Global Supply Chains without Principals*

Mengxiao Liu†

Abstract

I construct a novel database of 122,765 supplier-producer relationships and examine the extent to which these relationships feature backward integration, forward integration, and outsourcing. In this database, backward and forward integration coexist in 62% of all upstream-downstream industry pairs with integrated relationships. This coexistence has been ruled out both empirically and theoretically in previous literature. I show that ruling out coexistence results in estimation bias. To explain coexistence, I develop a property rights model without principal firms where each supplier-producer pair chooses from backward integration, forward integration, and outsourcing. I find that firm integration decisions, especially the cross-border ones, are strongly sensitive to contractual frictions as predicted by Property Rights Theory. By allowing coexistence, my model performs much better than all existing models with principals.

*I thank Daniel Trefler, Peter Morrow, Kunal Dasgupta, Davin Chor, Devashish Mitra, Mary Lovely, Joane Oxley, Ig Horstmann, Bernardo Blum, Heski Bar-Issac, and Nicolas Li for their help and suggestions. I appreciate the comments from seminar and conference participants at the University of Toronto, Syracuse University, Colgate University, U Arkansas at Fayetteville, UC Merced, U Florida, CEA, Midwest, and the International Atlantic Economic Society. I am grateful to the staff at Rotman Finance Lab and S&P Capital IQ for their help. All faults are my own.

†Department of Economics, Maxwell School of Citizenship and Public Affairs, Syracuse University, NY, 13244, USA. mliu34@syr.edu

1 Introduction

In 1987, Acer acquired Counterpoint Computers, a small Silicon Valley maker of multiuser systems. This acquisition marks the first step of Acer's effort towards transforming itself, through forward integration, from a Taiwanese OEM manufacturer to the sixth largest computer vendor of the world (Dobson and Yue, 1999). In the same year, PC's Limited, a dorm room headquartered company that builds IBM PC-compatible computers from stock components, changed its name to Dell, and started expanding both upstream and downstream. After 5 years, Dell made it to Fortune 500.

Backward and forward integration are powerful forces shaping multinational firms, yet many firm boundary studies make an unrealistic assumption about integration decisions. Specifically, they assume that integration is one-directional, i.e., it goes backward *or* forward, but not in both directions.[1] This assumption creates an identification problem: if backward and forward integration coexist and are driven by different mechanisms, any empirical exercise based on the one-directional assumption would be biased. For example, Property Rights Theory (Grossman and Hart, 1986) predicts that a supplier-producer pair chooses backward integration when the producer's non-contractible, relationship-specific investment is relatively important; it chooses forward integration when the supplier's non-contractible, relationship-specific investment is relatively important. An integrated supplier-producer pair can be explained by two opposite mechanisms. Without identifying the direction of integration, Property Rights Theory is not testable (Whinston, 2001). In other words, Property Rights Theory is unfalsifiable because one can support any significant relationship between integration decision and the relative importance of the supplier/producer's non-contractible, relationship-specific investment.[2]

[1] In the voluminous international trade literature inspired by Antràs (2003) and in the smaller literature testing integration theories using large-scale databases (e.g., Acemoglu et al., 2009), a principal firm decides whether or not to integrate an agent firm and, since the agent cannot integrate the principal, all integration is one-directional i.e., the coexistence of backward and forward integration is assumed away.

[2] It should be noted that this identification problem is not specific to Property Rights Theory. It applies to all firm integration theories where the direction of integration matters. Gibbons (2005) describes four theories of firm integration decisions: the rent-seeking theory, the property-rights theory, the incentive systems theory, and the adaptation theory. Except for the rent-seeking theory (more frequently referred to as the transaction cost theory), the direction of integration matters in all other theories. In this paper I focus my discussion on the Property Rights Theory because it is a predominant framework for studying firm integration decisions.

To see whether integration is one-directional, I compile a global database of seller-buyer relationships that allows one to distinguish between backward and forward integration for the first time.[3] For each seller-buyer relationship, I observe the parent information of both the seller firm and the buyer firm.[4] Using their parent information, I classify all seller-buyer relationships into three categories: seller integration of the buyer (forward integration), buyer integration of the seller (backward integration), and non-integration (outsourced). Note that each seller-buyer relationship features an industry pair. I calculate for each industry pair the fraction of buyer integration (of the seller) relationships relative to all integrated relationships. Figure 1 provides the unweighted and weighted histograms of all 4,713 industry pairs.[5] Panel (a) shows that most (74%) industry pairs seem to feature either seller *or* buyer integration. The one-directional assumption does not seem too aberrant. However, if each industry pair is weighted by its total number of seller-buyer relationships, 62% of the industry pairs feature the coexistence of backward and forward integration. Note that this has not been verifiable in previous databases because they do not contain ownership information, and thus cannot tell backward integration from forward integration.

[Figure 1 about here.]

I then search for a theory of the firm that allows for the coexistence of backward and forward integration. Interestingly, most studies on firm integration decisions assume that the integration decision is made by a principal firm towards an agent firm. Since the agent cannot integrate the principal, integration becomes one-directional.[6] However, the identity of a principal firm is not empirically defined, and has largely been customary.[7] If an industry pair features only one

[3]The database is compiled from S&P Capital IQ (now S&P Market Intelligence). The definition of a seller-buyer relationship is that the seller sells its product or services to the buyer. The seller can be considered as the upstream firm, and the buyer can be considered as the downstream firm.

[4]A parent company is a company that owns the majority ($\geq 50\%$) stakes of another company.

[5]There are 156 industries and 18,575 industry pairs, of which 4,713 contain integrated relationships. The industries here correspond to the 156 sub-industries in the GICS (Global Industry Classification Standard) developed by S&P.

[6]An exception is made in Acemoglu et al. (2010) where the authors develop a theoretical model that allows integration to go either forward or backward, but later have to assume away forward integration because their database does not contain information on the direction of integration.

[7]When Acer integrates it downstream partner Counterpoint Computers, Acer seems to be the principal that makes a forward integration decision. When Dell integrates Exanet, its upstream OEM NAS (Network-attached

direction of integration, the principal assumption does not lead to estimation bias. However, if an industry pair features the coexistence of backward and forward integration, as those illustrated by Figure 1, the principal assumption will lead to estimation bias.

To explain coexistence I construct a property rights model *without* the principal assumption. The key difference between my model and a model with principals lies in the determination of organizational form. In my model, both firms (buyer and seller) in a relationship bargain over three organizational forms: buyer integration (buyer owns seller), seller integration (seller owns buyer), and outsourcing (neither firm owns the other). This model retains the insight of the original property rights model in Grossman and Hart (1986) and does not require a principal assumption. I also extend this model to allow for heterogeneous productivity across different seller-buyer pairs, and derive testable predictions for both the homogeneous and the heterogeneous models. Both models find positive support in my database.

[Table 1 about here.]

Table 1 summarizes the key empirical results for the homogeneous model. I use the buyer firm's relative R&D intensity to proxy for the relative importance of the buyer firm's non-contractible, relationship-specific investment.[8] As Table 1 shows, a standard deviation increase in the buyer firm's relative R&D intensity is associated with a 0.193 increase in a firm pair's likelihood of choosing buyer integration and a 0.079 decrease in its likelihood of choosing seller integration. These results support the predictions of my model. In addition, I find that a standard deviation increase in the buyer firm's relative R&D intensity is associated with a 0.05 increase in the firm pair's likelihood of choosing integration. This result supports a property rights model with the buyer firm as the principal. If the coefficient were negative and significant,

Storage) software provider, Dell seems to be the principal that makes a backward integration decision. When Apple sources iPhone OLEDs from Samsung and LG, who is the principal? The literature norm has been to assume the principal status based on industry. The global sourcing literature assumes the downstream firm to be the principal, while the IO literature assumes the upstream firm to be the principal. Lafontaine and Slade (2007) summarize the assumptions and findings of various industry-level studies of firms' vertical integration decisions.

[8]The construction of this measure is elaborated in Section 2.3. R&D expenditure is a standard proxy for a firm's non-contractible, relationship-specific investment in the property rights literature. See for example Nunn (2007), Nunn and Trefler (2008, 2013), and Acemoglu et al. (2010). I use multiple ways to construct relative R&D intensity.

it would support a property rights model with the seller as the principal. This result verifies the conjecture of Whinston (2001) – a property rights model with the principal assumption is nonfalsifiable, because by switching the principal assumption, one can support any significant empirical relationship.

This paper contributes to the extensive literature on the global sourcing decisions of multinational firms.[9] The model in this paper follows the line of work by Antràs (2003) and Antràs and Helpman (2004, 2008). I show in Section 3.4 that by adding a step where the seller and buyer determine who becomes the principal, the model in Antràs and Helpman (2004) delivers the same predictions as my model. Existing literature assumes that the integration decisions are made by the headquarter company that is located in the downstream of a global supply chain. The empirical finding in this paper suggests that the possibility of an upstream headquarter company should also be considered. Another new insight from this paper is that international transaction relationships are less likely to be integrated, and are more sensitive to contractual frictions, as shown in Section 4.1.

This paper also contributes to the empirical literature examining firm integration decisions.[10] Of all the theories of firm integration decisions, Transaction Cost Theory has received the most empirical examinations. On the other hand, Property Rights Theory has received limited empirical examination despite its theoretical popularity. This paper provides the first cross-country, cross-industry support for Property Rights Theory. The results in this paper suggest that the reason for the lack of empirical evidence may be twofold. First, most empirical databases do not allow one to identify the direction of integration and are thus not suitable

[9]For example, Nunn (2007) and Nunn and Trefler (2008, 2013) investigate the role of contracting institutions in firms' location decisions. Antràs and Chor (2013), Alfaro and Chen (2014), and Alfaro et al. (2015) study the distribution of ownership structure along the global value chains. See Nunn and Trefler (2014), Helpman (2011), Antràs (2013, 2015), and Antràs and Yeaple (2014) for surveys of this literature.

[10]For examinations at the global level, see the recent papers by Alfaro and Charlton (2009), Alfaro and Chen (2014) and Alfaro et al. (2015) using the Dun&Bradstreet WorldBase data, Altomonte et al. (2012), Altomonte and Rungi (2013) and Prete and Rungi (2015) using the Orbis data. For country-level studies, see papers by Acemoglu et al. (2010) using UK plant-level data, Eaton et al. (2011) and Corcos et al. (2013) using French firm-level data, Tomiura (2009) using Japanese firm-level data, and Kohler and Smolka (2011, 2014) using Spanish firm-level data. In addition to firm-level databases, there are also transaction-level databases such as the U.S. Related Party Trade Data (Nunn and Trefler, 2008), the Chinese Customs Records data (Feenstra and Hanson, 2005; Fernandes and Tang, 2012), and the more recent U.S. shipment-level data used by Atalay et al. (2014).

for testing Property Rights Theory; second, the one-directional assumption on integration decisions leads to underestimation when industry pairs feature bilateral integration (Whinston, 2001; Williamson, 2002; Segal and Whinston, 2012).

The rest of this paper is organized as follows. Section 2 introduces the structure of the database, provides descriptive statistics, and describes the construction of key variables used in the empirical exercises. Section 3 builds a homogeneous model and a heterogeneous model, and derives testable predictions from both models. Section 4 reports empirical results for the homogeneous model with a series of robustness checks. Section 5 presents the empirical results for the heterogeneous model. Section 6 concludes the paper.

2 Data and Descriptive Statistics

2.1 Data

The data used in this paper are compiled from S&P Capital IQ over the years 2013, 2014, and 2016.[11] It has two components: the firm database and the relationship database.

The firm database contains industry, country, and financial information of 3,278,425 firms from 216 countries over the period 2002-2015. The generic industry classification system in this database is the Global Industry Classification Standard (GICS). It is developed in 1999 by MSCI Inc. (formerly known as Morgan Stanley Capital International) and Standard and Poor's. GICS contains 10 sectors, 24 industry groups, 67 industries, and 156 sub-industries[12]. In my empirical exercise, unless otherwise specified, I always use the sub-industry information. I also collect firms' balance sheet data including R&D expenditure, goodwill, total revenue, employment, total assets, etc. The financial information is missing for many companies. In my regression sample, I only use the financial information of 39,420 firms.

[11]S&P Capital IQ, now called S&P Global Market Intelligence, is a multinational financial information provider owned by S&P Global Inc.

[12]Capital IQ provides concordance tables between GICS and SIC, and GICS and NAICS. The concordance table maps GICS sub-industries with NAICS2-NAICS6 (one to many) and SIC4 (one to many). See the link below for the complete classification of GICS: `https://en.wikipedia.org/wiki/Global_Industry_Classification_Standard`.

The relationship database contains rich between-firm relationships[13]. In this paper I primarily use parent, direct subsidiary, customer, and supplier relationships. A parent company is a firm that owns majority (>=50%) stakes in another company. A customer is a company that receives products or services and that gives business. Direct subsidiary and supplier relationships are complementary to parent and customer relationships. I code all parent and direct subsidiary relationships into one firm-parent relationships database, and all customer and supplier relationships into one seller-buyer relationships database. Over three years of data collection, I have a total of 611,335 firm-parent relationships and 779,436 seller-buyer relationships. To better identify *integrated* relationships, I also collect information on firms' ultimate parent, holding company, merged entity, limited partner, investor and pending parent/investor relationships. The usage of these relationships will be elaborated later. Table 2 documents the number of various relationships in my database.

[Table 2 about here.]

To identify the ownership structures, I combine the seller-buyer relationships database with the firm-parent relationships database by matching firm names[14]. For each seller-buyer relationship, if the seller is the buyer's parent, I classify it as a seller integration (of the buyer) relationship. If the buyer is the seller's parent, I classify it as a buyer integration (of the seller) relationship[15]. A non-integrated relationship refers to a pair of buyer and seller that is not observed in any of the following relationships: parent, ultimate parent, holding company, merged entity, limited partner, investor, and pending parent/investor.

Recall that my data is collected over three years: 2013, 2014, and 2016. The above matching

[13]Capital IQ collects firms' business relationships from various sources including regulatory filings (SEC for American companies, SEDAR for Canadian companies, ASX for Australian companies, NZX for New Zealand companies), news aggregators (e.g., news articles, press release, corporate announcements, bankruptcy reports from newswires and papers etc.), and direct surveys with companies.

[14]This is done in a two-step procedure. I first match all firm names with an ID number in the Capital IQ system so all firms in the seller-buyer database and firm-parent database have a unique ID number. I then match each buyer firm and seller firm with their parent information using their ID. The first step of this procedure involves matching firm names in the relationships database with the firm names in the Capital IQ database. Since fuzzy string match does not produce ideal results, I first clean up firm names, then perform perfect string match.

[15]I drop a seller-buyer relationship when the two firms share a common parent because such relationships are not pertinent to my model.

process is performed for each year separately. The matching process thus generates three ownership samples. I then match each firm in each ownership sample with its industry, country, and financial information from the firm database. After deleting pairs in which either firm is missing important financial information, I end up with 82,893 seller-buyer relationships for 2013, 107,924 seller-buyer relationships for 2014, and 95,346 seller-buyer relationships for 2016. For my benchmark regressions, I use only relationships from 2016.[16]

Capital IQ's method for collecting customer and supplier relationships suggests that it is more likely to capture the relationships between large companies, especially those outside the firm boundary[17]. When a parent company buys from or sells to its subsidiary, it is less likely to be documented in any of the sources. This bias explains why in the 2016 ownership sample, only 1.37% of the 95,346 seller-buyer relationships are integrated. To correct for this bias, I impute 27,400 integrated seller-buyer relationships from the firm-parent relationships database. This imputation is performed using a now standard technique involving input-output tables, as in Acemoglu et al. (2009), Acemoglu et al. (2010), Antràs and Chor (2013), and Alfaro et al. (2015). Specifically, let $i(p)$ and $i(a)$ be the industries in which the parent and affiliate operate, respectively. Using the 2002 U.S. input output table, let $b_{i,i'}$ be industry i's share of intermediate inputs sourced from sector i'. If $b_{i(p),i(a)} > 0$, then the parent's industry buys from the affiliate's industry and I create a seller-buyer observation in which the parent is the buyer. If $b_{i(a),i(p)} > 0$, then the affiliate's industry buys from the parent's industry and I create a seller-buyer observation in which the affiliate is the buyer. If $b_{i(p),i(a)} = b_{i(a),i(p)} = 0$, then the parent-affiliate relationship is not vertical and thus not included in the data. In the literature just cited, 100% the buyer-supplier information is imputed in this way. I use this imputation for just (27400/122765=) 22.3% of the relationships in my database. Throwing away these imputed relationships does not change the qualitative results of this paper. For each benchmark regression using the extend sample (the sample with imputed relationships), I report in the Appendix the results from the same regression using the original sample (the sample without imputed relationships).

[16]All ownership samples, separately or combined, deliver similar results. I use the 2016 sample only to get a cleaner identification.

[17]For example, the SEC segment reporting (form 8-K) rules state "a company generally must report separately

Table 3 tabulates the three types of ownership structures in the 2016 ownership sample. Integrated relationships now constitute 23.4% (instead of 1.37%) of all seller-buyer relationships.

[Table 3 about here.]

2.2 Pros and Cons of this Database

The advance of numerous large-scale databases provides unprecedented opportunities for firm boundary studies. In this section I compare my database with some popular databases used in the literature and highlight the pros and cons of each type of these databases.

The databases used in firm boundary studies can be divided into two groups: firm/plant-level databases and transaction-level databases. The former includes the Bureau of Economic Analysis (BEA) annual survey of U.S. Direct Investment Abroad (Antràs et al., 2008), the Dun & Bradstreet WorldBase (Acemoglu et al. 2009, Alfaro and Charlton 2009, Alfaro and Chen 2014, and Alfaro et al. 2015), the UK Annual Respondenta Database (Acemoglu et al., 2010), and French firm-level data (Corcos et al. 2013, Defever and Toubal 2013). The latter includes the U.S. imports and exports database (Bernard et al. 2009 and Nunn and Trefler 2013), the import and export data from the Customs General Administration of China (Feenstra and Hanson 2005 and Li 2013) and more recently, the U.S. Commodity Flow Survey used by Atalay et al. (2014).

Firm/plant-level databases have detailed information on the firm, but not on its partners. For example, the BEA database contains much information on the multinational firm and its subsidiaries, but not on this multinational firm's outsourced partners. The content of the Dun & Bradstreet database varies across papers, but it mostly contains a firm's balance sheet data, the industries that a firm operates in, but does not have information on the firm's partners, i.e., customers, suppliers. The UK Census of Production data has information on the firm and

information about an operating segment that...its reported revenue, including both sales to external customers and intersegment sales and transfers, is 10 percent or more of the combined revenue of all reported operating segments, whether generated inside or outside the company. This statement suggests that large subsidiaries or major customers are likely to be reported in a company's regulatory filing. Newswires and papers on the other hand are more likely to report the signing of major deals between large companies, especially those that are not related.

its plants, but does not list the firm's outsourced partners either. However, the BEA database contains the transaction values between the multinational firm and its subsidiaries, and the UK Annual Respondents Database has information on the input costs and output levels of each UK plant. Such information is not available in my data.

Transaction/shipment databases usually contain detailed information on one side of each transaction/shipment, but not on the other side. The U.S./China Customs data contains information on the U.S./Chinese firm, but not on the other side of the transaction/shipment. The U.S. Commodity Flow Survey identifies the shipping company, but not the receiving company. These databases, however, contain information on the characteristics of the transaction/shipment, such as the price, quantity, and characteristics of the good. Such information is not available in my data.

Last but not least, some of the above mentioned databases are comprehensive in the sense that they cover the universe of all firms in a certain industry (BEA and UK Annual Response Database), all transactions/shipments to and from a country within a certain time period (U.S./China Customs data, U.S. Commodity Flow Survey). My database is a sample of the important relationships between large companies.

Compared to the previously mentioned databases, the most distinct feature of my database is that it combines features from both types of databases. Firm/plant-level databases contain information on integrated relationships (directly or indirectly), but not on outsourced ones. Transaction/shipment databases contain a large number of relationships (and sometimes integrated relationships as well, e.g., U.S. Commodity Flow Survey), but in most cases do not allow one to identify the ownership structure of a relationship. Both types of databases contain detailed information on one side of the relationship, but not on the other side. My database combines the strengths of both types of databases by joining a relationship database with a firm database. This is the first database that contains symmetric information on both sides of a seller-buyer relationship. Such feature is especially important for testing a PRT model because the center of PRT is that firms allocate property rights based on the *relative* importance of each firm's non-contractible, relationship-specific investment. Without enough information on both parties involved, researchers have to compromise by making assumptions either on integration decisions, i.e., the principal assumption, or on the empirical proxies for relationship-specificity. Such assumptions, as I will elaborate in Section 4, prevent one from properly testing the PRT. These reasons explain the scarcity of empirical work testing PRT relative to TCE (the Transaction Cost approach to firm boundary decisions as proposed by Coase (1937)).

2.3 Proxies for Non-contractible, Relationship-Specific Investments

The key variable in the PRT model is the relative importance of each firm's non-contractible, relationship-specific investment. I follow similar methodology to Antràs (2003), Nunn and Trefler (2008), and Alfaro et al. (2015) in using R&D expenditure as the main proxy for a firm's non-contractible, relationship-specific investment. I use goodwill as an alternative to R&D expenditure[18].

The R&D intensity of buyer firm b, rd_b, is calculated as

$$rd_b = \frac{\text{Buyer's average R\&D expenditure}}{\text{Buyer's average total revenue}},$$

where the buyer's average R&D expenditure and total revenue are averaged over the period 2007-2014.[19] The R&D intensity of seller s, rd_s, is calculated in the same way. In addition to total revenue, I also use employment and total asset as the denominators. They all deliver similar results.

Since the buyer and seller's organizational choice depends on the *relative* importance of their respective non-contractible, relationship-specific investment, I construct the following pair-level variable to proxy for the relative importance of the buyer firm's non-contractible, relationship-

[18]The goodwill measure I use is obtained from a firm's balance sheet. It is the sum of a firm's goodwill, intangible assets, and other intangibles. This combined term includes a firm's gross and accumulated amortization of goodwill, intangible assets, intellectual properties, purchased intangible assets excluding goodwill, intellectual property, trademarks and trade names, non-competition agreements, etc. Many of these items are non-contractible and relationship specific.

[19]A firm may have missing entries for its R&D expenditures. If a firm has missing R&D expenditures for the whole period of 2007-2014, its average R&D expenditure is treated as missing; if a firm's R&D expenditure is missing for some years and positive for the others, then the missing years are treated as 0.

specific investment in a seller-buyer pair i:

$$rd_i \equiv \frac{rd_b}{rd_b + rd_s}. \tag{1}$$

By construction, $(1 - rd_i)$ is the relative importance of the seller firm's non-contractible, relationship-specific investment[20]. I use rd_i only in my benchmark regressions because rd_i is the sufficient statistic for pair i. The relative importance of the buyer firm's goodwill intensity, gdw_i, is constructed in a similar manner.

Note that in other databases, since the authors do not observe both rd_b and rd_s, they have to either construct rd_b and rd_s at the industry level, or assume that one side does not have to make any non-contractible, relationship-specific investment. My database is the first to construct a *pair-level* R&D intensity measure.

2.4 Descriptive Statistics

As mentioned in Section 2, I use the 2016 ownership sample for my benchmark regressions. After deleting all seller-buyer relationships with missing financial information on either side of the relationship, I end up with 122,765 seller-buyer relationships between 39,420 firms. These 39,420 firms come from 138 countries and 156 industries. Table 4 shows the distribution of these firms across five geographic regions. 49% of these firms are located in Europe.

[Table 4 about here.]

Firms in this database are classified using GICS (the Global Industry Classification System), which contains 10 sectors[21]. The top-5 sector pairs that host the most seller-buyer relationships are: (Information Technology, Information Technology), (Industrials, Industrials), (Consumer Discretionary, Consumer Discretionary), (Information Technology, Consumer Discretionary), and (Information Technology, Industrials)[22]. The first item in each bracket refers to the seller

[20]The construction of rd_i automatically drops out two types of seller-buyer relationships: (1) $rd_b = rd_s = 0$; (2) rd_b or rd_s is missing.

firm's sector. See Table E.1 in for the list of sector-pairs that host more than 1000 seller-buyer relationships.

[Table 5 about here.]

Table 5 provides summary statistics for the seller firm, the buyer firm and an average firm (irrespective of whether it is a seller or a buyer) across all three types of ownership structures: buyer integration, seller integration and non-integration/outsourced relationships. In seller integration relationships, buyer firms are younger, smaller, spend less on R&D and goodwill, and outnumber seller firms by two thirds. The opposite is true in buyer integration relationships: buyer firms are older, bigger, spend more on R&D and goodwill, and are outnumbered by seller firms by one-half. In outsourced relationships, the difference between seller firms and buyer firms are much smaller, although seller firms are slightly bigger and spend relatively more on R&D and goodwill. The extended sample (the sample with imputed relationships) display similar patterns.

3 Model

In this section, I construct two property rights models: a homogeneous model without productivity measures, and a heterogeneous model with heterogeneous pair-level productivity.

The models in this paper are based on Antràs (2003) and Antràs and Helpman (2004), henceforth the AH models, with two important differences. First, in my model, a buyer firm and a seller firm collectively choose from buyer integration, seller integration, and non-integration. In the AH models (and many other PRT models), the integration decision is made by a principal firm. The essential difference here is not the existence of a principal firm, but the choice set. As I will show later in this section, by adding a step where the buyer and seller first determine who becomes the principal, an AH model delivers identical predictions as my model.

[21] The 10 sectors are: Consumer Discretionary, Consumer Staples, Energy, Financials, Healthcare, Industrials, Information Technology, Materials, Telecommunication Services, and Utilities.

[22] Consumer Discretionary contains 5 industry groups: Automobiles and Components, Consumer Durables and Apparel, Consumer Services, Media, and Retailing.

Second, in the heterogeneous model in Antràs and Helpman (2004), firms make heterogeneous organizational decision because of the assumption that the fixed cost of production is higher under vertical integration than outsourcing. I assume that firm pairs draw their fixed costs of production from an i.i.d. distribution. Heterogeneous organizational decision no long depends on an ad hoc assumption on the ranking of fixed costs, but is purely driven by property rights factors.

3.1 Setup

Consider an economy that is resided by a unit measure of consumers with CES preference

$$U = q_0 + \frac{1}{\omega}\int_{j=1}^{J} Q_j^{\omega} dj, \quad 0 < \omega < 1,$$

where $1/(1-\omega)$ is the representative consumer's cross-industry elasticity of substitution, q_0 is the consumption of a homogeneous good, and Q_j is a CES aggregate of the consumption of all varieties in industry j:

$$Q_j = \left(\int_{i=1}^{N_j} q_{ji}^{\alpha} di\right)^{1/\alpha}, \quad 0 < \alpha < 1.$$

N_j is the number of varieties in industry j, which is endogenous in a general equilibrium model but taken as exogenous in a partial equilibrium model. q_{ji} is the consumption level of variety i in industry j, and $\sigma \equiv 1/(1-\alpha) > 1$ is this consumer's within-industry elasticity of substitution. σ is constant across all industries.

The inverse demand function for variety i in industry j is

$$p_{ji} = Q_j^{\omega-\alpha} q_{ji}^{\alpha-1}.$$

Producing q_{ji} requires inputs from a producer (henceforth the buyer) b and a supplier (henceforth the seller) s. b and s are identities, not indexes. The buyer is located in the final good industry j and the seller is located in a related industry which does not produce any final good[23]. Let $x_{ji,b}$ and $x_{ji,s}$ denote the buyer and seller's investment levels. For simplicity, I assume that $x_{ji,b}$ and $x_{ji,s}$ are non-contractible and relationship-specific, which implies that the two firms cannot write a contract specifying their investment levels, and that $x_{ji,b}$ and $x_{ji,s}$ have zero values outside relationship i. The final good is produced by the following Cobb-Douglas production function

$$q_{ji} = \left(\frac{x_{ji,b}}{\eta_{ji}}\right)^{\eta_{ji}} \left(\frac{x_{ji,s}}{1-\eta_{ji}}\right)^{1-\eta_{ji}}, \quad \eta_{ji} \in (0,1). \tag{2}$$

η_{ji} and $(1-\eta_{ji})$ represent the relative importance of the buyer and the seller's non-contractible, relationship-specific investments in pair i, industry j. Let $c_{ji,b}$ and $c_{ji,s}$ denote the buyer and the seller's marginal costs of investment. Both firms take their marginal costs of investment as given.

In an industry equilibrium, the industry subscript j can be taken out of the production function:

$$q_i = \left(\frac{x_{i,b}}{\eta_i}\right)^{\eta_i} \left(\frac{x_{i,s}}{1-\eta_i}\right)^{1-\eta_i}, \quad \eta_i \in (0,1). \tag{3}$$

Since $x_{i,b}$ and $x_{i,s}$ are non-contractible, the two firms cannot write a contract specifying their investment levels[24]. Although firms cannot contract on their investment levels, they can write a contract on the ownership of these investments. More specificaly, they can choose from three alterantives: buyer integration (of the seller), seller integration (of the buyer), and non-integration[25]. Under buyer integration, the buyer owns both firms' investments; under seller integration, the seller owns both firms' investments; under non-intergation, the two firms own their respective investments. I follow Grossman and Hart (1986) in defining ownership as a collection of residual control rights - rights that are either too costly or impossible to write into a contract. Ownership comes into play during the renegotiation stage when the two firms bargain over the division of their total revenue. Under buyer integration, a negotiation breakdown means that the buyer gets to seize both $x_{i,b}$ and $x_{i,s}$, leaving the seller with nothing.

[23]That the seller's industry does not produce any final good is a simplifying assumption. Alternatively, one may relax this assumption allow allow sellers to produce final goods. In this case, a close resemblance to the seller-buyer relationships in this model is an input-output table, or the network model constructed by Acemoglu et al. (2012).

However, the buyer cannot use $x_{i,s}$ as efficiently as the seller, and can only produce an output of $\delta_i q_i$, where $0 < \delta_i < 1$ and q_i is defined in equation (5). Under seller integration, the seller owns both $x_{i,b}$ and $x_{i.s}$. If the negotiation breaks down, the seller produces $\delta_i q$, leaving the buyer with nothing. Under non-integration, if the negotiation breaks down, the two firms take away their respective investments. The Cobb-Douglas production function implies that they both end up with 0 output. It is now obvious that ownership structure affects the two firms' bargaining powers and thus their ex-post revenue shares. And their ex-post revenue shares will affect their ex-ante investment incentives.

The production in industry j (ignoring the industry subscript) goes as follows:

1. Buyer firms and seller firms form one-to-one matches.

2. Each firm pair draws a set of fixed costs $\{f_i^{BI}, f_i^{SI}, f_i^{NI}\}$ from an i.i.d. distribution

$$f_i^k = f^k + \varepsilon_i^k, \quad k \in \{BI, SI, NI\},$$

   where f_i^k is a fixed cost that is specific to organizational form k. BI, SI, and NI stand for buyer integration of the seller, seller integration of the buyer, and non-integration, respectively. There is no assumption on the ranking of f^{BI}, f^{SI}, and f^{NI}. ε_i^k is a random variable that follows a zero-mean Gumbel distribution that is independent of i and k.

[24] In other words, any contract specifying investment levels is not renegotiation-proof. Following the work of Antràs and Helpman (2008), this assumption can be readily extended to allow for partial-contractibility, where the two firms can write a contract on their contractible investments, and simultaneously choose their non-contractible investment levels based on their respective incentives.

[25] Without loss of generality, I assume that ownership is *indivisible*. Partial-ownership is easily achieve by allowing ownership to be any value between 0 and 1. Although in the current setup it is difficult to distinguish between partial-ownership and *cross*-partial ownership.

[26] One can think of δ as a measure of the court's enforcement power. The higher δ is, the more investment the owner can retrieve, and the higher outside option the ownership delivers.

[27] There is one additional ownership beside buyer integration, seller integration and non-integration, which is cross ownership, where the buyer owns the seller's investment and the seller owns the buyer's investment. This type of ownership structure does not seem to exist in reality, and it generates identical results as non-integration, and is thus left out of the discussion. The reason this generates identical results as the non-integration case is because property rights are assumed to be indivisible. When *partial* ownership over a firm is allowed, cross ownership, where the two firms are allowed to control only part of the other firm's asset can allow the firms to achieve the first-best joint surplus. This is because under cross-ownership, the two firms can perfectly control each-other's ex post share of total revenue, and hence their ex ante investment incentives. Allowing for the existence of *cross partial* ownership may be an interesting extension, so far there is no work modeling partial ownership in a property rights framework.

3. After drawing the fixed costs, the buyer and seller in each firm pair collectively determine their ownership structure via Nash bargaining. The bargaining weights are γ for the buyer and $(1-\gamma)$ for the seller. $0 < \gamma < 1$. At this step, both firms' outside options are 0. The bargaining results in the two firms choosing an ownership structure $k_i \in \{BI, SI, NI\}$, and a transfer payment $t_i \in \mathbf{R}$ from the buyer to the seller.

4. After k_i is chosen and t_i is paid, the buyer and seller simultaneously choose their investment levels $x_{i,b}$ and $x_{i,s}$. Their marginal costs are c_b and c_s.

5. After investments are made, the two firms bargain over the division of their future revenue. Suppose the bargaining weights are β for the buyer and $(1-\beta)$ for the seller. $0 < \beta < 1$.

6. The final good is produced and sold on the market. The two firms split the total revenue according to their bargaining agreement in step 5.

3.2 Industry Equilibrium of the Homogeneous Model

In the homogeneous model, all firms face the same $(\beta, \delta, \eta, c_b, c_s)$. Heoteogenerity in organizational decisions is driven solely by ε_i^k whose distribution is not firm-specific. I will take out the firm index i for now and introduce it back in when f_i^k comes up.

Assume that $Q_j = 1$, the inverse demand function facing firm pair i (ignoring the firm pair index i) is

$$p = q^{\alpha-1}. \tag{4}$$

The production function (ignoring the pair index i) is

$$q = \left(\frac{x_b}{\eta}\right)^{\eta} \left(\frac{x_s}{1-\eta}\right)^{1-\eta}. \tag{5}$$

The above two equations give the following revenue function

$$R = pq = q^{\alpha} = \left(\frac{x_b}{\eta}\right)^{\alpha\eta} \left(\frac{x_s}{1-\eta}\right)^{\alpha(1-\eta)}. \tag{6}$$

I use backward induction to derive the industry equilibrium[28].

In step 6, the two firms simply produce the product, sell it, and divide the revenue. The output function and revenue function are as specified in equations (5) and (6).

In step 5, the two firms bargain over the division of revenue R, taking k, x_b and x_s as given. As discussed before, under buyer integration ($k = BI$), the buyer's outside option is his revenue from δq units of output, which is $\delta^\alpha R$. The seller's outside option is 0. The pair's Nash surplus is thus $R - \delta^\alpha R - 0 = (1 - \delta^\alpha)R$. The buyer's Nash surplus equals his outside option plus his share (β) of the Nash surplus, that is, $\delta^\alpha R + \beta(1 - \delta^\alpha)R = [\beta + (1 - \beta)\delta^\alpha]R$. The seller's Nash surplus is $0 + (1 - \beta)(1 - \delta^\alpha)R(i) = (1 - \beta)(1 - \delta^\alpha)R$. Therefore, under buyer integration, the buyer gets a share $[\beta + (1 - \beta)\delta^\alpha]$ of total revenue R, and the seller gets a share of $(1 - \beta)(1 - \delta^\alpha)$.

The two firms' revenue shares under seller integration ($k = SI$) and non-integration ($k = NI$) can be derived using the same logic. Let β^k represent the buyer's share of total revenue under ownership structure k. The seller's share of total revenue is then $(1 - \beta^k)$. Their revenue shares under different ownership structures are summarized in Table 6.

[Table 6 about here.]

It is clear from Table 6 that $\beta^{BI} > \beta^{NI} > \beta^{SI}$: the buyer has the highest revenue share under $k = BI$ and the lowest revenue share under $k = SI$. On contrary, $(1-\beta^{BI}) < (1-\beta^{NI}) < (1 - \beta^{SI})$ implies that the seller has the highest revenue share under $k = SI$ and the lowest revenue share under $k = BI$. This is driven by their rankings of outside options across the three ownership structures.

In step 4, the two firms choose their investment levels x_b and x_s to maximize their respective surpluses, taking β^k and $(1 - \beta^k)$ as given. The marginal costs of investments are c_b for the buyer and c_s for the seller. Both firms take their marginal costs as given.

[28] Backward induction has become a standard technique thanks to the groundwork laid by Antràs (2003) and Antràs and Helpman (2004, 2008).

The buyer's problem is

$$\max_{x_b} \beta^k \left(\frac{x_b}{\eta}\right)^{\alpha\eta} \left(\frac{x_s}{1-\eta}\right)^{\alpha(1-\eta)} - c_b x_b, \tag{7}$$

The seller's problem is

$$\max_{x_s}(1-\beta^k) \left(\frac{x_b}{\eta}\right)^{\alpha\eta} \left(\frac{x_s}{1-\eta}\right)^{\alpha(1-\eta)} - c_s x_s. \tag{8}$$

The equilibrium levels of x_b and x_s simultaneously solve the buyer's problem and the seller's problem. The buyer's investment level is

$$x_b = \left\{\alpha \left(\frac{\beta^k \eta}{c_b}\right)^{1-\alpha+\alpha\eta} \left[\frac{(1-\beta^k)(1-\eta)}{c_s}\right]^{\alpha-\alpha\eta}\right\}^{1/(1-\alpha)}, \tag{9}$$

and the seller's investment level is

$$x_s^{1-\alpha} = \left\{\alpha \left(\frac{\beta^k \eta}{c_b}\right)^{\alpha\eta} \left[\frac{(1-\beta^k)(1-\eta)}{c_s}\right]^{1-\alpha\eta}\right\}^{1/(1-\alpha)}. \tag{10}$$

Substituting the solutions to x_b and x_s back into the buyer and the seller's problems gives their respective surpluses:

$$\psi_b(\beta^k, \eta) = \frac{(1-\alpha\eta)\beta^k}{\left[(1/\alpha)\left(\frac{c_b}{\beta^k}\right)^{\eta}\left(\frac{c_s}{1-\beta^k}\right)^{1-\eta}\right]^{\frac{\alpha}{1-\alpha}}}, \tag{11}$$

and

$$\psi_s(\beta^k, \eta) = \frac{(1-\alpha+\alpha\eta)(1-\beta^k)}{\left[(1/\alpha)\left(\frac{c_b}{\beta^k}\right)^{\eta}\left(\frac{c_s}{1-\beta^k}\right)^{1-\eta}\right]^{\frac{\alpha}{1-\alpha}}}. \tag{12}$$

The pair's joint surplus is

$$\psi(\beta^k, \eta) \equiv \psi_b(\beta^k, \eta) + \psi_s(\beta^k, \eta) = \frac{1-\alpha[\beta^k\eta + (1-\beta^k)(1-\eta)]}{\{(1/\alpha)\left(\frac{c_b}{\beta^k}\right)^{\eta}\left(\frac{c_s}{1-\beta^k}\right)^{1-\eta}\}^{\frac{\alpha}{1-\alpha}}}. \tag{13}$$

Equations (11) (12) and (13) imply that the buyer and the seller's surpluses from step 4 can also be expressed as fractions of their joint surplus. Let $\mu(\beta^k, \eta)$ be the buyer's fraction and let

$1-\mu(\beta^k,\eta)$ be the seller's fraction:

$$\mu(\beta^k,\eta)=\frac{(1-\alpha\eta)\beta^k}{1-\alpha[\beta^k\eta+(1-\beta^k)(1-\eta)]};\quad 1-\mu(\beta^k,\eta)=\frac{(1-\beta^k)(1-\eta)}{1-\alpha[\beta^k\eta+(1-\beta^k)(1-\eta)]}. \quad (14)$$

Lemma 1. *The variable profit function $\psi(\beta,\eta)$ is supermodular in β and η for $\beta\in(0,1)$ and $\eta\in(0,1)$.*

Lemma 1 simply states that if the firm pair were allowed to choose β as a continuous variable, with a higher η, the buyer's non-contractible, relationship-specific investment is relatively more important, the pair finds it optimal to choose a larger β because it provides the buyer with a higher ex-post revenue share, hence increases the buyer's ex-ante investment incentive.

In fact, the optimal ex-post revenue share (if it were continuous) can be solved as a function of η,

$$\beta^*(\eta)=\frac{\eta(\alpha\eta+1-\alpha)-\sqrt{\eta(1-\eta)(1-\alpha\eta)(\alpha\eta+1-\alpha)}}{2\eta-1},$$

where $\beta^*(\eta)$ is strictly increasing in η.[29]

In this model, firms are not allowed to freely choose their revenue shares. Instead, they can only choose from three discrete values of β: β^{BI}, β^{SI} and β^{NI}. Given the ranking of these three values ($\beta^{BI}>\beta^{NI}>\beta^{SI}$) and the supermodularity of $\psi(\beta,\eta)$, β^{BI} is optimal when η is high, β^{SI} is optimal when η is low, and β^{NI} is optimal when η is at the intermediate level, as formally stated in Lemma 2.

Lemma 2. *There are two threshold values of η, $\underline{\eta}$ and $\bar{\eta}$, with $0<\underline{\eta}<\bar{\eta}<1$, such that the following statements hold.*

1. *When $\eta>\bar{\eta}$, $\psi(\beta^{BI},\eta)>\psi(\beta^{NI},\eta)>\psi(\beta^{SI},\eta)$.*
2. *When $\underline{\eta}<\eta<\bar{\eta}$, $\psi(\beta^{NI},\eta)>\max\{\psi(\beta^{BI},\eta),\psi(\beta^{SI},\eta)\}$.*
3. *When $\eta<\underline{\eta}$, $\psi(\beta^{SI},\eta)>\psi(\beta^{NI},\eta)>\psi(\beta^{BI},\eta)$.*

$\underline{\eta}$ and $\bar{\eta}$ are implicitly solved by $\psi(\beta^{SI},\underline{\eta})=\psi(\beta^{NI},\underline{\eta})$ and $\psi(\beta^{BI},\bar{\eta})=\psi(\beta^{NI},\bar{\eta})$.

[29] $\beta^*(\eta)$ is actually the solution to a model where partial ownership is allowed.

See Appendix A for proofs of Lemma 1 and Lemma 2.

Lemma 2 is a rendition of the key mechanism highlighted in the original PRT paper by Grossman and Hart (1986). To put it in words: the firm pair allocates property rights to the buyer when the buyer's non-contractible, relationship-specific investment is relatively important, and to the seller when the seller's non-contratible, relationshpi-specific investment is relatively important. In the intermediate case, the firm pair chooses non-integration. In the literature that follows Grossman and Hart (1986), researchers often adopt the assumption that the principal is the one making the integration versus outsourcing decision. In most cases, this is because data does not allow one to distinguish between the two types of integration. Lemma 2 demonstrates that such an assumption results in the conflation of two opposite cases (buyer and seller integration), and will result in estimation bias. I provide empirical evidence for this estimation bias in Section 4.2.

In step 3, the firm pair chooses an ownership structure k to maximize their joint profit. Denote this optimal choice by k_i^*. This problem can be written as

$$k_i^* = \underset{k \in \{BI, SI, NI\}}{\arg\max} \; \pi_i(\beta^k, \eta), \tag{15}$$

where $\pi_i(\beta^k, \eta)$ is the firm pair's joint profit,

$$\pi_i(\beta^k, \eta) \equiv \psi(\beta^k, \eta) - f_i^k, \tag{16}$$

where $f_i^k = f^k + \varepsilon_i^k$ is as defined in step 2.

Without fixed cost f_i^k, the ranking of the profit functions is identical to the ranking of the variable profit functions as in Lemma 2. Keeping η constant, all firms choose the same k. There is no heterogeneity in firms' ownership decisions. In Antràs and Helpman (2004), heterogeneous ownership decision depend two assumptions: (1) heterogeneous productivity, and (2) the fixed cost of vertical integration is higher than the fixed cost of outsourcing. In my model, a stochastic fixed cost assumption drives heterogeneous ownership decisions.

Recall that the random term ε_i^k follows an i.i.d. zero-mean type I Gumbel distribution. By McFadden (1973), the probability of a firm-pair i choosing k as its ownership structure is

$$\Pr\{k_i^* = k\} = \frac{\exp\{\psi(\beta^k, \eta) - f^k\}}{\sum_{l \in \{BI,SI,NI\}} \exp\{\psi(\beta^l, \eta) - f^l\}}. \tag{17}$$

Theorem 1 and Corollary 1 describe some properties of the above model.

Theorem 1. *$\Pr\{k_i^* = BI\}$ is increasing in η; $\Pr\{k_i^* = SI\}$ is decreasing in η.*

Theorem 1 states that the pair's probability of choosing buyer integration decreases in η, and its probability of choosing seller integration increases in η. This is a *stochastic* version of Lemma 2. I do not repeat the intuition here.

Corollary 1. *Define the log odds-ratio of pair i choosing BI and SI as*

$$p^{BI} \equiv \ln\left(\frac{\Pr\{k^* = BI\}}{\Pr\{k^* = NI\}}\right) = \psi(\beta^{BI}, \eta) - \psi(\beta^{NI}, \eta) - f^{BI} + f^{NI},$$

and

$$p^{SI} \equiv \ln\left(\frac{\Pr\{k^* = SI\}}{\Pr\{k^* = NI\}}\right) = \psi(\beta^{SI}, \eta) - \psi(\beta^{NI}, \eta) - f^{SI} + f^{NI},$$

then p^{BI} is increasing in η and p^{SI} is decreasing in η.

The last step to solving this game is t^k – the transfer payment from the buyer to the seller under ownership structure k. It changes the two firms surpluses from step 4 to step 3. In other words, t is simply the difference between the difference between the buyer's surpluses in steps 3 and 4. Note that the two firms' surpluses in step 4 are respectively μ^k and $(1 - \mu^k)$ shares of the total surplus. Their surpluses in step 3 are respectively γ and $(1 - \gamma)$ shares of the total surplus. The joint surplus remains constant from step 3 to step 4. The transfer payment t can thus be expressed as a share of total surplus as well:

$$t(\beta^k, \eta) \equiv \mu(\beta^k, \eta)\psi(\beta^k, \eta) - \gamma\psi(\beta^k, \eta) = [\mu(\beta^k, \eta) - \gamma]\psi(\beta^k, \eta).$$

More specifically,

$$t(\beta^k, \eta) = \frac{(1-\alpha\eta)\beta^k - \gamma\{1 - \alpha[\beta^k\eta + (1-\beta^k)(1-\eta)]\}}{\{(1/\alpha)(c_b/\beta^k)^\eta(c_s/(1-\beta^k))^{1-\eta}\}^{\alpha/(1-\alpha)}}.$$

When $t(\beta^k, \eta) > 0$, the buyer pays the seller; when $t(\beta^k, \eta) < 0$, the seller pays the buyer. $t(\beta^k, \eta) > 0$ if and only if $\mu(\beta^k, \eta) > \gamma$. It is shown in Appendix C that $\mu(\beta^k, \eta)$ is strictly increasing in β^k and strictly decreasing in η. This implies that given k, a higher η decreases $\mu(\beta^k, \eta)$ and thus reduces the payment from the buyer to the seller. The reason is simple: when the buyer is more important, and the pair cannot change k to incentivize the buyer (k is kept constant), it would instead *lure* the buyer into this relationship by asking the buyer to pay a lower payment. If η is high enough, the seller is willing to pay the buyer for him to enter this relationship. The threshold value of η under ownership structure k is:

$$\eta(\beta^k, \gamma) \equiv \frac{\beta^k - [1 - \alpha(1-\beta^k)]\gamma}{\alpha[\beta^k(1-\gamma) + (1-\beta^k)\gamma]}.$$

For a given ownership structure k, $t(\beta^k, \eta) > 0$ if $\eta < \eta(\beta^k, \gamma)$ and $t(\beta^k, \eta) < 0$ if $\eta > \eta(\beta^k, \gamma)$.

Note that $t(\beta^{k_i^*}, \eta)$, the *optimal* transfer payment, is not necessarily decreasing in η. The reason is because $t(\beta^k, \eta)$ is increasing in β and decreasing in η, and $\beta^{k_i^*}$ is weakly increasing in η. An increase in η directly decreases $t(\beta^{k_i^*}, \eta)$, but indirectly increases $\beta^{k_i^*}$. Which effect dominates is indeterminate. However, around $\underline{\eta}$ and $\bar{\eta}$ from Theorem 1, an increasing in η weakly decreases the transfer payment, but an increase in β increases the transfer payment. The latter is likely to dominate the former. Therefore, the optimal transfer payment $t(\beta^{k_i^*}, \eta)$ is likely to be a step function that is strictly decreasing in η in each segment.

3.3 Industry Equilibrium of the Heterogeneous Model

In the heterogeneous model I introduce a productivity term θ_i that varies across firm pairs. The new production function is

$$q_i = \theta_i \left(\frac{x_{i,b}}{\eta}\right)^{\eta} \left(\frac{x_{i,s}}{1-\eta}\right)^{1-\eta}, \tag{18}$$

There is now one extra step between steps 1 and 2, where each pair of firms draws a productivity level from an i.i.d. distribution function. From step 2 onward, the game the same as before. I highlight only the important results in this new equilibrium.

The heterogeneous versions of equations (15), (16) and (17) are

$$k_i^* = \underset{k\in\{BI,SI,NI\}}{\arg\max}\ \pi(\beta^k, \theta_i, \eta), \tag{19}$$

$$\pi(\beta^k, \theta_i, \eta) = \theta_i^{\alpha/(1-\alpha)}\psi(\beta^k, \eta) - f_i^k, \tag{20}$$

and

$$\Pr\{k_i^* = k\} = \frac{\exp\{\theta_i^{\alpha/(1-\alpha)}\psi(\beta^k, \eta) - f^k\}}{\sum_{l\in\{BI,SI,NI\}} \exp\{\theta_i^{\alpha/(1-\alpha)}\psi(\beta^l, \eta) - f^l\}}. \tag{21}$$

$\psi(\beta^k, \eta)$ is the same as defined in equation (13), and f_i^k follows the same distribution as in the homogeneous model.

Theorem 2. *There are two threshold values of η, $\underline{\eta}$ and $\bar{\eta}$, where $\underline{\eta}$ and $\bar{\eta}$ are the same as defined in Theorem 1, such that*

1. *When $\eta < \underline{\eta}$, $\Pr(k_i^* = SI)$ increases in θ_i and $\Pr(k_i^* = BI)$ decreases in θ_i;*
2. *When $\underline{\eta} < \eta < \bar{\eta}$, $\Pr(k_i^* = NI)$ increases in θ_i;*
3. *When $\eta > \bar{\eta}$, $\Pr(k_i^* = BI)$ increases in θ_i and $\Pr(k_i^* = SI)$ decreases in θ_i.*

3.4 Comparison to a Principal-Agent Framework

The key difference between my model and a PRT model with the principal assumption is not the existence of a principal firm, but the choice set available to a pair of firms. If I break step 3 into the following two steps, my model delivers the same results *with* the existence of a principal firm:

– The buyer and seller bargain over who becomes the principal with bargaining weights γ and $(1-\gamma)$. $0<\gamma<1$. The bargaining results in a principal firm $p \in \{B, S\}$ being chosen and a transfer payment $t \in \mathbf{R}$ from the buyer to the seller.

– The principal firm chooses between integration and outsourcing. Denote the principal's choice by $k \in \{I, O\}$. Together with this choice of organizational form, the principal firm pays the agent firm an upfront payment of $\tau \in \mathbf{R}$.

With this new setup, there are four possible equilibria: the buyer becomes the principal and chooses integration (BI), the buyer becomes the principal and chooses outsourcing (BO), the seller becomes the principal and chooses integration (SI), and the seller becomes the principal and chooses outsourcing (SO). BI and SI are identical to the previous models. BO and SO combined corresponds to NI in the previous models. The equivalence to Lemma 2 in this new model would contain three instead of two threshold values of η. As η increases from 0 to 1, the firm pair's organizational choice is ordered as $SI \rightarrow SO \rightarrow BO \rightarrow BI$. I do not use a principal-agent framework in this paper because my database naturally allows me to identify three types of organizational forms (BI, SI and NI) instead of four (BI, BO, SI and SO).

It is now clear why the key difference between my model and a model with the principal assumption does not center around the existence of a principal firm. What I have been referring to as the PRT models with the principal assumption should be more accurately described as PRT models with unilateral-integration assumption. For consistency, I will continue to refer to these models as the PRT models with the principal assumption.

4 Empirical Results for the Homogeneous Model

In this section I test the predictions of the homogeneous model by estimating a multinomial logit model and a system of linear probability models. I address the reverse causality issue of the relative R&D/goodwill intensity by running an IV regression and estimating a multinomial logit model over a sample of seller-buyer relationships that were outsourced in the beginning and later become integrated.

4.1 Multinomial Logit Model

Recall that in the homogeneous model, if the error term ε_i^k follows a Gumbel distribution, the probability of pair i choosing ownership structure k is

$$\Pr\{k_i^* = k\} = \frac{\exp\{\psi(\beta^k, \eta) - f^k\}}{\sum_{j \in \{BI,SI,NI\}} \exp\{\psi(\beta^j, \eta) - f^j\}}.$$

Theorem 1 predicts that pair i's probability of choosing buyer integration is increasing in η and its probability of choosing seller integration is decreasing in η. To test this prediction, I estimate the following multinomial logit model

$$\Pr(y_i = k) = \frac{\exp\{\varphi_0^k \cdot rd_i + X_i'\varphi^k\}}{\sum_{j \in \{BI,NI,SI\}} \exp\{\varphi_0^j \cdot rd_i + X_i'\varphi^j\}}, \tag{22}$$

where y_i is the observed ownership structure of pair i and $\varphi_0^k \cdot rd_i + X_i'\varphi^k$ is an empirical proxy for $\psi(\beta^k, \eta) - f^k$. rd_i is the buyer firm's relative R&D intensity in pair i as defined in Section 2.3, which is an empirical proxy for η. X_i is a vector of pairwise characteristics including logs of the buyer and seller's sales, age, employment, and pairwise sector and region fixed effects[30]. φ_0^k is the coefficient for rd_i and φ^k is the vector coefficient for X_i. Theorem 1 predicts that $\varphi_0^{BI} > 0$ and $\varphi_0^{SI} < 0$.

[Table 7 about here.]

[30] As mentioned before, there are 10 sectors and 5 regions in my database. I could not add in higher dimensions of sector-pair and region-pair fixed effects due to well-known convergence difficulties in multinomial logit models.

Table 7 summarizes the regression results for equation (22) with non-integration as the base outcome[31]. rd_i is the buyer firm's relative R&D intensity in pair i, as defined in Section 2.3. itl and $crind$ are dummy variables that equal one when the two firms in pair i are located in different countries and different industries, respectively. Columns (1)-(3) report the regression results with rd_i as the variable of interest. Columns (4)-(6) report the regression results with gdw_i as the variable of interest. In both groups, a higher rd_i or gdw_i is associated with a higher probability of pair i choosing buyer integration, and a lower probability of pair i choosing seller integration. These effects remain significant after controlling for pair-specific characteristics, i.e., logs of buyer and seller's sales, age, and employment, and adding in pairwise sector and region fixed effects. These results support the predictions of Theorem 1. Since the regression results for rd_i and gdw_i are very similar, in my ensuing interpretations I will focus on rd_i under the premise that what applies to rd_i also applies to gdw_i.

The regression coefficients in Table 7 do not provide any information on the magnitude of the effect of rd_i on pair i's probability of choosing various ownership structures. Table 8 reports the average marginal effects of different changes in rd_i on pair i's probability of choosing various ownership structures.

[Table 8 about here.]

As Table 8 shows, the average marginal effect of rd_i on $\Pr(y_i = BI)$ is 0.176 and on $\Pr(y_i = SI)$ is -0.100. When rd_i changes from its minimum to its maximum, it creases pair i's probability of choosing buyer integration by 0.168, and decreases its probability of choosing seller integration by 0.095. When rd_i changes from $rd_i - 0.5$ to $rd_i + 0.5$, pair i's probability of choosing buyer integration increases by 0.175, and its probability of choosing seller integration decreases by 0,075. When rd_i changes from $rd_i - 0.5SD$ to $rd_i + 0.5SD$, where SD is rd_i's standard deviation, which is 0.441, $\Pr(y_i = BI)$ increases by 0.078 and $\Pr(y_i = SI)$ decreases by 0.034. The effect of rd_i on pair i's organizational choice is nontrivial at the least.

To further assess the predictive power of rd_i, I construct several *interference* variables in a

[31] See Table E.4 for the regression results on the original sample (the sample without the imputed relationships).

similar manner as rd_i, namely, I calculate the buyer's relative R&D expenditure, relative total revenue, relative employment and relative age, and run a multinomial logistic regression of y_i on rd_i and these interference variables. I then plot in Figure 2 the average predicted probabilities of firm pairs choosing buyer integration (bi) and seller integration (si) at various levels of each of these variables. Compared to relative total revenue, relative employment and relative age, relative R&D intensity has the largest influence on pair i's organizational decision[32].

[Figure 2 about here.]

Another interesting finding in Table 7 is the interaction between rd_i and two dummy variables itl and $crind$. itl is a dummy that equals one when the buyer and seller are located in different countries. $crind$ is a dummy that equals one when the buyer and seller operate in difference industries. The coefficients on $rd_i * itl$ are of the same signs as the coefficients on rd_i, indicating that international seller-buyer relationships are more sensitive to contractual frictions. The coefficient for the dummy itl is negative for both $\Pr(y_i = BI)$ and $\Pr(y_i = SI)$, suggesting that compared to domestic seller-buyer relationships, international seller-buyer relationships are less likely to be integrated.

Similar analysis on $crind$ suggests that cross-industry seller-buyer relationships are less likely to choose integration relative to within-industry relationships. Cross-industry buyer-integration relationships are more sensitive to contractual frictions than within-industry relationships. However, there is no strong evidence that cross-industry seller-integration relationships are more sensitive to contractual frictions than the within-industry ones.

[Figure 3 about here.]

To see how the itl and $crind$ dummies affect firms' organizational choice at different levels of rd_i, I plot in Figure 3 the average probabilities of firm pairs choosing buyer integration (bi) and seller integration (si) for different groups of seller-buyer relationships at different values of rd_i.

[32] See the regression results and the plot for buyer's relative R&D expenditure in Table E.2 and Figure E.1.

Panel (a) plots the average predicted probabilities for three groups of firms: all relationships in the regression sample (all), international relationships where the buyer and seller are located in different countries (itl), and domestic relationships where the buyer and seller are located within the same country (dom). First, at lower values of rd_i, domestic relationships are more likely to integrate; at higher values of rd_i, international relationships are more likely to integrate. Second, international relationships are more responsive (have a larger curvature) to changes in rd_i than domestic relationships, suggesting that international relationships are more sensitive to contractual frictions. Third, buyer integration relationships are more response to changes in rd_i than seller integration relationships.

Panel (b) also plots the average predicted probabilities for three groups of relationships: all relationships in the regression sample (all), cross-industry relationships (crind), and within-industry relationships (wtind). At all levels of rd_i, within-industry relationships are more likely to be integrated, and are more sensitive to changes in rd_i than within-industry relationships.

[Table 9 about here.]

To see how the international and cross-industry dummies interact with each other, Table 9 divides all seller-buyer relationships into 4 groups based on whether the two firms in a seller-buyer relationship are located in the same country or the same industry, and report the average predicted probabilities of firm pairs choosing various ownership structures for each group of relationships. As Table 9 shows, for both international and domestic relationships, cross-industry seller-buyer relationships are much less likely to be integrated than within-industry ones. For both cross-industry and within-industry relationships, international seller-buyer relationships are slightly *more* likely to be integrated. This result seems to contradict with the regression results in Table 7, which suggests that international relationships are *less* likely to be integrated. Note that the difference here is that Table 8 computes the average probability while Table 7 computes the log odds ratios. The two seemingly contradictory results are not necessarily mutually exclusive.

To summarize, I find positive support for Theorem 1. I also find that international seller-buyer relationships are significantly more sensitive to contractual frictions than domestic relationships. On average, cross-country and within-industry seller-buyer relationships are more likely to choose integration than the domestic and cross-industry ones.

4.2 Linear Probability Models

A multinomial logit model is the closest to the structure of my theoretical model. However, it does not provide a comparison between my model and a model with a principal assumption, nor does it allow me to control for fixed effects at higher dimensions. Linear probability models help fix these issues.

A PRT model with the principal assumption suggests the following linear probability model

$$\Pr(VI_i = 1) = \varphi_0^V \cdot rd_i + X_i' \varphi^V + \varepsilon_i^V, \tag{23}$$

where i is a pair index. VI_i is a dummy that equals one if pair i is integrated and zero if pair i is outsourced. rd_i and X_i are the same as previously defined. φ_0^V is the coefficient of interest. The theory's prediction on φ_0^V depends on the principal assumption: if the buyer is assumed to be the principal, φ_0^V should be positive; if the seller is assumed to be the principal, φ_0^V should be negative[33].

It is clear from equation (23) that a model with the principal assumption can support *any* significant value of φ_0^V. If φ_0^V is positive and significant, it supports a PRT model with the buyer as the principal; if φ_0^V is negative and significant, it supports a PRT model with the seller as the principal. As Whinston (2001) points out: "the discussion...sidesteps one important issue by assuming that any observed integration is buyer integration. When this is not clear a priori, the PRT's prediction...will depend on whether the type of integration (buyer vs. seller) is observable."

[33]The idea is simple. Compared to outsourcing, integration provides the principal with a higher ex-post bargaining power and hence a higher ex-ante investment incentive. Outsourcing provides the agent with higher ex-post bargaining power and ex-ante investment incentive. The principal chooses integration when the principal's non-contractible, relationship-specific investment is relatively important, and outsourcing when the agent's non-contractible, relationship-specific investment is relatively important.

Whinston's statement explains both the source of bias, and the reason why the principal assumption is so widely used in the current literature – most databases do not allow researchers to observation the direction of integration. With the direction of integration observable in my database, I am able to estimate the following two linear probability models:

$$\Pr(BI_i = 1) = \varphi_0^B \cdot rd_i + X_i' \varphi^B + \varepsilon_i^B, \tag{24}$$

and

$$\Pr(SI_i = 1) = \varphi_0^S \cdot rd_i + X_i' \varphi^S + \varepsilon_i^S, \tag{25}$$

where i, rd_i and X_i are the same as defined above. BI_i is a dummy variable that equals one if pair i chooses buyer integration. SI_i is a dummy variable that equals one if pair i chooses seller integration. Theorem 1 predicts that $\varphi_0^B > 0$ and $\varphi_0^S < 0$.

Note that $VI_i = 1$ whenever $BI_i = 1$ or $SI_i = 1$. Since $BI_i = 1$ and $SI_i = 1$ are mutually exclusive, $\Pr(VI_i = 1) = \Pr(BI_i = 1) + \Pr(SI_i = 1)$. This implies that if there is only one type of integration, either φ_0^B or φ_0^S equals 0, my model would deliver the same result as the PRT model with the principal assumption. However, if buyer integration and seller integration coexist, φ_0^V is simply a combination of φ_0^B and φ_0^S. Equation (23) is no longer a test of PRT.

[Table 10 about here.]

Table 10 reports the regression results from equations (23), (24) and (25) for three types of industry pairs: all industry pairs, industry pairs with no buyer integration relationships, and industry pairs with no seller integration relationships. Panels (a) uses rd_i as the variable of interest and (b) uses gdw_i as the variable of interest. Columns (1) and (2) suggest that my model is supported by the linear regressions: with a higher rd_i or gdw_i is associated with a higher likelihood of pair i choosing buyer integration and a lower likelihood of pair i choosing seller integration. Column (3) shows that with a higher rd_i or gdw_i, pair i is more likely to be integrated. This result appears to support a PRT model with the buyer as the principal. However, as explained before, column (3) is a conflation of columns (1) and (2). A closer inspection reveals that the coefficients in column (3) are simply the sums of the coefficients in columns (1) and (2): $0.072 = 0.25 + (-0.178)$, and $0.075 = 0.223 + (-0.148)$. In other words, the reason I was able to find positive support for the former is not because buyers are indeed the principals, but because the coefficient in column(1) is larger in magnitude than the coefficient in column (2). If they were close in magnitudes, I would not be able to find any support for the PRT model regardless of who is assumed to be the principal.

To test if the patterns in Table 10 depend on the linearity assumptions of the linear probability models, I report in Table E.3 the logit regression results for all seller-buyer relationship. See Table E.5 and Table E.6 for the linear and logit regression results for the original sample (the sample without imputed relationships).

4.3 Robustness Checks

One concern with the relative R&D/goodwill intensity measure is reverse causality. This concern is consistent with PRT. A firm pair chooses buyer/seller integration to incentivize the buyer/seller, but this comes at the expense of the seller/buyer. After buyer/seller integration, we expect to see an increase in the buyer/seller's R&D/goodwill intensity, and a decrease in the seller/buyer's R&D/goodwill intensity. In this section, I address the reverse causality concern in two ways. In Section 4.3.1, I construct a series of instrumental variables for the buyer's R&D/goodwill intensity at the industry-country level. In Section 4.3.2, I utilize the fact that I have ownership samples from three years, and run the benchmark regressions for the seller-buyer relationships that are *later* observed to be integrated. In both cases, I am able to find positive support for my model.

4.3.1 IV regressions

The first instrumental variable I construct for the buyer firm's relative R&D/goodwill intensity is the buyer firm's relative industry-country level R&D/goodwill intensity.[34] The exclusion

[34]The industry-country level R&D intensity is calculated by dividing the total R&D expenditure at the industry-country level by the total sales at the industry-country level. The buyer's relative R&D intensity

restriction holds if a firm's industry-country level R&D/goodwill intensity affects the firm's R&D intensity but does not affect the ownership structure between this firm and other firms.

To address this concern, I construct two more instrumental variables: the industry-country-ownership level R&D/goodwill intensity, and the *weighted* industry-country-ownership level R&D/goodwill intensity. The former calculates the buyer and seller's industry-country level R&D/goodwill intensity for each ownership type. The latter calculates the buyer and seller's industry-country level R&D/goodwill intensity weighted by the number of ownership structures at the pairwise industry-country levels. These variables correct for the industry-country level R&D/goodwill intensity's effect on firms' choices of ownership structures.

Table 11 reports the IV regression results. R&D/goodwill intensity remain strong and of the correct signs.

[Table 11 about here.]

4.3.2 Seller-buyer relationships with ex-post ownership

Recall that I have ownership samples from three years: 2013, 2014 and 2016. Another way to address the reverse causality concern is by keeping only seller-buyer relationships that are previously non-integrated and later become integrated. As shown in Table 12 shows, of all 2,499 seller-buyer relationships, 983 come from the 2013 sample, 861 come from the 2014 sample, and 655 come from the 2016 sample[35].

[Table 12 about here.]

Table 13 reports the multinomial logistic regression results for the ex-post ownership sample. rd_i, gdw_i and itl remain significant and of the same signs as before. Other variables are no longer significant. Note that the sample size has dropped to roughly 1.17% of the sample in

is the ratio between the buyer's industry-country level R&D intensity and the sum of the buyer and seller's industry-country level R&D intensity.

[35]The 655 seller-buyer relationships from 2016 are *non-recent* seller-buyer relationships. Capital IQ defines a seller-buyer relationship as "not recent" if it was previous documented but no longer observed for the past two years.

Table 7 (2110 versus 122,765). The new sample contains 128 buyer-integration relationships (6%), 168 seller-integration relationships (8%), and 1,814 non-integration relationships (84%). The results in Table 7 and Table 13 are not directly comparable. I conclude that my model *does* find positive support in the ex-post ownership sample.

[Table 13 about here.]

5 Empirical Results for the Heterogeneous Model

In this section, I test the predictions of Theorem 2 with a multinomial logistic model and a system of linear probability models. I find positive support for Theorem 2 in both models. I present my linear regression results first because it is easier to explain my empirical specifications with the linear probability models.

5.1 Linear Regression Results

Theorem 2 demands an empirical proxy for pair i's productivity level, $prod_i$. I define $prod_i$ as the weighted average of the logs of the buyer and seller firm's productivity levels:

$$prod_i = \frac{size_b}{size_b + size_s} \ln prod_b + \frac{size_s}{size_b + size_s} \ln prod_s, \tag{26}$$

where $prod_b$ and $prod_s$ are the buyer and seller's sales per employee. $size_b$ and $size_s$ are the buyer and seller's sales. The theory suggest the following empirical specifications

$$\Pr(BI_i = 1) = prod_i \sum_{q=1}^{Q} \lambda_q^B rd_q + X_i' \nu^B + \xi_i^B; \tag{B}$$

$$\Pr(SI_i = 1) = prod_i \sum_{q=1}^{Q} \lambda_q^S rd_q + X_i' \nu^S + \xi_i^S; \tag{S}$$

$$\Pr(NI_i = 1) = prod_i \sum_{q=1}^{Q} \lambda_q^N rd_q + X_i' \nu^N + \xi_i^N. \tag{N}$$

BI_i, SI_i, NI_i, and X_i are the same as defined before. $prod_i$ is the empirical proxy for pair i's productivity level as defined in equation (26). rd_q is the qth quantile of rd_i. λ_q^k is the coefficient for the qth quantile of rd_i in equation (k), where $q = 1, ..., Q$ and $k \in \{B, S, N\}$. ν^k is the vector of coefficients for X_i in equation (k), and ξ_i^k is the error term for equation (k).

Recall that rd_i is the empirical proxy for η in the model. Theorem 2 makes three predictions:

1. At lower levels of q, $\lambda_q^S > 0$ and $\lambda_q^B < 0$;
2. At intermediate levels of q, $\lambda_q^N > 0$;
3. At higher levels of q, $\lambda_q^B > 0$ and $\lambda_q^S < 0$.

Since the distributions of rd_i and $(1 - rd_i)$ are different, the cutoffs for these two variables are also different. If I replace rd_q in the above equations with $(1 - rd)_q$, the qth quantile of the *seller*'s relative R&D intensity, then Theorem 2's predictions are reversed at the two ends of q:

1. At lower levels of q, $\lambda_q^S < 0$ and $\lambda_q^B > 0$;
2. At intermediate levels of q, $\lambda_q^N > 0$;
3. At higher levels of q, $\lambda_q^B < 0$ and $\lambda_q^S > 0$.

[Table 14 about here.]

Table 14 reports the linear regression results for tertiles of rd_i and $(1 - rd_i)$. Panel (a) reports the results for rd_i. At the first tertile, more productive firm pairs are more likely to choose seller integration and less likely to choose buyer integration. At the third tertile, more productive firm pairs are more likely to choose buyer integration and less likely to choose seller integration. The regression results for the second tertile are similar to those for the third tertile. These results support the predictions of Theorem 2 at the higher and lower levels of η, but there is no support for the intermediate levels of η. Panel (b) reports the results for $(1 - rd_i)$. At the first tertile, more productive firm pairs are more likely to choose buyer integration and less likely to choose seller integration. At the third tertile, more productive firm pairs are more likely to choose seller integration and less likely to choose buyer integration. At the second tertile, more productive firm pairs are *less* likely to choose non-integration, but the coefficients are less significant. Again, there are supports for Theorem 2 at the higher and lower levels of η, but there is no support for the intermediate level of η. See Table E.7 and Table E.8 for quintiles and septiles of rd_i and $(1 - rd_i)$. The patterns are similar to those in Table 14: there are positive supports for higher and lower levels of η for buyer and seller integration, but no suppoer for the intermediate levels of η.

The reason for the lack of support for Theorem 2 at the intermediate levels of η is twofold. First, the threshold levels of η are not clear-cut. It is easier to look for patterns when BI_i and SI_i are the dependent variables because the predictions regarding buyer integration and seller integration concern the two extremes of the quantiles. However, it is difficult to tell where exactly are the *intermediate* levels of rd_i and $(1-rd_i)$ located. Second, the heterogeneous model predicts a nominal outcome with three values, not a dummy as in the linear probability models. The linearity assumption may not fit the model very well, especially at the intermediate levels of the independent variables. To eliminate the second concern, I present in the next section the multinomial logit model for the heterogeneous model.

5.2 The Multinomial Logistic Regression Results

[Table 15 about here.]

Table 15 reports the multinomial logistic regression results. The results are similar to those from Table 14. Columns (1)-(3) document the regression results for tertiles of rd_i without and with pairwise sector and region fixed effects. At the first tertile of rd_i, more productive firm pairs are less likely to choose buyer integration and more likely to choose seller integration. At the third tertile of rd_i, more productive firm pairs are more likely to choose buyer integration and less likely to choose seller integration. The second tertile is similar to the third tertile in that more productive firm pairs are more likely to choose buyer integration and less likely to choose seller integration. Columns (4)-(6) report the regression results for tertiles of $(1 - rd_i)$.

In the first tertile, more productive firm pairs are more likely to choose buyer integration and less likely to choose seller integration. In the third tertile, more productive firm pairs are less likely to choose buyer integration and more likely to choose seller integration. The second tertile is similar to the first tertile in the sense that more productive firm pairs are more likely to choose buyer integration and less likely to choose seller integration.

Since non-integration is used as the base outcome for the multinomilal logistic regressions, there is no results regarding the third prediction of Theorem 2. I present in Table 16 the average marginal effects of $prod_i$ on firm pairs' probabilities of choosing various ownership structures at tertiles of rd_i and $(1 - rd_i)$. At the first $(q = 1)$ and third $(q = 3)$ tertiles, the marginal effects of $prod_i$ are just what Theorem 2 predicts. In the second $(q = 2)$ tertile, however, the effects of $prod_i$ are different for rd_2 and $(1 - rd)_2$. For rd_2, $prod_i$ becomes insignificant for all outcomes. For $(1-rd)_2$, the effects of $prod_i$ are similar to those at the first tertile. For both rd_q and $(1 - rd)_q$, there does not seem to be a clear relatoinships between $\Pr(y_i = NI)$ and $prod_i$. This again illustrates the difficulty of finding empirical support for non-integration decisions.

[Table 16 about here.]

To summarize, the multinomial logistic regressions provide positive support for two out of three predictions of Theorem 2. The third prediction is empirically difficult to test.

6 Conclusion

In this paper, I compile the first relationship-level database that shows the coexistence of backward and forward integration within a pair of industries. Of all the industry pairs with integrated relationships, 28% of which contain both types of integration relationships. If the industry pairs are weighted by their total numbers of relationships, this percentage goes up to 61%.

To explain this coexistence, I combine the elements from Grossman and Hart (1986) and Antràs and Helpman (2004) to construct a PRT model with bi-lateral integration decisions. I use the buyer firm's relative R&D/goodwill intensity, rd_i, to proxy for the relative importance of the buyer firm's non-contractible, relationship-specific investment in a seller-buyer relationship. I find that on average, a standard deviation increase in the buyer's relative R&D intensity is associated with a 0.078 increase in the firm pair's likelihood of choosing buyer integration, and a 0.034 decrease in the firm pair's likelihood of choosing seller integration. I also find that international seller-buyer relationships are less likely to be integrated, but are more sensitive to contractual frictions (as measured by a change in the buyer firm's relative R&D intensity).

I show that a PRT model with the principal assumption suffers from estimation bias in industries where seller and buyer integration coexist. In the linear probability model, the magnitude of the coefficient for rd_i in a PRT model with the principal assumption is roughly one-third of that in my model without the principal assumption. This result suggest that previous empirical literature testing PRT models with the principal assumption is likely to suffer from under-estimation.

By incorporating heterogeneous firm-pair productivity, I find (theoretically and empirically) that in seller-buyer relationships where the buyer firm's non-contractible, relationship-specific investment is relatively important, more productive firm pairs are more likely to choose buyer integration and less likely to choose seller integration. In seller-buyer relationships where the seller firm's non-contractible, relationship-specific investment is relative important, more productive firm pairs are more likely to choose seller integration and less likely to choose buyer integration.

To conclude, my analysis shows that PRT is a useful framework for analyzing the ownership structure of global supply chains. There are many potential areas for future research. First, it should be noted that the testable predictions in this paper are derived from a partial equilibrium model – firms' industry and location decisions are taken as exogenous. My future work will extending the current model to a general equilibrium framework allows one to study how firms' location and make-or-buy decisions are affected by country characteristics such as contracting/financial institutions, corporate tax laws, labor market frictions, etc. Second, embedding the one-to-one seller-buyer relationships in an input-output network provides a useful framework for analyzing the implications of firm's make-or-buy decisions on the distribution inequality along the global value chain. Third, combining this global database with a firm-level database may generate interesting implications for labor market outcomes, e.g., unemployment, wage inequality, labor demand.

DEPARTMENT OF ECONOMICS, MAXWELL SCHOOL OF CITIZENSHIP AND PUBLIC AFFAIRS, SYRACUSE UNIVERSITY

Appendix

A Proofs of Lemma 1 and Lemma 2

The proof of this theorem is very similar to that in Antràs and Helpman (2004). Recall that pair i's the joint surplus is

$$\psi(\beta,\eta) = \frac{1-\alpha[\beta\eta+(1-\beta)(1-\eta)]}{\{(1/\alpha)\left(\frac{c_b}{\beta}\right)^{\eta}\left(\frac{c_s}{1-\beta}\right)^{1-\eta}\}^{\frac{\alpha}{1-\alpha}}}.$$

If the firm pair is allowed to choose any $\beta \in (0,1)$ to maximize $\psi(\beta,\eta)$, the optimal β, $\beta^*(\eta)$, solves the first order condition, and

$$\beta^*(\eta) = \frac{\eta(\alpha\eta+1-\alpha)-\sqrt{\eta(1-\eta)(1-\alpha\eta)(\alpha\eta+1-\alpha)}}{2\eta-1}.$$

Figure A.1 plots the optimal β as a function of η. As the figure shows, $\beta^*(\eta)$ is increasing in η, with $\lim_{\eta\to 1}\beta^*(\eta)=0$ and $\lim_{\eta\to 1}\beta^*(\eta)=0$. By Implicit Function Theorem, $\psi(\beta,\eta)$ is supermodular in (β,η).

[Figure A.1 about here.]

Since the firm pair cannot choose β as a continuous variable, it has to choose from one of three discrete values of β: β^{BI}, β^{SI} and β^{NI}, the firm pair's actual choice of β is an increasing step function of η. In other words, when $\eta > \bar{\eta}$, the seller and buyer choose β^{BI} by choosing $k = BI$. When $\eta < \underline{\eta}$, the seller and buyer choose β^{SI} by choosing $k = SI$. When $\underline{\eta} < \eta < \overline{\eta}$, the seller and buyer choose β^{NI} by choosing $k = NI$.

B Proof of Theorem 1

After introducing the i.i.d fixed costs, the profit functions of the homogeneous model can be written as

$$\pi_i(\beta^k, \eta) = \psi(\beta^k, \eta) - f_i^k,$$

where $\psi(\beta^k, \eta)$ is as defined in the previous section, and

$$f_i^k = f^k + \varepsilon_i^k.$$

Since ε_i^k follows an i.i.d. type-I Gumbel distribution with zero mean, according to McFadden (1973), the probability that buyer integration maximizes pair i's profit is

$$\Pr\{k^* = BI\} = \frac{\exp\{\psi(\beta^{BI}, \eta) - f^{BI}\}}{\exp\{\psi(\beta^{BI}, \eta) - f^{BI}\} + \exp\{\psi(\beta^{SI}, \eta) - f^{SI}\} + \exp\{\psi(\beta^{NI}, \eta) - f^{NI}\}},$$

or

$$\Pr\{k^* = BI\} = \frac{1}{1 + \exp\{\psi(\beta^{SI}, \eta) - \psi(\beta^{BI}, \eta) - f^{SI} + f^{BI}\} + \exp\{\psi(\beta^{NI}, \eta) - \psi(\beta^{BI}, \eta) - f^{NI} + f^{BI}\}}.$$

$\Pr\{k^* = BI\}$ is increasing in η if $\psi(\beta^{BI}, \eta) - \psi(\beta^{SI}, \eta)$ and $\psi(\beta^{BI}, \eta) - \psi(\beta^{NI}, \eta)$ are increasing in η. Given that $\beta^{BI} > \beta^{NI} > \beta^{SI}$, a sufficient condition for the previous property is if $\psi(\beta, \eta)$ satisfies increasing differences in (β, η). I showed in the previous section that $\psi(\beta, \eta)$ is supermodular in (β, η), which is a sufficient condition for increasing differences. Therefore, $\Pr\{k^* = BI\}$ is increasing in η.

Similarly,

$$\Pr\{k^* = SI\} = \frac{1}{1 + \exp\{\psi(\beta^{BI}, \eta) - \psi(\beta^{SI}, \eta) - f^{BI} + f^{SI}\} + \exp\{\psi(\beta^{NI}, \eta) - \psi(\beta^{SI}, \eta) - f^{NI} + f^{SI}\}}.$$

Since $\beta^{BI} > \beta^{NI} > \beta^{SI}$ and $\psi(\beta, \eta)$ is supermodular in (β, η), $\psi(\beta^{BI}, \eta) - \psi(\beta^{SI}, \eta)$ and $\psi(\beta^{NI}, \eta) - \psi(\beta^{SI}, \eta)$ are increasing in η. Therefore, $\Pr\{k^* = SI\}$ is decreasing in η.

C Properties of Joint Surplus Shares

μ^k is strictly increasing in β^k and strictly decreasing in η. Recall that

$$\mu^k \equiv \frac{(1-\alpha\eta)\beta^k}{1 - \alpha[\beta^k\eta + (1-\beta^k)(1-\eta)]}$$

as defined in equation (14). Taking the derivative of μ^k with respect to β^k and η gives

$$\frac{\partial \mu^k}{\partial \beta^k} = \frac{(1-\alpha\eta)(1-\alpha+\alpha\eta)}{\{1 - \alpha[\beta^k\eta + (1-\beta^k)(1-\eta)]\}^2}$$

and

$$\frac{\partial \mu^k}{\partial \eta} = \frac{\alpha(\alpha-2)\beta^k(1-\beta^k)}{\{1 - \alpha[\beta^k\eta + (1-\beta^k)(1-\eta)]\}^2}$$

$\frac{\partial \mu^k}{\partial \beta^k} > 0$ and $\frac{\partial \mu^k}{\partial \eta} < 0$ because $0 < \alpha, \beta, \eta < 1$.

D Proof of Theorem 2

Pair i's joint profit in the heterogeneous model is

$$\pi(\beta^k, \theta_i, \eta) = \theta_i^{\alpha/(1-\alpha)}\psi(\beta^k, \eta) - f_i^k,$$

where $\psi(\beta^k, \eta)$ and f_i^k are the same as previously defined. The pair's probability of choosing buyer integration is

$$\Pr\{k_i^* = BI\} = \frac{1}{1 + \exp\{\theta_i^{\alpha/(1-\alpha)}(\psi^{SI} - \psi^{BI}) - f^{SI} + f^{BI}\} + \exp\{\theta_i^{\alpha/(1-\alpha)}(\psi^{NI} - \psi^{BI}) - f^{NI} + f^{BI}\}},$$

where ψ^k is short for $\psi(\beta^k, \eta)$.

According to Lemma 2, when $\eta < \underline{\eta}$, $\psi^{SI} > \psi^{NI} > \psi^{BI}$, so $\psi^{SI} - \psi^{BI} > 0$ and $\psi^{NI} - \psi^{BI} > 0$. The denominator is increasing in θ, and $\Pr\{k_i^* = BI\}$ is decreasing in θ. When $\eta > \bar{\eta}$, $\psi^{BI} > \psi^{NI} > \psi^{SI}$, so $\psi^{SI} - \psi^{BI} < 0$ and $\psi^{NI} - \psi^{BI} < 0$. The denominator is decreasing in θ_i and $\Pr\{k_i^* = BI\}$ is increasing in θ_i. When $\underline{\eta} < \eta < \bar{\eta}$, $\psi^{NI} > \min\{\psi^{BI}, \psi^{SI}\}$. Although $\psi^{NI} - \psi^{BI} > 0$, $\psi^{SI} - \psi^{BI}$ cannot be signed. It is not clear whether $\Pr\{k_i^* = BI\}$ is increasing or decreasing in θ.

Similarly, pair i's probability of choosing seller integration is

$$\Pr\{k_i^* = SI\} = \frac{1}{1 + \exp\{\theta_i^{\alpha/(1-\alpha)}(\psi^{BI} - \psi^{SI}) - f^{BI} + f^{SI}\} + \exp\{\theta_i^{\alpha/(1-\alpha)}(\psi^{NI} - \psi^{SI}) - f^{NI} + f^{SI}\}},$$

where ψ^k is short for $\psi(\beta^k, \eta)$.

Use similar logic, one can show that $\Pr\{k_i^* = SI\}$ is increasing in θ_i when $\eta < \underline{\eta}$ and decreasing in θ_i when $\eta > \bar{\eta}$, but indeterminate when $\underline{\eta} < \eta < \bar{\eta}$.

Lastly, pair i's probability of choosing non-integration is

$$\Pr\{k_i^* = NI\} = \frac{1}{1 + \exp\{\theta_i^{\alpha/(1-\alpha)}(\psi^{BI} - \psi^{NI}) - f^{BI} + f^{NI}\} + \exp\{\theta_i^{\alpha/(1-\alpha)}(\psi^{SI} - \psi^{NI}) - f^{SI} + f^{NI}\}},$$

where ψ^k is short for $\psi(\beta^k, \eta)$.

When $\underline{\eta} < \eta < \bar{\eta}$, $\psi^{NI} > \min\{\psi^{BI}, \psi^{SI}\}$ implies $\psi^{BI} - \psi^{NI} < 0$ and $\psi^{SI} - \psi^{NI} < 0$. $\Pr\{k_i^* = NI\}$ is increasing in θ_i. When $\eta < \underline{\eta}$, $\psi^{BI} - \psi^{NI} < 0$ and $\psi^{SI} - \psi^{NI} > 0$. When $\eta > \bar{\eta}$, $\psi^{BI} - \psi^{NI} > 0$ and $\psi^{SI} - \psi^{NI} < 0$. In both cases, $\Pr\{k_i^* = NI\}$ is indeterminate in θ_i.

To summarize, when $\eta < \underline{\eta}$, $\Pr\{k_i^* = SI\}$ is increasing in θ_i and $\Pr\{k_i^* = BI\}$ is decreasing in θ_i; when $\eta > \bar{\eta}$, $\Pr\{k_i^* = BI\}$ is increasing in θ_i and $\Pr\{k_i^* = SI\}$ is decreasing in θ_i; when

$\underline{\eta} < \eta < \bar{\eta}$, $\Pr\{k_i^* = NI\}$ is increasing in θ_i.

E Supplementary Tables and Figures

[Table E.1 about here.]

[Table E.2 about here.]

[Table E.3 about here.]

[Table E.4 about here.]

[Table E.5 about here.]

[Table E.6 about here.]

[Table E.7 about here.]

[Table E.8 about here.]

[Table E.9 about here.]

[Table E.10 about here.]

[Figure E.1 about here.]

References

Acemoglu, Daron, Philippe Aghion, Rachel Griffith, and Fabrizio Zilibotti, "Vertical Integration and Technology : Theory and Evidence," *Journal of the European Economic Association*, 2010, *8* (5), 989–1033.

_ , **Simon Johnson, and Todd Mitton**, "Determinants of Vertical Integration: Financial Development and Contracting Costs," *The Journal of Finance*, 2009, *64* (3), 1251–1290.

_ , **Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi**, "The Network Origins of Aggregate Fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Alfaro, Laura and Andrew Charlton, "Intra-Industry Foreign Direct Investment," *The American Economic Review*, 2009, *99* (5), 2096–2119.

_ **and Maggie Xiaoyang Chen**, "The Global Agglomeration of Multinational Firms," *Journal of International Economics*, 2014, *94* (2), 263–276.

_ , **Pol Antràs, Davin Chor, and Paola Conconi**, "Internalizing Global Value Chains: A Firm-Level Analysis," 2015.

Altomonte, Carlo and Armando Rungi, "Business Groups as Hierarchies of Hirms: Determinants of Vertical Integration and Performance," 2013.

_ , **Filippo Di Mauro, Gianmarco I. P. Ottaviano, Armando Rungi, and Vincent Vicard**, "Working Paper Series Global Value Chains During the Great Trade Collapse a Bullwhip Effect ?," 2012.

Antràs, Pol, "Firms, Contracts, and Trade Structure," *The Quarterly Journal of Economics*, 2003, *118* (4), 1375—-1418.

_ , "Grossman-Hart (1986) Goes Global: Incomplete Contracts, Property Rights, and the International Organization of Production," *Journal of Law, Economics, and Organization*, 2013, *30* (suppl 1), i118—-i175.

_ , *Global Production: Firms, Contracts, and Trade Structure*, Princeton University Press, 2015.

_ **and Davin Chor**, "Organizing the Global Value Chain," *Econometrica*, 2013, *81* (6), 2127–2204.

_ **and Elhanan Helpman**, "Global Sourcing," *Journal of Political Economy*, 2004, *112* (3), 552—-580.

_ **and** _ , "Contractual Frictions and Global Sourcing," in "The Organization of Firms in a Global Economy" 2008, pp. 9–54.

_ **and Stephen R. Yeaple**, "Multinational Firms and the Structure of International Trade," *Handbook of International Economics*, 2014, *4*, 55–130.

_ **, Mihir A. Desai, and Fritz C. Foley**, "Multinational Firms, FDI Flows and Imperfect Capital Markets," *The Quarterly Journal of Economics*, 2008, *124* (3), 1171–1219.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson, "Vertical Integration and Input Flows," *The American Economic Review*, 2014, *104* (4), 1120–1148.

Bernard, Andrew B., Bradford J. Jensen, and Peter K. Schott, "Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods," in "Producer Dynamics: New Evidence from Micro Data" number 202 2009, pp. 513–552.

Coase, R.H., "The Nature of the Firm," *Economica*, 1937, *4* (16), 386–405.

Corcos, Gregory, Delphine M. Irac, Giordano Mion, and Thierry Verdier, "The Determinants of Intrafirm Trade: Evidence From French Firms," *The Review of Economics and Statistics*, 2013, *95(3)* (4), 825–838.

Defever, Fabrice and Farid Toubal, "Productivity, Relationship-specific Inputs and the Sourcing Modes of Multinationals," *Journal of Economic Behavior and Organization*, 2013, *94*, 345–357.

Dobson, Wendy and Chia Siow Yue, *Multinationals and East Asian Integration*, Vol. 10, Intl Development Research, 1999.

Eaton, Jonathan, Samuel S. Kortum, and Francis Kramarz, "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 2011, *79* (5), 1453–1498.

Feenstra, Robert C. and Gordon H. Hanson, "Ownership and Control in Outsourcing to China: Estimating the Property-Rights Theory of the Firm," *The Quarterly Journal of Economics*, 2005, *120* (2), 729–761.

Fernandes, Ana P. and Heiwai Tang, "Determinants of Vertical Integration in Export Processing: Theory and Evidence from China," *Journal of Development Economics*, 2012, *99* (2), 396–414.

Gibbons, Robert, "Four Formal(izable) Theories of the Firm?," *Journal of Economic Behavior and Organization*, 2005, *58* (2), 200–245.

Grossman, Sanford J. and Oliver D. Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 1986, *94* (4), 691—-719.

Helpman, Elhanan, *Understanding Global Trade* 2011.

Kohler, Wilhelm K. and Marcel Smolka, "Sourcing Premia with Incomplete Contracts: Theory and Evidence," *The B.E. Journal of Economic Analysis and Policy*, 2011, *11* (1).

_ **and** _ , "Global Sourcing and Firm Selection," *Economics Letters*, 2014, *124* (3), 411–415.

Lafontaine, Francine and Margaret Slade, "Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*, 2007, *45* (3), 629–685.

Li, Zhiyuan, "Task Offshoring and Organizational Form: Theory and Evidence from China," *Journal of Economic Behavior and Organization*, 2013, *94*, 358–380.

Long, J Scott and Jeremy Freese, *Regression Models for Categorical Dependent Variables using Stata* 2014.

McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in "Frontiers in Econometrics" 1973, pp. 105–142.

Nunn, Nathan, "Relationship-Specificity , Incomplete Contracts , and the Pattern of Trade," *The Quarterly Journal of Economics*, 2007, *122* (2), 569–600.

_ **and Daniel Trefler**, "The Boundaries of the Multinational Firm: An Empirical Analysis," in "The Organization of Firms in a Global Economy" 2008, pp. 55—-83.

_ **and** _ , "Incomplete Contracts and the Boundaries of the Multinational Firm," *Journal of Economic Behavior and Organization*, 2013, *94*, 330–344.

_ **and** _ , "Domestic Institutions as a Source of Comparative Advantage," in "Handbook of International Economics," Vol. 4 2014, p. 263.

Prete, D Del and A Rungi, "Organizing the Global Value Chain: a firm-level test," 2015.

Segal, Ilya and Michael D. Whinston, "Property Rights," in "Handbook of Organizational Economics" 2012, pp. 100–158.

Tomiura, Eiichi, "Foreign versus Domestic Outsourcing: Firm-level Evidence on the Role of Technology," *International Review of Economics and Finance*, 2009, *18* (2), 219–226.

Whinston, Michael D., "Assessing the Property Rights and Transaction-Cost Theories of Firm Scope," *The American Economic Review*, 2001, *91* (2), 184–188.

Williamson, Oliver E., "The Theory of the Firm as Governance Structure: From Choice to Contract," *Journal of Economic Perspectives*, 2002, *16* (3), 171–195.

Table 1: Average Marginal Effects of a One Standard Deviation Increase in Buyer's Relative R&D Intensity in PRT Models with and without the Principal Assumption

	Model without principal assumption		Model with principal assumption
	Buyer integration	Seller integration	Integration
One std. dev. increase in buyer's relative R&D intensity	0.193 (0.000)	-0.079 (0.000)	0.050 (0.000)

Notes: Buyer's relative R&D intensity is calculated by dividing the buyer's R&D intensity by the sum of buyer and seller's R&D intensities. A firm's R&D intensity is the ratios between the firm's total R&D expenditure and its total sales over the period 2005-2014. The standard deviation of the buyer firm's relative R&D intensity is 0.441. The numbers in this table are calculated using the logistic regression results. Numbers in parentheses report p-values.

Table 2: Number of Various Business Relationships

Relationship Type	Number of Relationships
Buyer-seller relationships	779,436
Firm-parent relationships	611, 335
Investor	843,405
Ultimate Parent	618,634
Holding Company	532,723
Merged Entity	181,104
Limited Partner	34,621
Pending Parent/Investor	23,531

Table 3: Ownership Distribution of the 2016 Sample

Ownership structure	Frequency	Percentage
Buyer/backward integration	14,400	11.73
Seller/forward integration	14,325	11.67
Non-integration	94,040	76.60
Total	122,765	100

Table 4: Distribution of 2016 Ownership Sample Firms across Regions

Regions	Frequency	Percent
Europe	19,230	49
Asia / Pacific	9,815	25
United States and Canada	8,384	21
Latin America and Caribbean	1,072	3
Africa / Middle East	919	2
Total	39,420	100

Notes: This table shows the distribution of firms in the 2016 ownership sample across five geographic regions. Note that only firms that have non-missing financial information and appear in the relationships database are included in this table. In the firms database, 39% of the 3,244,612 firms come from Europe, 31% come from United States and Canada, and 24% come from Asia/Pacific.

Table 5: Firm Characteristics by Organizational Forms

	Seller Integration Relationships			Buyer Integration Relationships			Outsourced Relationships		
	Buyer	Seller	Firm	Buyer	Seller	Firm	Buyer	Seller	Firm
R&D expenditure	5.346	81.76	55.86	109.0	7.721	75.00	18.68	24.07	15.33
	(38.69)	(520.7)	(427.8)	(630.9)	(51.04)	(518.7)	(225.8)	(257.0)	(203.1)
Total revenue	1530.4	5898.3	4297.2	7001.8	1613.2	5070.5	1632.9	1863.5	1364.7
	(10860.5)	(21288.6)	(17898.7)	(24989.1)	(11898.6)	(20978.4)	(8842.0)	(9924.9)	(7986.2)
Goodwill	56.84	598.5	414.6	658.7	49.95	452.7	181.4	221.8	150.1
	(465.5)	(2850.7)	(2358.3)	(2917.2)	(329.5)	(2406.2)	(1540.6)	(1735.4)	(1388.0)
R&D intensity	0.0124	0.0615	0.0454	0.0469	0.00984	0.0347	1.020	1.123	0.873
	(0.158)	(0.864)	(0.714)	(0.860)	(0.0810)	(0.705)	(74.50)	(82.32)	(67.03)
Goodwill intensity	0.0650	0.202	0.155	0.260	0.0429	0.187	0.216	0.541	0.349
	(0.598)	(2.297)	(1.911)	(3.428)	(0.232)	(2.807)	(46.38)	(47.73)	(53.53)
Age	40.11	56.28	50.63	57.48	42.58	52.07	46.60	45.87	45.37
	(34.82)	(47.40)	(44.17)	(46.37)	(37.49)	(43.96)	(43.13)	(43.03)	(42.38)
Total employment	4137.6	15049.1	11031.3	16549.8	4461.9	12246.9	4905.1	5548.4	4169.8
	(21526)	(45156)	(38348)	(47210)	(22532)	(40665)	(24958)	(27554)	(22607)
Observations	6, 699	3, 989	10, 524	3, 279	4, 984	8, 131	148, 875	101, 430	197, 742

Notes: The first three columns summarize the buyer, seller and firm characteristics for all buyer-integration relationships. The next and last three columns summarize the same characteristics for all seller-integration and outsourced relationships, respectively. Numbers in parentheses report standard deviations. Total revenue, R&D expenditure and goodwill are averaged over the period 2007-2014, measured in millions of U.S. dollars. Employment is the number of employees for the year 2015. R&D and goodwill intensities are the ratios between R&D expenditure and total revenue, and goodwill and total revenue.

Table 6: Buyer and Seller's Revenue Shares under Different Ownership Structures

	$k = BI$	$k = SI$	$k = NI$
Buyer's revenue share (β^k)	$\beta + (1-\beta)\delta^\alpha$	$\beta(1-\delta^\alpha)$	β
Seller's revenue share ($1-\beta^k$)	$(1-\beta)(1-\delta^\alpha)$	$1-\beta+\beta\delta^\alpha$	$1-\beta$

Table 7: Multinomial Logistic Regression Results

		(1)	(2)	(3)		(4)	(5)	(6)
Pr(y_i = BI)								
	rd_i	1.979***	1.956***	1.668***	gdw_i	2.155***	2.164***	2.046***
		(0.287)	(0.258)	(0.241)		(0.208)	(0.210)	(0.225)
	rd_i*itl	2.077***	1.895***	1.645***	gdw_i*itl	0.875***	0.776***	0.739***
		(0.243)	(0.210)	(0.193)		(0.174)	(0.172)	(0.175)
	rd_i*crind	0.560*	0.527**	0.551**	gdw_i*crind	0.537**	0.337	0.526*
		(0.228)	(0.197)	(0.197)		(0.195)	(0.190)	(0.205)
	itl	-1.758***	-1.638***	-1.641***	itl	-0.639***	-0.612***	-0.879***
		(0.225)	(0.174)	(0.179)		(0.162)	(0.136)	(0.169)
	crind	-1.964***	-1.738***	-1.784***	crind	-2.012***	-1.589***	-1.748***
		(0.200)	(0.179)	(0.167)		(0.183)	(0.172)	(0.177)
Pr(y_i = SI)								
	rd_i	-1.528***	-1.449***	-1.024***	gdw_i	-1.614***	-1.529***	-1.193***
		(0.189)	(0.168)	(0.109)		(0.140)	(0.135)	(0.124)
	rd_i*itl	-1.272***	-1.245***	-1.031***	gdw_i*itl	-0.608***	-0.585***	-0.563***
		(0.205)	(0.178)	(0.153)		(0.168)	(0.175)	(0.161)
	rd_i*crind	0.312	-0.071	-0.046	gdw_i*crind	0.410**	0.316*	0.251
		(0.204)	(0.159)	(0.129)		(0.159)	(0.144)	(0.131)
	itl	-0.204	-0.173	-0.453***	itl	-0.178	-0.162	-0.467***
		(0.154)	(0.129)	(0.093)		(0.172)	(0.143)	(0.093)
	crind	-1.764***	-1.511***	-1.501***	crind	-1.774***	-1.559***	-1.565***
		(0.114)	(0.068)	(0.074)		(0.125)	(0.070)	(0.076)
	Control variables	Y	Y	Y	Control variables	Y	Y	Y
	Pairwise sector FE	N	Y	Y	Pairwise sector FE	N	Y	Y
	Pairwise region FE	N	N	Y	Pairwise region FE	N	N	Y
	Observations	122765	122765	122765	Observations	106402	106402	106402
	pseudo R-squared	0.385	0.427	0.466	pseudo R-squared	0.373	0.424	0.478

Notes: This table summarizes the regression results for equation (22). rd_i and gdw_i are the buyer's relative R&D and goodwill intensities in firm-pair i, as defined in Section 2.3. *itl* is a dummy that equals one when the buyer and seller are located in the same country. *crind* is a dummy that equals one when the buyer and seller are operating in difference industries. The control variables include the logs of the buyer and seller's sales, age, and employment. There are a total of 100 pairwise sector fixed effects and 25 pairwise region fixed effects. All regressions are clustered at pairwise sector and region levels. Numbers in parentheses report standard errors.
* p<0.05, ** p<0.01, *** p<0.001.

Table 8: Average Marginal Effects of rd_i in column (3), Table 7

	$\Delta \Pr(y_i = BI)$	$\Delta \Pr(y_i = SI)$	$\Delta \Pr(y_i = NI)$
Marginal	0.176	-0.100	-0.077
	(0.000)	(0.000)	(0.000)
min->max	0.168	-0.095	-0.073
	(0.000)	(0.000)	(0.000)
+1 centered	0.175	-0.100	-0.075
	(0.000)	(0.000)	(0.000)
+SD centered	0.078	-0.044	-0.034
	(0.000)	(0.000)	(0.000)

Notes: This table calculates the average marginal effects of various changes in rd_i on pair i's probability of choose buyer integration, seller integration, and non-integration (Long and Freese, 2014). The four rows in turn report the effect of rd_i at its marginal value, when rd_i changes from the minimum (0) to the maximum (1), when rd_i changes from $rd_i - 0.5$ to $rd_i + 0.5$, when rd_i changes from $rd_i - 0.5SD$ to $rd_i + 0.5SD$, where $SD = .441$ is rd_i's standard deviation. The effects are calculated for each observation and then averaged across observations. Numbers in parentheses report p-values.

Table 9: Average Predicted Probabilities for Different Groups of Seller-Buyer Relationships

	Within country, within industry	Within country, cross indutry	Cross country, within industry	Cross country, cross industry
$\Pr(y_i\text{=BI})$	0.230	0.077	0.251	0.087
	(0.003)	(0.001)	(0.002)	(0.001)
$\Pr(y_i\text{=NI})$	0.541	0.847	0.497	0.827
	(0.004)	(0.002)	(0.003)	(0.001)
$\Pr(y_i\text{=SI})$	0.229	0.077	0.252	0.086
	(0.003)	(0.001)	(0.002)	(0.001)

Notes: This table reports the average predicted values of $\Pr(y_i = BI)$, $\Pr(y_i = BI)$, and $\Pr(y_i = BI)$ for four groups of firms based on whether they are located in the same country or the same industry. Parameters used for calculating these numbers are from a multinomial logistic regression similar to column (1) in Table 7, but excluding the control variables. Numbers in parentheses report standard errors.

Table 10: Linear Regression Results

	All			No buyer integration			No seller integration		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variables:	BI_i	SI_i	VI_i	BI_i	SI_i	VI_i	BI_i	SI_i	VI_i
rd_i	0.250***	-0.178***	0.072***	N/A	-0.027*	-0.027*	0.035***	N/A	0.035***
	(0.004)	(0.004)	(0.004)	(.)	(0.011)	(0.011)	(0.008)	(.)	(0.008)
Control variables	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise country FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	118747	118747	118747	5943	5943	5943	9122	9122	9122
R-squared	0.478	0.459	0.465	.	0.616	0.616	0.469	.	0.469

(a) Buyer's relative R&D intensity

	All			No buyer integration			No seller integration		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variables:	BI_i	SI_i	VI_i	BI_i	SI_i	VI_i	BI_i	SI_i	VI_i
gdw_i	0.223***	-0.148***	0.075***	N/A	-0.017**	-0.017**	0.048***	N/A	0.048***
	(0.009)	(0.005)	(0.005)	(.)	(0.007)	(0.007)	(0.006)	(.)	(0.006)
Control variables	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise country FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	153493	153493	153493	7709	7709	7709	11013	11013	11013
R-squared	0.501	0.489	0.553	.	0.574	0.574	0.459	.	0.459

(b) Buyer's relative goodwill intensity

Notes: Panel (a) uses rd_i as the variable of interest. Panel (b) uses gdw_i as the variable of interests. The dependent variables BI_i, SI_i and VI_i are dummies for buyer integration, seller integration, and integration. Each panel runs linear regressions for three types of industry pairs: all industry pairs, industry pairs with no buyer integration relationships, and industry pairs with no seller integration relationships. The control variables include logs of the buyer and seller's sales, employment and age. All regressions are clustered at pairwise industry and country levels. Numbers in parentheses report standard errors. Since each group of regressions contain different numbers of fixed effects, I report the number of fixed effects and level of clustering for columns (1)-(3) in each panel.
Panel (a): 9,856 pairwise industry fixed effects, 2,690 pairwise country fixed effects, clustered at 67,794 pairwise industry-country fixed effects. Panel (b): 11,498 pairwise industry fixed effects, 3522 pairwise country fixed effects, clustered at 84,277 pairwise industry-country fixed effects.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 11: IV Regression Results

	Industry-country				Industry-country-ownership				Weighted industry-country-ownerhip			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	rd_i	BI_i	SI_i	VI_i	rd_i	BI_i	SI_i	VI_i	rd_i	BI_i	SI_i	VI_i
Instrumental $variable_i$	0.838***				0.884***				0.837***			
	(0.013)				(0.007)				(0.012)			
rd_i		0.181***	-0.058***	0.123***		0.270***	-0.140***	0.130***		0.183***	-0.061***	0.123***
		(0.018)	(0.009)	(0.017)		(0.030)	(0.021)	(0.016)		(0.018)	(0.009)	(0.017)
Control variables	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise sector FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise region FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	114964	114964	114964	114964	114964	114964	114964	114964	114964	114964	114964	114964
F statistic	6921.988	53.003	19.831	27.471	9823.049	50.494	18.780	28.481	6848.295	52.127	20.011	27.293
Sargan's J test		0.000	0.000	0.000		0.000	0.000	0.000		0.000	0.000	0.000

(a) R&D Intensity

	Industry-country				Industry-country-ownership				Weighted industry-country-ownerhip			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	gdw_i	BI_i	SI_i	VI_i	gdw_i	BI_i	SI_i	VI_i	gdw_i	BI_i	SI_i	VI_i
Instrumental $variable_i$	0.155				0.756***				0.683***			
	(0.118)				(0.012)				(0.019)			
gdw_i		0.211***	-0.043**	0.168***		0.396***	-0.225***	0.171***		0.224***	-0.056***	0.168***
		(0.024)	(0.014)	(0.023)		(0.048)	(0.037)	(0.020)		(0.023)	(0.012)	(0.023)
Control variables	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise sector FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pairwise region FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	145328	145328	145328	145328	145328	145328	145328	145328	145328	145328	145328	145328
F statistic	139.779	55.245	22.241	34.879	2889.483	63.858	22.818	36.967	1152.085	56.050	23.098	35.376
Sargan's J test		0.000	0.000	0.000		0.000	0.000	0.000		0.000	0.000	0.000

(b) Goodwill Intensity

This table reports the IV regression results for rd_i and gdw_i. There are three instrumental variables for each variable: industry IV, industry-ownership IV, and weighted industry IV. The industry IV is calculated by dividing the buyer's industry-level R&D/goodwill intensity by the sum of the buyer and seller's industry-level R&D/goodwill intensities. The industry-ownership IV is calculated in a similar manner, only replacing the industry level R&D/goodwill intensity with industry-ownership level R&D/goodwill intensity, where the R&D/goodwill intensities are calculated for each type of ownership structure within the industry. The weighted industry R&D intensity calculates the R&D/goodwill intensity for each industry by replace the industry-level R&D/goodwill intensities by the number of ownership structures in each industry pair. There are 100 pairwise sector fixed effects and 25 pairwise region fixed effects. All regressions are clustered at pairwise sector and region levels. Numbers in parentheses report standard errors. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 12: Parent and Relationship sources of the Ex-post Ownership Sample

Relationship Source	Parent Source			Total
	2014	2014 2016	2016	
2013	52	239	692	983
2014	30	193	638	861
2016	0	0	655	655
Total	82	432	1,985	2,499

Notes: This table tabulates the sources of seller-buyer relationships and firm-parent relationships in the ex-post ownership sample–seller-buyer relationships that later become integrated.

Table 13: Multinomial Logistic Regression Results for Seller-buyer Relations with Ex-post Ownership

		(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
$Pr(y_i = BI)$										
	rd_i	1.236***	0.878**	1.413***	0.993*	gdw_i	1.802***	1.803***	1.873***	1.896***
		(0.223)	(0.268)	(0.360)	(0.408)		(0.289)	(0.366)	(0.461)	(0.533)
	rd_i*itl		1.001*		0.966	gdw_i*itl		-0.091		-0.101
			(0.502)		(0.506)			(0.602)		(0.606)
	rd_i*crind			-0.326	-0.192	gdw_i*crind			-0.154	-0.176
				(0.458)	(0.467)				(0.591)	(0.596)
	itl		-1.082**		-1.028*	itl		-0.217		-0.189
			(0.417)		(0.421)			(0.518)		(0.520)
	crind			0.505	0.371	crind			0.362	0.353
				(0.366)	(0.373)				(0.515)	(0.518)
$Pr(y_i = SI)$										
	rd_i	-1.396***	-1.502***	-1.107**	-1.192**	gdw_i	-1.496***	-1.804***	-1.879***	-2.315***
		(0.220)	(0.261)	(0.355)	(0.399)		(0.238)	(0.284)	(0.470)	(0.517)
	rd_i*itl		-0.047		-0.126	gdw_i*itl		0.897		0.936
			(0.519)		(0.521)			(0.522)		(0.526)
	rd_i*crind			-0.497	-0.467	gdw_i*crind			0.504	0.686
				(0.452)	(0.462)				(0.545)	(0.555)
	itl		-0.653**		-0.569**	itl		-1.030***		-0.996***
			(0.202)		(0.205)			(0.243)		(0.245)
	crind			0.640**	0.556**	crind			0.464*	0.339
				(0.201)	(0.204)				(0.218)	(0.222)
	Observations	2110	2110	2110	2110	Observations	1713	1713	1713	1713
	Pseudo R-sq	0.042	0.052	0.048	0.056	Pseudo R-sq	0.062	0.075	0.068	0.080

Notes: This table reports the multinomial logistic regression results for the seller-buyer relationships that later became integrated. Numbers in parentheses report standard errors. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 14: Linear Regression Results for Tertiles of rd_i and $(1 - rd_i)$

Dependent variables	BI_i	SI_i	NI_i
	(1)	(2)	(3)
$prod_i*rd_1$	-0.096***	0.048***	0.049***
	(0.003)	(0.002)	(0.003)
$prod_i*rd_2$	0.016***	-0.024***	0.008***
	(0.001)	(0.002)	(0.002)
$prod_i*rd_3$	0.038***	-0.071***	0.033***
	(0.002)	(0.002)	(0.002)
$\ln(age_b)$	0.034***	-0.065***	0.031***
	(0.002)	(0.002)	(0.002)
$\ln(age_s)$	-0.062***	0.053***	0.009***
	(0.002)	(0.002)	(0.002)
Pairwise industry FE	Y	Y	Y
Pairwise country FE	Y	Y	Y
Observations	118747	118747	118747
R-squared	0.425	0.401	0.462

(a) Buyer's relative R&D intensity

Dependent variables	BI_i	SI_i	NI_i
	(1)	(2)	(3)
$prod_i*(1\text{-}rd)_1$	0.055***	-0.104***	0.049***
	(0.002)	(0.002)	(0.002)
$prod_i*(1\text{-}rd)_2$	0.002	0.004**	-0.006**
	(0.001)	(0.001)	(0.002)
$prod_i*(1\text{-}rd)_3$	-0.093***	0.040***	0.052***
	(0.003)	(0.002)	(0.003)
$\ln(age_b)$	0.030***	-0.057***	0.027***
	(0.002)	(0.002)	(0.002)
$\ln(age_s)$	-0.061***	0.051***	0.010***
	(0.002)	(0.002)	(0.002)
Pairwise industry FE	Y	Y	Y
Pairwise country FE	Y	Y	Y
Observations	118747	118747	118747
R-squared	0.430	0.419	0.465

(b) Seller's relative R&D intensity

Notes: Panels (a) and (b) report the linear regression results as specified in equations (B), (S) and (N) for tertiles of rd_i and $(1 - rd_i)$. rd_q is the qth tertile of rd_i and $(1 - rd)_q$ is the qth tertile of $(1 - rd_i)$. $\ln(\text{age})_b$ and $\ln(\text{age})_s$ are the log of the buyer and seller's ages, respectively. There are 9,856 pairwise industry fixed effects and 2,690 pairwise country fixed effects. All regressions are clustered at pairwise industry and country levels.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 15: Multinomial Logistic Regression Results for Tertiles of rd_i and $(1 - rd_i)$

		(1)	(2)	(3)		(4)	(5)	(6)
$Pr(y_i = BI)$								
	$prod_i*rd_1$	-0.760***	-0.742***	-0.604***	$prod_i*(1\text{-}rd)_1$	0.464***	0.398***	0.449***
		(0.011)	(0.013)	(0.071)		(0.016)	(0.018)	(0.056)
	$prod_i*rd_2$	0.202***	0.155***	0.074	$prod_i*(1\text{-}rd)_2$	0.374***	0.302***	0.276***
		(0.012)	(0.014)	(0.050)		(0.014)	(0.015)	(0.061)
	$prod_i*rd_3$	1.950***	1.792***	1.967***	$prod_i*(1\text{-}rd)_3$	-0.895***	-0.857***	-0.744***
		(0.050)	(0.051)	(0.146)		(0.011)	(0.013)	(0.086)
	$ln(age_b)$	0.431***	0.337***	0.253***	$ln(age_b)$	0.474***	0.402***	0.342***
		(0.013)	(0.014)	(0.041)		(0.012)	(0.014)	(0.042)
	$ln(age_s)$	-0.430***	-0.567***	-0.563***	$ln(age_s)$	-0.450***	-0.589***	-0.592***
$Pr(y_i = SI)$								
	$prod_i*rd_1$	0.394***	0.382***	0.388***	$prod_i*(1\text{-}rd)_1$	-0.735***	-0.929***	-0.668***
		(0.017)	(0.017)	(0.046)		(0.016)	(0.018)	(0.106)
	$prod_i*rd_2$	0.221***	0.199***	0.098	$prod_i*(1\text{-}rd)_2$	-0.223***	-0.325***	-0.334***
		(0.014)	(0.016)	(0.067)		(0.012)	(0.013)	(0.064)
	$prod_i*rd_3$	-0.682***	-0.868***	-0.703***	$prod_i*(1\text{-}rd)_3$	0.404***	0.377***	0.410***
		(0.011)	(0.013)	(0.061)		(0.017)	(0.017)	(0.046)
	$ln(age_b)$	-0.720***	-0.771***	-0.759***	$ln(age_b)$	-0.748***	-0.821***	-0.804***
		(0.013)	(0.014)	(0.047)		(0.013)	(0.014)	(0.048)
	$ln(age_s)$	0.902***	0.680***	0.626***	$ln(age_s)$	0.855***	0.635***	0.603***
		(0.013)	(0.014)	(0.041)		(0.013)	(0.014)	(0.044)
Pairwise sector FE		N	Y	Y		N	Y	Y
Pairwise region FE		N	N	Y		N	N	Y
Observations		122765	122765	122765		122765	122765	122765
Pseudo R-squared		0.185	122765	0.338		0.160	0.246	0.317

Notes: Columns (1)-(3) report the multinomial logistic regression results for tertiles of rd_1; columns (4)-(6) report the results for tertiles of $(1 - rd_i)$. rd_q is the qth tertile of rd_i and $(1 - rd)_q$ is the qth tertile of $(1 - rd_i)$. $\ln(\text{age})_b$ and $\ln(\text{age})_s$ are the log of the buyer and seller's ages, respectively. There are 100 pairwise sector fixed effects and 25 pairwise region fixed effects. All regressions are clustered at pairwise sector- and region- levels. Numbers in parentheses report standard errors. * p<0.05, ** p<0.01, *** p<0.001.

Table 16: Average Marginal Effects of Pair-level Productivity

		rd_q			$(1\text{-}rd)_q$		
		$\Delta Pr(y_i = BI)$	$\Delta Pr(y_i = SI)$	$\Delta Pr(y_i = NI)$	$\Delta Pr(y_i = BI)$	$\Delta Pr(y_i = SI)$	$\Delta Pr(y_i = NI)$
q=1	Marginal	-0.072	0.025	0.048	0.029	-0.068	0.039
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	+SD	-0.068	0.030	0.038	0.037	-0.063	0.026
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	min->max	-0.946	0.612	0.334	0.755	-0.928	0.173
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
q=2	Marginal	0.005	0.005	-0.010	0.019	-0.030	0.010
		(0.158)	(0.149)	(0.083)	(0.000)	(0.000)	(0.080)
	+SD	0.006	0.006	-0.011	0.023	-0.029	0.006
		(0.167)	(0.162)	(0.089)	(0.000)	(0.000)	(0.280)
	min->max	0.104	0.110	-0.214	0.473	-0.567	0.095
		(0.169)	(0.178)	(0.086)	(0.000)	(0.000)	(0.526)
q=3	Marginal	0.069	-0.076	0.007	-0.084	0.028	0.056
		(0.000)	(0.000)	(0.296)	(0.000)	(0.000)	(0.000)
	+SD	0.126	-0.074	-0.052	-0.075	0.034	0.042
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	min->max	0.997	-0.941	-0.056	-0.977	0.649	0.328
		(0.000)	(0.000)	(0.014)	(0.000)	(0.000)	(0.000)

Notes: The parameters used for calculations come from column (3) and column (6) in Table 15. The first three columns calculate the average marginal effects of $prod_i$ on $\Pr(y_i = BI)$, $\Pr(y_i = BI)$ and $\Pr(y_i = BI)$ at different tertiles of rd_i. The next three columns calculate the same average marginal effects of $prod_i$ at different tertiles of $(1 - rd_i)$. For each tertile I calculate the marginal effect, the one standard deviation increase, and the min to max change of $prod_i$ on the probability of pair i choosing various ownership structures. Numbers in parentheses report p-values.

Table E.1: Sector-pairs Hosting more than 1000 Seller-buyer Relationships

Seller's sector	Buyer's sector	Total number of relationships	Buyer integration	Seller integration	Non-integration
Information Technology	Information Technology	19091	1754	1769	15568
Industrials	Industrials	12669	2815	2824	7030
Consumer Discretionary	Consumer Discretionary	8273	1177	1153	5943
Information Technology	Consumer Discretionary	7433	193	248	6992
Information Technology	Industrials	6455	446	339	5670
Healthcare	Healthcare	6204	1049	1046	4109
Materials	Materials	4221	1145	1144	1932
Information Technology	Financials	4191	38	45	4108
Industrials	Consumer Discretionary	3834	327	543	2964
Information Technology	Telecommunication Services	3632	90	21	3521
Industrials	Materials	3015	382	342	2291
Industrials	Information Technology	2773	338	440	1995
Energy	Energy	2564	378	380	1806
Consumer Discretionary	Industrials	2562	529	295	1738
Industrials	Utilities	2228	57	38	2133
Consumer Staples	Consumer Staples	2178	594	593	991
Materials	Industrials	2066	344	391	1331
Industrials	Energy	1643	124	53	1466
Consumer Discretionary	Information Technology	1589	244	170	1175
Information Technology	Healthcare	1522	49	38	1435
Utilities	Utilities	1299	287	285	727
Information Technology	Consumer Staples	1209	7	1	1201
Materials	Consumer Discretionary	1201	71	231	899
Information Technology	Utilities	1033	10	5	1018

Table E.2: Multinomial Logistic Regression Results

	y_i=BI	y_i=SI
Buyer's relative R&D intensity	3.927***	-1.714***
	(0.0642)	(0.0653)
Buyer's relative R&D expenditure	-0.809***	0.378***
	(0.0741)	(0.0790)
Buyer's relative total revenue	0.244***	-3.336***
	(0.0665)	(0.0784)
Buyer's relative employment	2.844***	-1.608***
	(0.0764)	(0.0786)
Buyer's relative age	0.352***	-1.177***
	(0.0565)	(0.0564)
Observations	124971	
Pseudo R-sq	0.330	

Table E.3: Logit Regression Results for the Extended Sample

	Dependent variable: BI_i				Dependent variable: SI_i				Dependent variable: VI_i			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
rd_i	4.031***	3.390***	3.268***	2.909***	-2.866***	-1.809***	-2.105***	-1.671***	0.603***	0.993***	0.765***	0.857***
	(0.263)	(0.211)	(0.201)	(0.199)	(0.258)	(0.177)	(0.159)	(0.135)	(0.101)	(0.107)	(0.104)	(0.107)
$ln(tr_b)$		-0.124**	-0.121**	-0.092*		-0.153***	-0.161***	-0.208***		-0.073*	-0.085**	-0.104***
		(0.046)	(0.038)	(0.038)		(0.033)	(0.032)	(0.032)		(0.031)	(0.028)	(0.026)
$ln(tr_s)$		0.068**	0.050*	0.029		0.180**	0.116*	0.171***		0.237***	0.202***	0.216***
		(0.022)	(0.021)	(0.019)		(0.057)	(0.049)	(0.049)		(0.026)	(0.022)	(0.022)
$ln(emp_b)$		0.150**	0.172***	0.154***		-0.302***	-0.299***	-0.242***		-0.164***	-0.149***	-0.111***
		(0.047)	(0.034)	(0.036)		(0.019)	(0.018)	(0.016)		(0.020)	(0.016)	(0.015)
$ln(emp_s)$		-0.433***	-0.438***	-0.384***		0.033	0.126***	0.114**		-0.287***	-0.261***	-0.242***
		(0.019)	(0.020)	(0.019)		(0.050)	(0.037)	(0.037)		(0.023)	(0.019)	(0.019)
$ln(age_b)$		0.230***	0.130**	0.091*		-0.107*	-0.200***	-0.226***		0.133*	0.027	-0.028
		(0.063)	(0.044)	(0.045)		(0.049)	(0.040)	(0.040)		(0.054)	(0.035)	(0.038)
$ln(age_s)$		-0.059	-0.163***	-0.198***		0.457***	0.184***	0.137***		0.271***	0.068*	-0.002
		(0.040)	(0.037)	(0.037)		(0.066)	(0.036)	(0.031)		(0.050)	(0.030)	(0.028)
Pairwise sector FE	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Pairwise region FE	N	N	N	Y	N	N	N	Y	N	N	N	Y
Observations	122345	122345	122345	122345	122345	122345	121981	121971	122345	122345	122345	122345
Pseudo R-sq	0.261	0.348	0.400	0.439	0.120	0.330	0.412	0.452	0.012	0.087	0.179	0.255

There are 89 pairwise sector fixed effects and 24 pairwise region fixed effects. All regressions are clustered at the pairwise sector- and region- levels. Numbers in parentheses report standard errors.* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table E.4: Multinomial Logit Regression Results for the Original Sample

		(1)	(2)	(3)		(4)	(5)	(6)
$Pr(y_i = BI)$								
	rd_i	1.727***	1.700***	1.598***	gdw_i	1.900***	1.731***	1.629***
		(0.339)	(0.318)	(0.298)		(0.395)	(0.369)	(0.363)
	rd_i*itl	1.134***	1.022***	1.017***	gdw_i*itl	-0.919***	-0.656**	-0.558**
		(0.310)	(0.291)	(0.275)		(0.615)	(0.613)	(0.642)
	rd_i*crind	-0.0599	0.0246	-0.0265	gdw_i*crind	0.649**	0.811***	0.860***
		(0.354)	(0.340)	(0.316)		(0.566)	(0.525)	(0.503)
	itl	-1.800***	-1.697***	-1.677***	itl	-0.254	-0.437*	-0.450*
		(0.283)	(0.255)	(0.236)		(0.469)	(0.478)	(0.488)
	crind	-1.307***	-1.039***	-1.003***	crind	-1.986***	-1.776***	-1.811***
		(0.319)	(0.278)	(0.257)		(0.454)	(0.419)	(0.397)
$Pr(y_i = SI)$								
	rd_i	-1.493***	-1.417***	-1.118***	gdw_i	-1.270***	-1.156***	-1.029***
		(0.261)	(0.250)	(0.203)		(0.208)	(0.207)	(0.183)
	rd_i*itl	-0.839**	-0.908**	-0.794**	gdw_i*itl	-0.110	-0.165	-0.153
		(0.288)	(0.270)	(0.262)		(0.307)	(0.299)	(0.250)
	rd_i*crind	0.885***	0.534*	0.452	gdw_i*crind	0.466*	0.360*	0.320
		(0.267)	(0.252)	(0.219)		(0.343)	(0.337)	(0.257)
	itl	-0.785***	-0.706***	-0.826***	itl	-1.069***	-0.985***	-1.012***
		(0.194)	(0.165)	(0.121)		(0.190)	(0.160)	(0.124)
	crind	-1.827***	-1.524***	-1.478***	crind	-1.604***	-1.335***	-1.327***
		(0.154)	(0.119)	(0.121)		(0.154)	(0.119)	(0.107)
	Control variables	Y	Y	Y	Control variables	Y	Y	Y
	Pairwise sector FE	N	Y	Y	Pairwise sector FE	N	Y	Y
	Pairwise region FE	N	N	Y	Pairwise region FE	N	N	Y
	Observations	95346	95346	95346	Observations	107560	107560	107560
	pseudo R-squared	0.161	0.208	0.257	pseudo R-squared	0.164	0.215	0.266

Notes: rd_i and gdw_i are the buyer's relative R&D and goodwill intensities in firm-pair i, as defined in Section 2.3. *itl* is a dummy that equals one when the buyer and the seller are located in the same country. *crind* is a dummy that equals one when the buyer and the seller are in difference industries. The controls variables include the logs of the buyer and seller's sales, age, and employment. There are a total of 100 pairwise sector fixed effects and 25 pairwise region fixed effects. All regressions are clustered at ?? pairwise sector and region fixed effects. Numbers in parentheses report standard errors. * p<0.05, ** p<0.01, *** p<0.001.

Table E.5: Linear Regression Results for the Original Sample

	Buyer's relative R&D intensity				Buyer's relative goodwill intensity		
	(1)	(2)	(3)		(4)	(5)	(6)
Dependent variables:	BI_i	SI_i	VI_i		BI_i	SI_i	VI_i
rd_i	0.016***	-0.014***	0.002	gdw_i	0.013***	-0.009***	0.004**
	(0.001)	(0.001)	(0.002)		(0.001)	(0.001)	(0.001)
Control variables	Y	Y	Y	Control variables	Y	Y	Y
Pairwise industry FE	Y	Y	Y	Pairwise industry FE	Y	Y	Y
Pairwise country FE	Y	Y	Y	Pairwise country FE	Y	Y	Y
Observations	95346	95346	95346	Observations	77765	77765	77765
R-squared	0.167	0.204	0.212	R-squared	0.152	0.180	0.199

Control variables include logs of the buyer and seller's respective sales, employment and age. There are 9,151 pairwise industry-level fixed effects and 2,553 pairwise country-level fixed effects for columns (1)-(3). There are 10,518 pairwise industry fixed effects and 3,225 pairwise country fixed effects. All regressions are clustered at pairwise industry- and country- levels. Numbers in parentheses report standard errors.* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table E.6: Logit Regression Results for the Original Sample

	Dependent variable: BI_i				Dependent variable: SI_i				Dependent variable: VI_i			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
rd_i	2.401***	2.221***	2.077***	1.949***	-1.497***	-0.895***	-1.220***	-1.002***	0.437***	0.665***	0.418***	0.495***
	(0.200)	(0.202)	(0.212)	(0.216)	(0.214)	(0.194)	(0.180)	(0.168)	(0.107)	(0.105)	(0.090)	(0.093)
$\ln(tr_b)$		0.056	0.039	0.038		-0.013	-0.050	-0.127***		0.050	0.014	-0.036
		(0.054)	(0.053)	(0.051)		(0.036)	(0.032)	(0.034)		(0.035)	(0.029)	(0.031)
$\ln(tr_s)$		0.105*	0.031	0.006		0.385***	0.293***	0.334***		0.281***	0.187***	0.204***
		(0.042)	(0.038)	(0.040)		(0.059)	(0.049)	(0.038)		(0.040)	(0.031)	(0.029)
$\ln(emp_b)$		-0.091	-0.057	-0.045		-0.225***	-0.206***	-0.150***		-0.190***	-0.158***	-0.114***
		(0.049)	(0.050)	(0.047)		(0.035)	(0.034)	(0.033)		(0.028)	(0.026)	(0.025)
$\ln(emp_s)$		-0.242***	-0.190***	-0.159***		-0.258***	-0.150**	-0.137***		-0.257***	-0.176***	-0.159***
		(0.040)	(0.044)	(0.041)		(0.049)	(0.046)	(0.041)		(0.032)	(0.032)	(0.031)
$\ln(age_b)$		0.169	0.119	0.130		-0.283***	-0.336***	-0.339***		-0.049	-0.106	-0.104
		(0.096)	(0.087)	(0.080)		(0.078)	(0.072)	(0.072)		(0.069)	(0.062)	(0.060)
$\ln(age_s)$		-0.113	-0.136	-0.181*		0.327**	0.107	0.091		0.142	0.016	-0.022
		(0.085)	(0.084)	(0.086)		(0.114)	(0.076)	(0.069)		(0.076)	(0.053)	(0.053)
Pairwise sector FE	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Pairwise region FE	N	N	N	Y	N	N	N	Y	N	N	N	Y
Observations	87278	87278	87278	87278	87278	87278	82205	81812	87278	87278	87278	87278
Pseudo R-sq	0.076	0.093	0.140	0.184	0.024	0.117	0.184	0.240	0.003	0.038	0.099	0.154

There are 59 pairwise sector fixed effects and 16 pairwise region fixed effects. All regressions are clustered at the pairwise sector- and region- levels. Numbers in parentheses report standard errors.* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table E.7: Linear Regression Results for Quintiles of Buyer and Seller's Relative R&D Intensities

Dependent variables	BI_i (1)	SI_i (2)	NI_i (3)
$prod_i$*rd_1	-0.189***	0.073***	0.116***
	(0.004)	(0.002)	(0.003)
$prod_i$*rd_2	0.002	0.020***	-0.022***
	(0.002)	(0.002)	(0.003)
$prod_i$*rd_3	0.009***	-0.009***	-0.000
	(0.001)	(0.002)	(0.002)
$prod_i$*rd_4	0.030***	-0.086***	0.055***
	(0.001)	(0.002)	(0.002)
$prod_i$*rd_5	0.016***	-0.039***	0.023***
	(0.002)	(0.003)	(0.003)
ln(age_b)	0.027***	-0.060***	0.032***
	(0.001)	(0.002)	(0.002)
ln(age_s)	-0.052***	0.050***	0.002
	(0.002)	(0.002)	(0.002)
Pairwise industry FE	Y	Y	Y
Pairwise country FE	Y	Y	Y
Observations	118747	118747	118747
R-squared	0.467	0.415	0.477

(a) Buyer's relative R&D intensity

Dependent variables	BI_i (1)	SI_i (2)	NI_i (3)
$prod_i$*$(1\text{-}rd)_1$	0.043***	-0.090***	0.046***
	(0.002)	(0.002)	(0.002)
$prod_i$*$(1\text{-}rd)_2$	0.051***	-0.087***	0.036***
	(0.003)	(0.004)	(0.004)
$prod_i$*$(1\text{-}rd)_3$	0.007***	-0.001	-0.007***
	(0.001)	(0.002)	(0.002)
$prod_i$*$(1\text{-}rd)_4$	-0.062***	0.036***	0.026***
	(0.002)	(0.002)	(0.003)
$prod_i$*$(1\text{-}rd)_5$	-0.125***	0.053***	0.072***
	(0.004)	(0.002)	(0.004)
ln(age_b)	0.031***	-0.059***	0.029***
	(0.002)	(0.002)	(0.002)
ln(age_s)	-0.062***	0.052***	0.010***
	(0.002)	(0.002)	(0.002)
Pairwise industry FE	Y	Y	Y
Pairwise country FE	Y	Y	Y
Observations	118747	118747	118747
R-squared	0.427	0.415	0.464

(b) Seller's relative R&D intensity

Table E.8: Linear Regression Results for Septiles of Buyer and Seller's Relative R&D Intensities

Dependent variables	BI_i	SI_i	NI_i
	(1)	(2)	(3)
$prod_i*rd_1$	-0.167***	0.061***	0.106***
	(0.003)	(0.002)	(0.003)
$prod_i*rd_2$	-0.014***	0.020***	-0.006
	(0.003)	(0.002)	(0.003)
$prod_i*rd_3$	-0.006*	0.024***	-0.018***
	(0.002)	(0.002)	(0.003)
$prod_i*rd_4$	0.018***	-0.007***	-0.011***
	(0.002)	(0.002)	(0.003)
$prod_i*rd_5$	0.031***	-0.077***	0.046***
	(0.001)	(0.002)	(0.002)
$prod_i*rd_6$	0.061***	-0.112***	0.051***
	(0.002)	(0.003)	(0.003)
$prod_i*rd_7$	-0.002	-0.012***	0.014***
	(0.002)	(0.003)	(0.003)
$ln(age_b)$	0.029***	-0.059***	0.030***
	(0.002)	(0.002)	(0.002)
$ln(age_s)$	-0.054***	0.050***	0.005*
	(0.002)	(0.002)	(0.002)
Pairwise industry FE	Y	Y	Y
Pairwise country FE	Y	Y	Y
Observations	118747	118747	118747
R-squared	0.456	0.415	0.474

(a) Buyer's relative R&D intensity

Dependent variables	BI_i	SI_i	NI_i
	(1)	(2)	(3)
$prod_i*(1-rd)_1$	0.044***	-0.092***	0.048***
	(0.002)	(0.002)	(0.002)
$prod_i*(1-rd)_2$	0.080***	-0.125***	0.045***
	(0.002)	(0.004)	(0.004)
$prod_i*(1-rd)_3$	-0.008***	-0.000	0.008**
	(0.002)	(0.003)	(0.003)
$prod_i*(1-rd)_4$	0.022***	-0.009***	-0.013***
	(0.002)	(0.002)	(0.003)
$prod_i*(1-rd)_5$	-0.002	0.021***	-0.019***
	(0.002)	(0.002)	(0.003)
$prod_i*(1-rd)_6$	-0.121***	0.047***	0.074***
	(0.003)	(0.002)	(0.003)
$prod_i*(1-rd)_7$	-0.074***	0.032***	0.042***
	(0.005)	(0.003)	(0.004)
$ln(age_b)$	0.030***	-0.058***	0.028***
	(0.002)	(0.002)	(0.002)
$ln(age_s)$	-0.057***	0.050***	0.007***
	(0.002)	(0.002)	(0.002)
Pairwise industry FE	Y	Y	Y
Pairwise country FE	Y	Y	Y
Observations	118747	118747	118747
R-squared	0.439	0.418	0.469

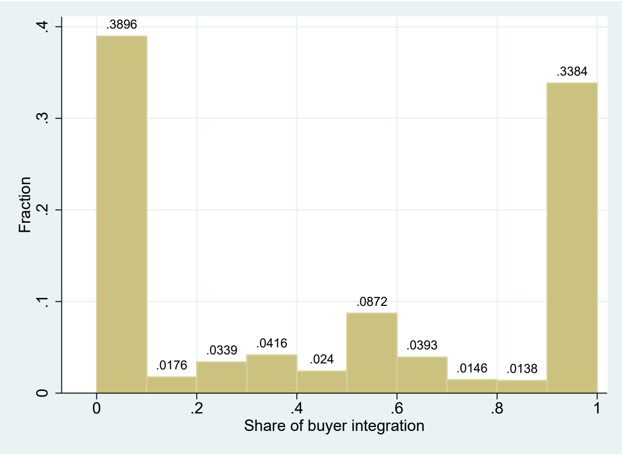
(b) Seller's relative R&D intensity

Table E.9: Average Marginal Effects of Pair-level Productivity with Quintiles of rd_i and $(1-rd_i)$

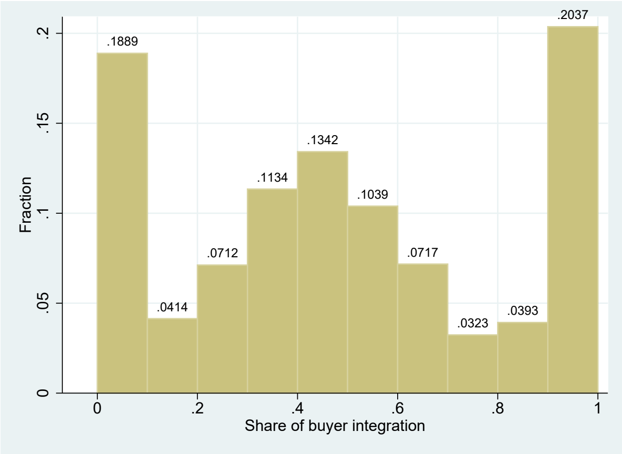
	rd_q			$(1\text{-}rd)_q$		
	$\Delta Pr(y_i = BI)$	$\Delta Pr(y_i = SI)$	$\Delta Pr(y_i = NI)$	$\Delta Pr(y_i = BI)$	$\Delta Pr(y_i = SI)$	$\Delta Pr(y_i = NI)$
q=1	-0.124	0.029	0.095	0.014	-0.021	0.008
	(0.000)	(0.000)	(0.000)	(0.000)	(0.041)	(0.327)
q=2	-0.058	0.026	0.032	0.048	-0.066	0.018
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)
q=3	0.010	0.000	-0.010	0.017	-0.009	-0.007
	(0.000)	(0.911)	(0.045)	(0.000)	(0.017)	(0.151)
q=4	0.033	-0.052	0.019	0.011	0.012	-0.023
	(0.000)	(0.001)	(0.185)	(0.006)	(0.000)	(0.000)
q=5	0.051	-0.067	0.015	-0.133	0.055	0.078
	(0.000)	(0.000)	(0.041)	(0.000)	(0.000)	(0.000)

Table E.10: Average Marginal Effects of Pair-level Productivity with Quintiles of rd_i and $(1 - rd_i)$

	rd_q			$(1\text{-}rd)_q$		
	$\Delta Pr(y_i = BI)$	$\Delta Pr(y_i = SI)$	$\Delta Pr(y_i = NI)$	$\Delta Pr(y_i = BI)$	$\Delta Pr(y_i = SI)$	$\Delta Pr(y_i = NI)$
q=1	-0.060	0.016	0.043	-0.005	0.002	0.003
	0.001	(0.000)	0.004	0.393	0.804	0.418
q=2	-0.087	0.032	0.055	0.067	-0.083	0.016
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	0.015
q=3	-0.002	0.011	-0.009	0.045	-0.059	0.014
	0.838	0.002	0.361	(0.000)	(0.000)	0.010
q=4	0.027	-0.012	-0.015	0.024	-0.009	-0.015
	(0.000)	0.041	0.043	(0.000)	0.146	0.041
q=5	-0.011	0.012	-0.001	-0.005	0.014	-0.009
	0.043	0.001	0.882	0.502	(0.000)	0.374
q=6	0.074	-0.096	0.022	0.003	0.012	-0.015
	(0.000)	(0.000)	0.007	0.503	(0.000)	(0.000)
q=7	0.058	-0.069	0.011	-0.114	0.044	0.070
	(0.000)	(0.000)	0.081	(0.000)	(0.000)	(0.000)



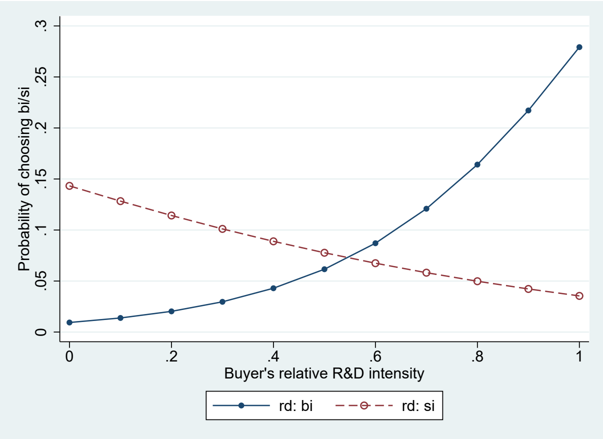
(a) Unweighted



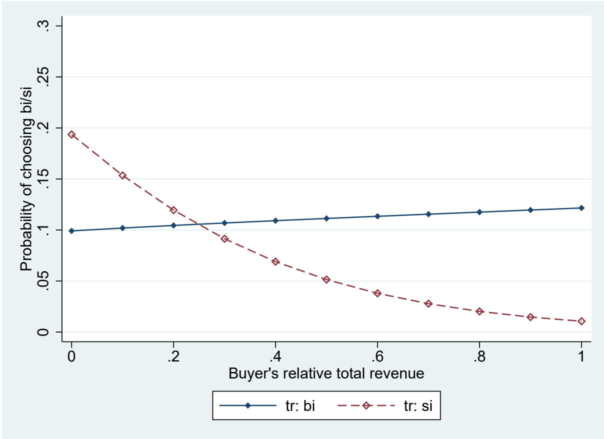
(b) Weighted

Figure 1: Histograms of Industry Pairs by Their Shares of Buyer Integration

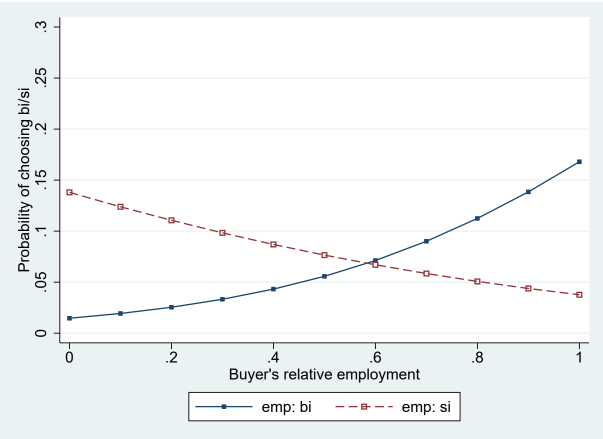
Notes: This figure contains histograms of industry pairs by their shares of buyer integration relative to all integrated relationships. There are a total of 18,575 industry pairs, 4,713 of them contain integrated relationships. Panel (a) is the unweighted histogram of industry pairs. Panel (b) is the weighted histogram of industry pairs where industry pairs are weighted by their total number of integrated relationships. Numbers over the bars are fractions.



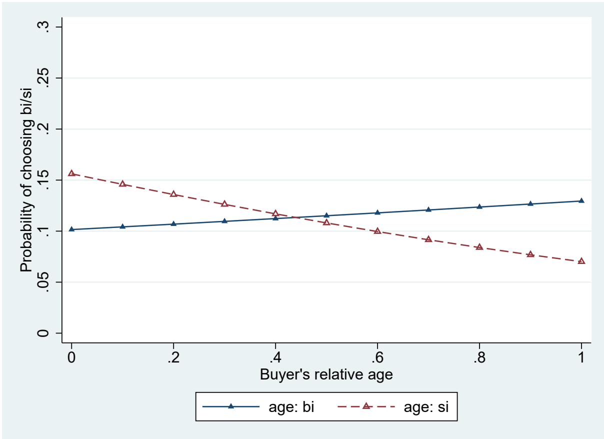
(a) Buyer's relative R&D intensity



(b) Buyer's relative total revenue

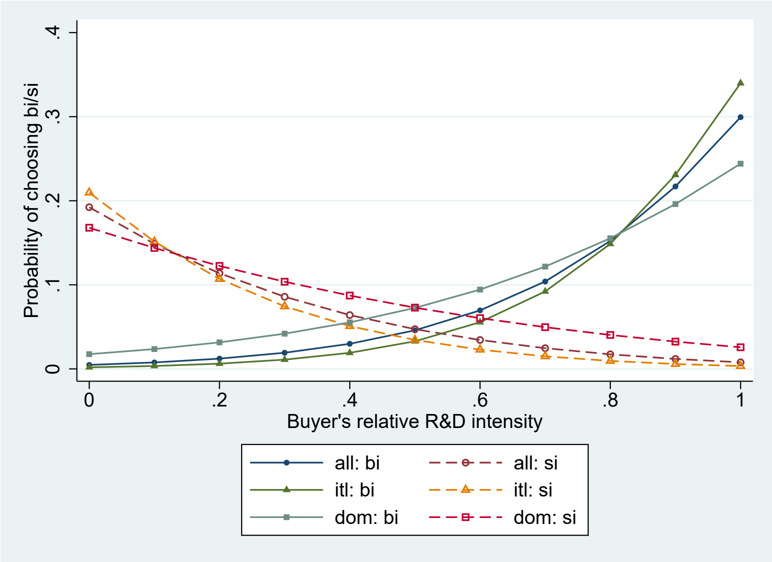


(c) Buyer's relative employment

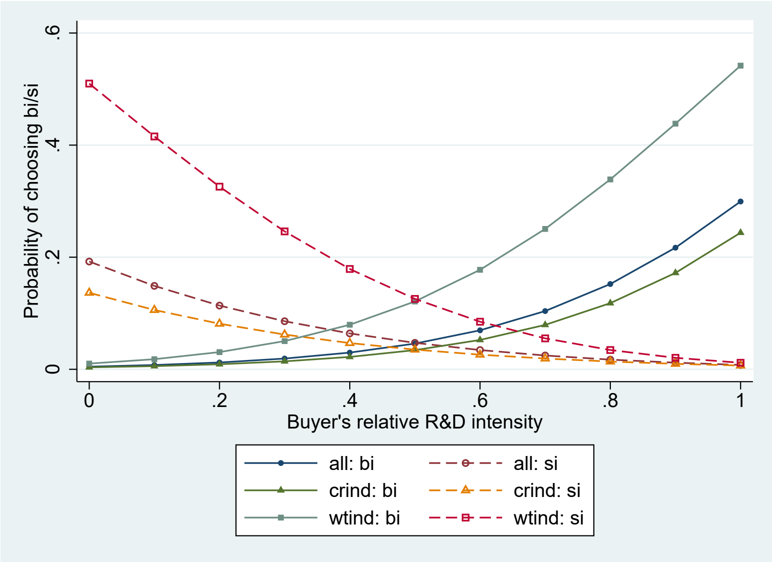


(d) Buyer's relative age

Figure 2: Comparison between Buyer's Relative R&D Intensity and Other Variables



(a) Domestic versus international relationships



(b) Within-industry versus cross-industry relationships

Figure 3: Predicted Probabilities Plots for Cross-country and Cross-industry Relationships

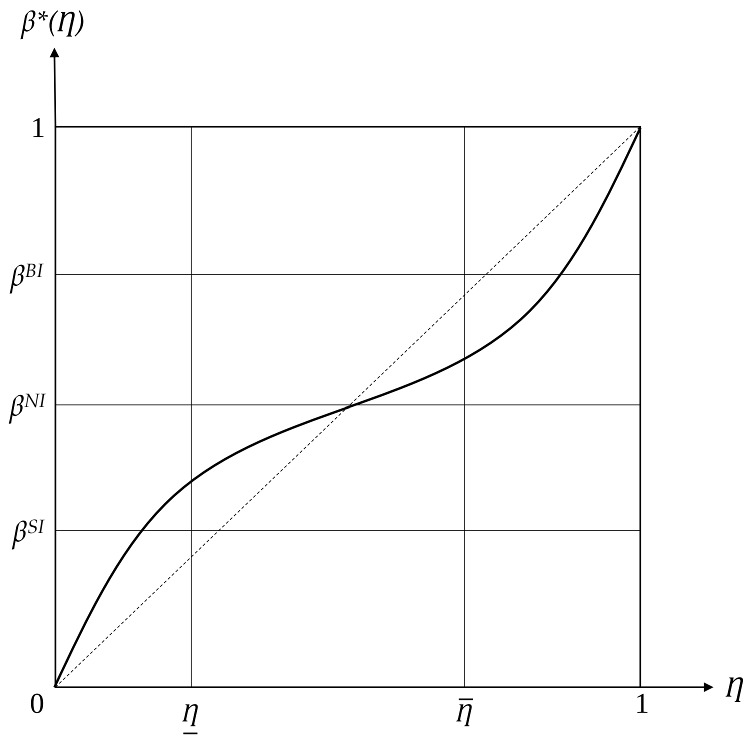


Figure A.1: The Optimal Revenue Share as a Function of η

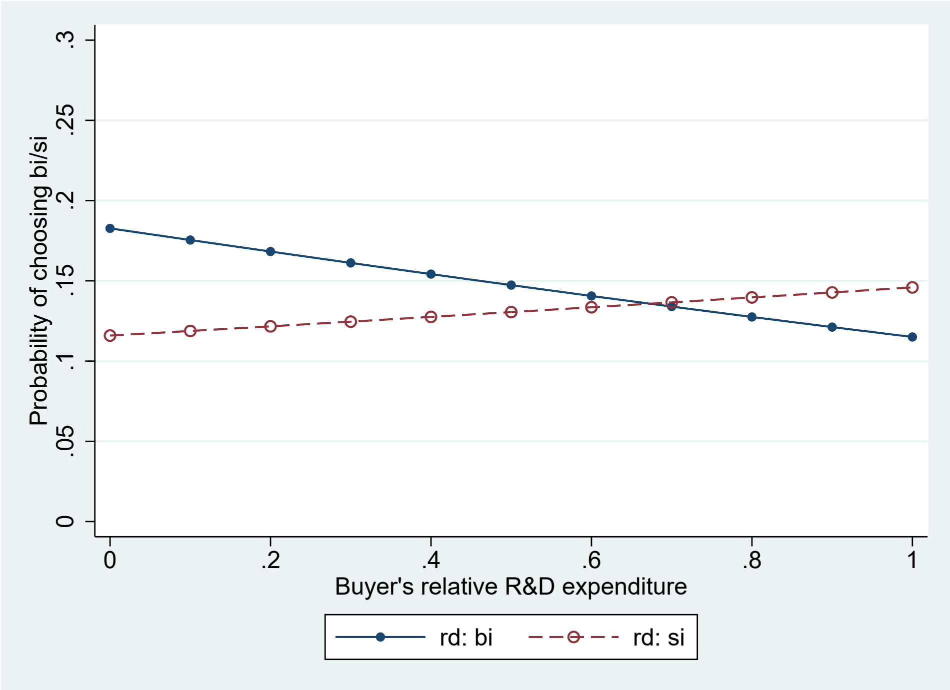


Figure E.1: Average Probabilities for Choosing Buyer/Seller Integration