31

32

-WILEY-

5309134, 0, Down

aded from https://onlinelibrary.wiley.com/doi/10.1111/jems.12564 by Centro De Investigacion Y Docencia Economicas, A.C., Wiley Online Library on [14/11/2023]. See the Terms

and Conditions (https://or

on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Next, consider the equilibrium structure, ⟨ N, I ⟩. As in the ⟨ N ⟩ equilibrium, u_L(<u>z</u>) ≤ u₀ ≤ u_L(<u>z</u>) imply that the lowest-productivity enterprise chooses to stay separate. There are two possibilities. First, u(b) intersects U_H(b) from below only once (as depicted in Figure 4c). Second, u(b) intersects U_L(b) from above only once. The former situation occurs if both u_N(b) and u_I(b) are steeper than U_H(b) for all b ∈ [b, b]. The sufficient condition for u'_N(b) > U'_H(b) is ξ(a, b) > h_N(z(a, b), u), and that for u'_I(b) > U'_H(b) is ξ(a, b) > h_I(z(a, b)). Thus,

$$\xi(a, b) > \max\{h_N(z(a, b), u), h_I(z(a, b))\}$$

guarantees that u(b) is steeper than $U_H(b)$ for all b, and hence, there is a unique intersection between u(b) and $U_H(b)$ at $b = b^* > \underline{b}$. Notice that u(b) may not intersect $U_H(b)$ from below even if the former is steeper than the latter for all b. This no-intersection may occur if \overline{b} is not very large. In this case, set $b^* = \overline{b}$, that is, in this knife-edge situation, we have all enterprises choosing N. Thus, to have $b^* < \overline{b}$, choose \overline{b} finite but sufficiently large.

Following the same argument as above, we can establish that a sufficient condition for u(b) intersecting $U_L(b)$ from above only once at $b = b^*$ is

$$\xi(a, b) < \min\{l_N(z(a, b), u), l_I(z(a, b))\}$$

which guarantees that u(b) is flatter than $U_L(b)$ for all $b \in [\underline{b}, \overline{b}]$. Again, for no-intersection, set $b^* = \overline{b}$. Also, note that the unique b^* at which u(b) intersects $U_H(b)$ is not necessarily the same as b^* at which u(b) intersects $U_L(b)$.

4. Finally, consider the equilibrium ownership structure, $\langle I, N \rangle$. As in the $\langle I \rangle$ equilibrium, either $u_0 < u_L(\underline{z})$ or $u_0 > u_H(\underline{z})$ guarantees that the lowest-productivity firm chooses to integrate. There are two possibilities. First, the $\langle I, N \rangle$ equilibrium emerges when $u_0 < u_L(\underline{z})$, u(b) intersects $U_L(b)$ from below only once, and u(b) does not intersect $U_H(b)$ (as depicted in Figure 4d). Thus, $u'_l(b) > U'_L(b)$, $u'_N(b) > U'_L(b)$ and $u'_N(b) < U'_H(b)$ together imply that u(b) intersects $U_L(b)$ from below at a unique $b = b^{**}$. Note that $\xi(a, b) > l_I(z(a, b))$ implies that $u'_l(b) > U'_L(b)$ for all b, $\xi(a, b) > l_N(z(a, b), u)$ implies that $u'_N(b) > U'_L(b)$ for all b, and $\xi(a, b) < h_N(z(a, b), u)$ guarantees that $u'_N(b) < U'_H(b)$ for all b. Therefore,

$$\max\{l_N(z(a, b), u), l_I(z(a, b))\} < \xi(a, b) < h_N(z(a, b), u)$$

imply that u(b) intersects $U_L(b)$ from below at a unique $b = b^{**} < \underline{b}$. As in the previous case, if u(b) does not intersect $U_L(b)$, then set $b^{**} = \overline{b}$.

The second possibility is that $u_0 > u_H(\underline{z})$, u(b) intersects $U_H(b)$ from above only once, and u(b) never intersects $U_L(b)$. Following a similar argument to the one above, we establish that

$$l_N(z(a, b), u) < \xi(a, b) < \min\{h_N(z(a, b), u), h_I(z(a, b))\}$$

is a sufficient condition for a single-crossing between u(b) and $U_H(b)$ at a unique $b = b^{**} \le \overline{b}$. As before, note that u(b) does not necessarily intersect $U_L(b)$ and $U_H(b)$ at the same b.