Vertical Supply Chains and International Business

by

Ran Jing

M.A., The University of British Columbia, 2004

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2009

Abstract

This dissertation comprises three essays, each of which studies a business phenomenon in the context of vertical supply chains. The first essay studies the equity stakes that downstream producers hold in their upstream input suppliers. It applies and extends a property-rights model to make predictions concerning the size of such equity stakes and tests three main hypotheses using firm-level data from the Japanese auto parts industry. The second essay studies the impact of multinational retailers on China's exports. As the multinational retailers enter China, they may increase Chinese exports through linkages with distant markets where they have operations. Meanwhile, the presence of multinational retailers may stimulate productivity growth among local Chinese suppliers. Chinese city-level exports data are utilized to test and distinguish these two mechanisms. The third essay synthesizes, extends and illustrates the theory of exclusionary contracts. Two new channels though which exclusionary contracts in vertical supply chains could be anti-competitive are proposed. A recent antitrust case illustrates all incentives—the two established channels and the two new ones.

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I would like to thank my advisors, Keith Head and Barbara Spencer, who are not only helpful, but also caring. I appreciate their research advice, guidance, and all the efforts. I am very grateful to Ralph Winter and John Ries for their great inspiration and support. Many thanks to Anming Zhang and the members of the Strategy and Business Economics Division, your comments, suggestions, and acts of kindness are most appreciated.

I owe special thanks to my parents and husband who always try their best to support me at anytime.

Dedication

To my parents and husband

Co-authorship Statement

My third chapter is co-authored with Professors Keith Head and Deborah Swenson. I developed the research idea, collected the data of the multinational retailers' operation in China, and performed the econometric analyses. I also wrote up the paper and revised the paper.

My fourth chapter is co-authored with Professor Ralph Winter. I initiated the research idea of studying exclusionary contracts in the context of vertical supply chains. Under the guidance of Professor Ralph Winter, I derived the models and showed that exclusionary contracts at one stage of the supply chain could extract rents at another stage of the supply chain and the mechanism for the dual contracts strategy. I also programmed all simulations and contributed to the manuscript preparation.

Chapter 1

Introduction

This dissertation contains three essays, each of which studies a phenomenon in the context of vertical supply chains. The first essay studies the equity stakes that downstream producers hold in their upstream input suppliers. It focuses on the characteristics that could have directly *causal* effects on the size of such equity stakes. The second essay tests whether the presence of multinational retailers has a positive impact on China's exports and further identifies the mechanism through which the multinational retailers increase exports. The third essay turns to exclusionary contracts. Besides synthesizing the theories in the previous literature, this essay also proposes two new mechanisms when exclusionary contracts exist in vertical supply chains.

The full title for the first essay is "Partial Equity Stakes in Vertical Supply Chains: Evidence from the Japanese Auto Parts Industry." In vertical supply chains, downstream producers often hold equity stakes in their upstream input suppliers. This paper extends a property-rights model developed by Acemoglu, Aghion, Griffith, and Zilibotti (forthcoming) to make predictions concerning the size of such equity stakes. Three main hypotheses are tested using firm-level data from the Japanese auto parts industry. Two-stage least squares estimations correct for endogeneity problems. The panel data fractional probit estimation method proposed by Papke and Wooldridge (2008) allows for the bounded nature of the fractional dependent variable. As predicted, equity stakes are reduced if suppliers serve multiple auto producers. Also, the scope of intermediate inputs purchased from a supplier increases the producer's equity stake in that supplier. Finally, this paper predicts conditions under which a greater geographic distance between suppliers and

downstream producers may increase or leave equity stakes unchanged. Distance and equity stakes are negatively correlated in the data, but when the analysis is corrected for endogeneity, the effect of distance disappears. This irrelevance of distance is consistent with the property-rights model of the firms as long as distance only acts to lower total surplus in the relationship.

The second essay is titled, "From Beijing to Bentonville: Do Multinational Retailers Link Markets?" The world's largest retail chains—Wal-Mart, Carrefour, Tesco, and Metro—all entered China in the last 15 years. Since each of these retailers procures large quantities of goods in China for sale in their worldwide outlets, they may increase Chinese exports through linkages with distant markets where they have operations. Meanwhile, the presence of multinational retailers may stimulate productivity growth among local Chinese suppliers by increasing competitive pressure on the suppliers or allowing them to enjoy economies of scale in production and distribution. Thus the presence of multinational retailers may improve the general export capabilities of Chinese regions. To test whether the recent growth in retailer presence has helped expand the exports of Chinese regions, we use the econometric specification which controls for city-country, city-year, and country-year fixed effects. Our results suggest that retailer presence promotes China's exports by improving regional general export capabilities.

The third essay of this dissertation is "Exclusionary Contracts." The economic literature, notably Aghion and Bolton (1987), has developed three types of incentives for exclusionary contracts, contracts through which an incumbent supplier can profitably deter the entry of a rival into a market. When do participants in a market have the incentive to enter into agreements that exclude potential entrants? This paper synthesizes, extends and illustrates the theory of exclusionary contracts. Two new channels are extended from the previous literature. We first consider contracts between an incumbent and buyers within the conventional framework of a first-mover advantage by the incumbent. A "Chicago" benchmark yields no incentive for exclusionary long-term

contracts. Departures from the benchmark in each of three directions yield predictions of long-term exclusionary contracts. These include the two existing theories (Aghion and Bolton, 1987) as well as a third theory—a long-term contract at one stage of the supply chain may extract rents at another stage of the supply chain. We then consider contracts with upstream suppliers of critical inputs, for which incumbency does not yield a first-mover advantage. Two firms bid simultaneously for the rights to upstream inputs, with bids for exclusive rights being an available strategy, and then compete in a downstream market. The equilibrium allocation of input rights, when it exists, maximizes combined profits of all firms in the industry (subject to the constraint of competition downstream). With sufficient complementarity upstream and substitutability downstream, the equilibrium allocates all inputs to a single firm—excluding the other firm from the market. We examine an antitrust case that illustrates all four channels through which exclusionary incentives flow—the two established channels and the two that we introduce.

In the appendix of chapter 4, we also considered the case where the incumbent offers long-term contracts both upstream and downstream in the conventional framework of a first-mover advantage by the incumbent. We showed that there is an additional source of gains from exclusionary contracting. Each contract serves to relax the participation constraint in the other contract by reducing the probability of successful entry in the absence of the other contract.

Bibliography

Acemoglu, D., P. Aghion, R. Griffith, and F. Zilibotti. Forthcoming. Vertical Integration and Technology Theory and Evidence. *Journal of European Economic Association*.

Aghion, P. and P. Bolton (1987). Contracts as a barrier to entry. *The American Economic Review 77*, 388–401.

Papke, L. and J. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*, 121–133.

Chapter 2

Partial Equity Stakes

2.1 Introduction

In vertical supply chains, downstream producers often hold equity stakes in their upstream input suppliers.[1] Denso is the world's largest listed auto parts supplier.[2] According to a 2001 report,[3] 54.2% of its outputs were sold to Toyota Motor, and Toyota held a 24.9% equity stake in Denso. Calsonic Kansei Corp is a major part supplier for Nissan Motor. In 2001, Nissan held a 32% equity stake in Calsonic Kansei.[4] ArcelorMittal is the world's largest integrated metals and mining company, with over 310,000 employees in more than 60 countries.[5] In 2008, it increased its stake in Macarthur Coal Limited to 19.9%. Although downstream producers frequently hold equity stakes in their upstream input suppliers, analysis of the size of such equity stakes is not well developed.

This paper examines the determinants of the size of equity stakes that downstream producers hold in their upstream input suppliers. It uses and extends a property-rights model built by Acemoglu, Aghion, Griffith, and Zilibotti (forthcoming) and tests hypotheses formally predicted by the model. The three characteristics studied in the paper are the number of customers that an upstream supplier serves, the scope or variety of intermediate inputs involved in a supplier–customer relationship, and the geographic dis-

[1] A version of this chapter has been submitted for publication. Jing, R. Partial Equity Stakes in Vertical Supply Chains: Evidence from the Japanese Auto Parts Industry.

[2] The Daily Yomiuri(Tokyo), April 22, 2008

[3] *Japan Auto Parts Industry*, 2001 published by Brown & Company Ltd./Yano Research Institute Ltd.

[4] *Japan Auto Parts Industry*, 2001 published by Brown & Company Ltd./Yano Research Institute Ltd.

[5] *Business Wire*, June 29, 2008

tance between suppliers and downstream producers. The first hypothesis is that equity stakes are reduced if suppliers serve multiple downstream producers. The second hypothesis predicts that the scope of intermediate inputs purchased from a supplier increases the producer's equity stake in that supplier. The greater the scope of intermediate inputs, the higher the likelihood that the downstream producer will be held up by that supplier. The downstream producer uses the equity stake to alleviate the hold-up problem.[6] The third hypothesis concerns the effect of geographic distance on the size of equity stakes. Depending on particular assumptions about distance costs, distance may increase or leave the equity stake unchanged.

To test the predictions of the theory, this paper uses three years' (1990, 1997, and 2001) data from the Japanese auto parts industry. These data are published by the Dodwell Marketing Consultants Company, a Japanese consultancy. In the data set, the purchasing links between the 11 largest Japanese auto makers and about 600 Japanese suppliers in 243 auto parts markets are directly available. The data set also includes the percentages of these suppliers' equity stakes held by those 11 auto makers respectively,[7] which allows me to investigate impacts of key variables on the *size* of equity stakes.

Endogeneity problems hinder the investigation of the determinants of partial equity stakes in vertical supply relationships. Two points are particularly important. One is simultaneity bias. High equity stakes endow shareholders with strong control rights. When a supplier's downstream customer has such rights, the customer can influence the supplier's operation decisions. Therefore simple regression results reveal only the partial correlations between the characteristics of interest and the size of equity stakes. They cannot tell whether the variable of interest is the reason for acquiring equity stakes or the consequence of equity stakes. To address simultaneity bias, three instrumental variables for three key variables are constructed. The construction is based on exogenous features

[6] Hold up arises when part of the return on an agent's relationship-specific investment is ex post expropriable by his trading partner. See Che and Sákovics (2008).

[7] Arm's-length is defined as zero, and full integration is denoted as 100.

of different auto parts: transportability and economies of scope in production. The second concern is the unobserved heterogeneity in suppliers' attributes. It is difficult to guarantee that all supplier attributes that affect equity stakes are controlled for in the specification. The influencing factors omitted could be correlated with the variables included in the model, thereby making the estimates subject to omitted variable bias. To address this concern, models with supplier fixed effects are applied.

The bounded nature of equity stakes is another econometric challenge hindering the study of partial equity stakes. Equity stakes are between 0 and 100% inclusive. This nature determines that the marginal effects of key variables are not likely to be constant. For example, when the equity stakes are very low, it is unlikely that the negative effect of customer is as strong as that when the equity stakes are high. In order to deal with this issue, panel data fractional probit estimator proposed by Papke and Wooldridge (2008) is implemented.

This paper provides strong evidence for the hypothesis that the number of customers that an upstream supplier serves has a direct negative impact on the size of equity stakes that downstream producers hold in that upstream supplier. After instrumenting for the three key variables, the estimation results show that equity stakes decrease by 2.1% as the number of customers increases by 1, and this coefficient is significant at the 1% level. This paper also provides reasonable evidence for the direct positive effect of the scope or variety of intermediate inputs ordered by downstream producers. It is found that as the scope of intermediate inputs transacted in a relationship between a supplier and a downstream customer increases by 1%, equity stakes increase by 1.6%. This paper also shows that the geographic distance between suppliers and customers does not have a causal effect on the size of equity stakes, even though distance and the size of equity stakes are negatively associated when reverse causation is not taken out.

This paper contributes to the literature in three main ways. First, it addresses the endogeneity problem in empirical analysis and identifies *causal* effects of characteristics

on vertical integration. Reverse causation and unobserved heterogeneity are dealt with carefully. Second, the paper introduces a continuous measure of vertical integration. The previous literature (Ahmadjian and Oxley, 2006; Oxley, 1997; Pisano, 1989), focuses on the incidence of partial ownership, treating low (e.g. 1%) and high (e.g. 99%) equity stakes as the same. In this paper, equity stakes are measured by their size, and the bounded nature of fractions is addressed by a estimation method recently available—panel data fractional probit. Third, this paper introduces the number of customers actually served by a supplier into the empirical study.[8] Ahmadjian and Oxley (2006) examine the impact of the total number of suppliers in each auto part market on the incidence of partial ownership in the Japanese auto parts industry. They emphasize that this variable captures the vulnerability of suppliers in their relationships with downstream customers. They expect the variable to have a positive effect on the likelihood of equity stakes. However, their empirical analysis does not find *direct* evidence for the hypothesis.[9] In this paper, the vulnerability of suppliers is captured by a more specific measure, namely the actual number of customers served by each supplier. Acemoglu, Aghion, Griffith, and Zilibotti (forthcoming) point out that this measure should be associated with lower equity stakes, but due to limitations in their data set, their empirical analysis is still restricted to an industry-level measure—the relative number of downstream customers to upstream suppliers.

A theoretical contribution of this paper is to formally model the effect of distance on the likelihood of vertical integration through the lens of property right theory. Predictions about the distance effect are different depending on the assumptions about the way

[8]Pisano (1989) and Oxley (1997) find that the likelihood of equity stakes increases with the scope of transactions, which is measured by dummies (whether the collaboration involves multiple projects, or whether single or multiple technology or product areas are involved). Ahmadjian and Oxley (2006) explicitly introduce the numerical count of different types of auto parts transacted in each supplier-assembler dyad, which makes the analysis of the marginal effect of the scope variable feasible.

[9]Ahmadjian and Oxley (2006) find the interaction terms between the total number of suppliers and their specific investment measures (supplier dependence, model specificity, etc.) are significantly positive. However, the coefficient of the total number of suppliers in auto part markets is insignificant or even negative when they only include the total number of suppliers in the specification without the interaction term.

in which distance affects profits.[10] When we assume that distance reduces the marginal effect of the supplier's effort, the model predicts that the greater the distance between the supplier and its customer, the higher the equity stake that the downstream customer holds in its upstream supplier. However, if distance costs are assumed to decrease the level of quasi-rents in which neither the supplier nor the assembler exerts effort, the model predicts that distance has no effect on the size of equity stakes. The fact that different assumptions about distance costs give different predictions demonstrates the importance of a formal analytic model in the study of vertical integration. Such differences inform the debate about the difference between the transaction cost theory (Williamson, 1975, 1985) and the property-rights theory (Grossman and Hart, 1986; Hart and Moore, 1990). Whinston states: "Transaction cost economics is concerned primarily with the level of quasi-rents; property-rights approach's focus is on the marginal returns to non-contractible investments."(2003, p. 12)[11] Predictions based on the property-rights theory are more complicated than those based only on the transaction cost theory.

The remainder of the paper is structured as follows. Section 2.2 derives the theoretical model and proposes the hypotheses. Section 2.3 explains the estimation strategies, data, and variables. Section 2.4 presents and discusses the empirical results, before and after correcting reverse causation, with or without addressing the bounded nature of the dependent variable. Conclusions are drawn in section 2.5.

2.2 Hypotheses

Three hypotheses are tested in this paper. The first two hypotheses are drawn directly from Acemoglu, Aghion, Griffith, and Zilibotti (forthcoming). The third one is obtained by extending their model. The first hypothesis contends that equity stakes are reduced

[10] More precisely, profits represent quasi-rents.

[11] The other difference emphasized by Whinston (2003) is that the effect of the marginal return to investments on the likelihood of integration depends heavily on the nature of investments (self-investment vs. cooperative investment) and the ownership structure under which the investment is made.

if suppliers serve multiple assemblers. The second prediction is that the scope of intermediate inputs purchased from a supplier increases the producer's equity stake in that supplier. The third hypothesis examines the effect of geographic distance between suppliers and downstream producers on the likelihood of vertical integration. Given the way in which distance affects profits, two predictions about distance effect are proposed. Section 2.2.1 explains the first two hypotheses. The effect of distance on the likelihood of vertical integration is discussed in section 2.2.2.

2.2.1 Numbers of Customers and Scopes of Parts

Following the property-rights theory initiated by Grossman and Hart (1986) and Hart and Moore (1990), Acemoglu, Aghion, Griffith, and Zilibotti (forthcoming) (AAGZ) study three organizational forms between a supplier and a downstream producer: backward integration (VIB), in which the buyer or downstream producer owns its upstream input supplier, non-integration (NI), in which the two parties remain independent; and forward integration (VIF), in which the supplier owns its downstream producer. AAGZ, as well as Whinston (2001, 2003), point out that the relevant comparison in practice is between backward integration and non-integration. This paper focuses on these two organizational forms in both analytic and empirical analysis.[12]

AAGZ predict the effects of two characteristics on the likelihood of backward integration.[13]

1. As the share of the revenue that the supplier obtains if she sells the input outside the specific relationship increases, backward integration is less likely relative to non-integration.

2. As the share of the cost of the supplier's inputs in the producer's total cost increases, backward integration is more likely relative to non-integration.

[12]In other words, the term "vertical integration" used afterwards refers to backward integration specifically.

[13]AAGZ provide a detailed proof for the following two hypotheses in Proposition 1.

The underlying mechanism of the first hypothesis is as follows. Its key parameter is the ratio of supplier's revenue if she sells the input outside the specific relationship relative to the revenue if she sells the input inside. When this ratio is high, it means that the supplier has a better outside option. According to the symmetric Nash bargaining solution, the party who has a better outside option enjoys a higher revenue. Thus a supplier with a better outside option has a stronger incentive to exert effort. This parameter is *not* relevant under backward integration, because under backward integration the supplier does not have the residual right of control over the inputs it has produced. Under the assumptions of no financial constraints on each party and feasible ex ante transfers, the objective is to choose the organizational form that maximizes the total output generated by the relationship. Given all other variables are constant, a higher share of the revenue that the supplier obtains if she sells the input outside the specific relationship relative to inside makes non-integration more favorable than backward integration because it induces higher effort from the supplier.

In this paper, the likelihood of vertical integration is measured by the size of equity stakes that downstream producers hold in their upstream suppliers. The previous literature (Ahmadjian and Oxley, 2006; Oxley, 1997; Pisano, 1989, and Acemoglu et al., forthcoming) all study the incidence of partial ownership. In the Japanese auto parts industry, it is widely observed that suppliers are owned by their downstream producers with very low equity stakes. However, suppliers owned by assemblers with very low equity stakes are more likely to think of themselves as non-integrated.[14] In the meantime, suppliers with *relatively* high equity stakes are more likely to behave as if they are integrated. The traditional dichotomous measure treats very low equity stakes as full vertical integration, which may introduce measurement errors in this paper given this paper's hypotheses and context. A variety of thresholds for dichotomies are available in

[14] Ahmadjian and Oxley (2006) emphasize that institutional features of the Japanese economy lead to a de facto inability of assemblers to easily sell their equity stakes in suppliers. Therefore even very small partial equity stakes constitute relatively *robust hostages* in the Japanese context.

the literature. Three types of cutoffs are shown Appendix 1.3. The differences across these classifications make it hard to judge the level of equity stakes above which the partial ownership should be regarded as full vertical integration. A continuous measure of the size of equity stakes seems to be a natural way to deal with this problem.

In AAGZ, the proxy of the revenue share in the first hypothesis is the number of downstream producing firms relative to supplying firms at the industry–level. The intuition behind the measure is that with more downstream producers relative to the upstream suppliers, it is easier for the supplier to find a suitable buyer for her input in the secondary market after a breakup.[15] One concern with this measure is that it is easy to imagine that when a downstream producer fully or partially owns its supplier of a particular part, it is less likely that this downstream producer buys the part from the other suppliers.[16] The other concern with this measure is that it implicitly assumes that the suppliers in the upstream industry have the same market shares. Given the fact that the size distribution of suppliers in the Japanese auto parts industry is rather skewed, this measure could introduce errors.

This paper measures the supplier's revenue share in the first hypothesis by the number of customers that a supplier actually has. When a customer refuses to buy the inputs that she has originally promised to buy, the supplier can easily ask all her other customers and try to find a suitable customer. The more customers the supplier has, the higher the probability of finding a new buyer. The supplier thus has a better outside option.

Based on the first prediction of proposition 1 in Acemoglu et al. (2005), the specific hypothesis to be tested in this paper is

Hypothesis 1: *The number of customers that a supplier serves reduces the size of equity stakes that downstream producers/assemblers hold in this supplier.*

The second prediction made by AAGZ tested here is that as the share of the cost of the supplier's inputs in the producer's total cost increases, backward integration is more likely

[15] AAGZ caution that the relative number of producers to suppliers could be endogenous.

[16] It is consistent with the idea of market thickness theory (See McLaren, 2000).

relative to non-integration. This prediction is easy to understand. A greater share of costs (of the supplier's inputs in the producer's total costs) increases the degree to which the producer is held up by the supplier under non-integration. Hence under non-integration, the downstream producer has a lower incentive to exert effort ex ante; meanwhile this cost share does not affect the producer's effort decision under the organizational form of backward integration because under backward integration the producer owns the inputs that the supplier has produced. Therefore, a high share of the cost of the supplier's inputs in the producer's total cost makes non-integration less favorable relative to backward integration; in other words, a high cost share makes backward integration more likely.

In AAGZ, the cost share studied in the second hypothesis is measured at 2/3-digit industry level from an Input-Output table. This macro-level measure implicitly assumes that all supplier-customer pairs are similar and may introduce measurement error.

In this paper, the cost share is measured by the scope of intermediate inputs transacted in supplier-assembler relationships. It seems reasonable that the cost share of a supplier's inputs in its customer's total costs increases with the scope of intermediate inputs transacted in the relationship. More essentially, a greater scope of intermediate inputs that the downstream producer buys from the supplier, the higher likelihood that the downstream producer will be held up by the supplier in the absence of backward integration. The control power associated with the equity stake that downstream producer holds in the supplier helps alleviate the hold-up problem. In the following empirical analysis, the scope of intermediate inputs is measured by the number of different auto parts. As before, the likelihood of vertical integration is measured by the size of equity stakes. Thus the hypothesis to be tested is

Hypothesis 2: *The scope of auto parts in a supplier-assembler relationship increases the equity stake that the downstream producer holds in the upstream supplier.*

Ahmadjian and Oxley (2006) also study the effect of the number of varieities on the likelihood of equity stakes and show that varieties have a significantly positive effect on

the likelihood of equity stakes. They build their conceptual framwork on Williamson's (1983) hostage model and regard partial equity stakes in the Japanese auto parts industry as a hostage-based governance mechanism that downstream assemblers use to show their commitments to otherwise vulnerable suppliers. The number of varieties that a supplier provides in a bilateral relationship is regarded as a measure of the supplier's vulnerability. The current paper examines equity stakes through the lens of the property rights theory and focuses on whether the number of varieties could have a *causal* effect on the size of equity stakes. The number of varieties that a supplier provideds to a particular assembler represents the importance of the supplier's efforts in the assembler's production. Intuitively, it is a measure of the assembler's vulnerability. The empirical results from the two stage least square estimations suggest that this approach captures important aspects of the vertical integration decisions of the Japanese auto assemblers.

2.2.2 Distance

There is no consensus about how geographical distance affects the likelihood of vertical integration. Previous research predicts that distance could have either a direct positive (Williamson, 1975, 1985) or negative effect (Head and Ries, 2008; Lafontaine, 1992; Minkler, 1990) on the decision of vertical integration.[17] This paper examines the effects of distance on the likelihood of vertical integration under two different assumptions about distance costs. Each assumption gives a unique prediction about the impact of distance on the likelihood of vertical integration. The empirical results will then show us which assumption is consistent with the data set used in this paper.

The model built by AAGZ emphasizes the role of efforts exerted by a supplier and its downstream producer. In their model, the effort levels exerted by the supplier and downstream producer are denoted by e_s and e_p respectively. The marginal effects of

[17] In the franchising literature, it is said that distance makes monitor difficult, and incentives have to be increased to "increase the attractiveness of contractual (versus integrated) solutions (Oxley, 1997, footnote 21)." Head and Ries (2008) assume inspection cost increases with geographic and cultural distances and predicts compensations paid to subsidiaries increases with distance.

their efforts are s and p correspondingly. s is the marginal effect of the supplier's effort; p is the marginal effect of the downstream producer's effort. The value of intermediate inputs in which neither the supplier nor the downstream producer has exerted efforts is normalized to 1. The share of the cost of the supplier's input in the producer's total cost is denoted by ϕ. x_s shows whether or not the inputs are eventually supplied to the downstream producer. It is equal to 1 if it is supplied; otherwise, x_s equals 0. The production technology of the vertical relationship assumed by AAGZ is

$$F(x_s, e_p, e_s) = (pe_p + se_s + 1)\phi x_s + (pe_p + 1)(1 - \phi), \tag{2.2.1}$$

The first half of the right-hand side of the equation above is the output from the intermediate parts; the second half refers to the output produced by the assembler alone.

The first method of modelling distance costs is to assume that distance decreases only the value of the parts in which neither the supplier nor the assembler has exerted any effort. The point in making such an assumption is that when the supplier increases a unit of effort, the positive marginal effect of her effort on the total output is not dampened by distance. Since the value of the part without the supplier or the assembler's effort is normalized to 1, in order to implement this assumption, the "1" in the first parenthesis is replaced with "$1/d$", where d is the distance between the supplier and the assembler. The new production technology function is

$$F(x_s, e_p, e_s) = (pe_p + se_s + \frac{1}{d})\phi x_s + (pe_p + 1)(1 - \phi). \tag{2.2.2}$$

This model predicts that distance has no effect on the likelihood of backward integration. The proof is shown in Appendix 1.1. This result is essentially consistent with the finding emphasized by Whinston (2001, 2003), that in property-rights theory, it is the *marginal* effect of effort rather than *the level of quasi-rents* that influences the decision of vertical integration.

Under the assumption that distance decreases only the value of the parts in which neither the supplier nor the assembler exerts any effort, it is predicted that distance has no effect on the likelihood of backward integration. Given the likelihood of vertical integration is measured by the size of equity stakes in this paper, the hypothesis to be tested is

Hypothesis 3A: *Distance has no effect on the size of equity stakes.*

The second method of incorporating distance in the model is to assume that distance has a negative effect *only* on *the marginal effect of the supplier's effort.* This can be understood intuitively. The supplier exerts effort in the production of intermediate inputs. The value of this effort cannot be fully utilized by the assembler unless the assembler understands all those techniques. The longer the distance, the lower the proportion of techniques transmitted from the supplier to the assembler. Importantly, the fewer techniques that the assembler understands, the less valuable the supplier's effort becomes to the assembler. In essence, distance reduces the "effectiveness" of the supplier's effort. In order to implement this assumption, the marginal value of supplier's effort denoted by s is divided by d, where d is the distance. Thus the new production technology function is

$$F(x_s, e_p, e_s) = (pe_p + \frac{se_s}{d} + 1)\phi x_s + (pe_p + 1)(1 - \phi). \tag{2.2.3}$$

Under this assumption, the model predicts that as distance increases, backward integration becomes more preferable. The proof is in Appendix 1.3. The underlying intuition is as follows. The purpose of allowing the supplier to be non-integrated is to induce its effort. When the supplier's effort becomes less effective, non-integration loses its value, and then backward integration becomes more preferable.[18] This prediction is derived directly from a standard property-rights model; however, it also conforms with Williamson's pre-

[18] As distance increases, the optimal efforts of independent suppliers and non-independent suppliers both decrease. However, a controlled supplier reduces efforts relatively less than an independent supplier, because the reduction in the total payoff of the controlled supplier due to distance is partially offset by its assembler. Given the assembler's effort does not change with distance, backward integration elicits higher effort and generate higher total output.

diction, that, when distance is long, the transaction becomes more complex, and it is better to be vertically integrated.

Under the assumption that distance has a negative effect *only* on the *marginal effect of the supplier's effort*, the model predicts that distance increases the likelihood of backward integration. Since the likelihood of vertical integration is measured by the size of equity stakes in this paper, the hypothesis to be tested is

Hypothesis 3B: *Distance increases the size of equity stakes.*

The empirical analysis will test the validity of these alternative predictions of distance costs.

2.3 Specification and Data

2.3.1 Econometric Strategies

The following specification is employed to test the hypotheses.

$$\begin{aligned}\text{Own}\%_{ast} =& \text{Constant} + \alpha_c \text{Customers}_{st} + \alpha_v \ln \text{ Variety}_{ast} + \alpha_d \ln \text{ Distance}_{ast} \\ &+ \beta X_{st} + D_a + D_t + \upsilon_s + \varepsilon_{sat} \end{aligned} \tag{2.3.1}$$

s : supplier $\qquad$ t : year $\qquad$ a : assembler

where X_{st} is a vector of time-varying control variables for supplier s at time t. It includes L_{st} (number of employees), K_{st}/L_{st} (capital intensity), and Age_{st}. D_a is the dummy variable for assemblers. Assembler Daihatsu is taken as default. D_t is the dummy variable for three years, of which 1990 is excluded. ε_{sat} is an idiosyncratic error. The υ_s represents supplier-specific unobservable characteristics that influence the size of equity stakes. In some specifications, υ_s will be captured by dummy variables. Instrumental variables for customers_{st}, variety_{ast}, and distance_{ast} will be employed to address endogeneity associated with each variable.

2.3.2 Data

The data come from three specific years' (1990, 1997, and 2001) market reports published by Dodwell Marketing Consultants Company. Based on its information, two data sets are constructed: one comprises the information on equity stakes that the 11 largest Japanese auto assemblers hold in about 600 Japanese auto parts suppliers, and the characteristics of these suppliers. The other data set details the purchasing links between these assemblers and suppliers in 243 auto parts markets.

As for the data set of equity stakes and characteristics, the equity stake is defined as the equity stake of a supplier held *directly* by an assembler.[19] Each supplier has 11 observations in each year. Each observation describes the size of equity stake that one of the 11 Japanese auto assemblers holds in the supplier. As we can expect, dyads with zero equity stakes amount for a large proportion of the data set. The characteristics of the suppliers include the number of employees, the year established, address, etc. Among these characteristics, distance between plants of an assembler and supplier is calculated by using the formula of great circle distances. Longitudes and latitudes of the Japanese cities where the headquarters of suppliers are located and assemblers build their plants are used. Over the three sample periods, some suppliers came into the market, and some dropped out. Thus the data set of equity stakes and suppliers' characteristics is imbalanced.

As for the other data set, the purchasing links are defined as whether or not there is a transaction between a supplier and an assembler. The value or the quantity of the transaction is not available. One advantage of this data set is that the supply relationship is available in each auto part market. This feature makes it feasible to construct instrumental variables based on heterogeneities of different auto parts. Among the 14399 assembler–supplier–year observations who have complete information on all the variables involved in Specification 2.3.1, only 4648 observations have positive transaction links. In

[19] This measure neglects the scenario in which there are indirect equity stakes.

the result section, most regressions are run on the full sample including the observations who do not have positive transaction links in that year. The reason is that each of the 11 Japanese assemblers could potentially be a customer of each supplier. A robustness test is conducted on the sample in which only the observations with positive transaction links are included.

Dependent Variable and Key Variables

Similar to Ahmadjian and Oxley (2006), the unit of observation is a supplier–assembler dyad rather than a transaction for each auto part.[20] The dependent variable is the percentage of a supplier's equity stake held by an assembler in the year. This variable is represented by $\text{Own}\%_{ast}$, where the subscripts a, s, and t refer to assemblers, suppliers, and year, respectively. $\text{Own}\%_{ast}$ ranges from 0 to 100. Table 2.1 presents the distribution of equity stakes in the sample with all dyads, and Table 2.2 presents the data for dyads with supply relationships. In each table, the top panel shows the proportion of zero equity stakes; the bottom panel lists the proportion of each critical level of equity stakes *within* the subsample of dyads with positive equity stakes. In Table 2.1, the majority of dyads have zero equity stakes. Among the dyads with positive equity stakes, about 50% have their equity stakes below 20%; around 70% have equity stakes below one-third. Similar patterns are also exhibited in Table 2.2, i.e. the sample confined to supply relationships. The main difference is that zero equity stakes account for a lower proportion.[21]

Table 2.3 gives the definition of all variables and the predicted signs of the three key variables. The three key variables are customer_{st}, distance_{ast}, and variety_{ast}. Customer_{st} is the number of customers that a supplier serves in the year. This measure disregards information about how many different auto parts are bought by a customer. As long

[20] Oxley (1997) points out that it is the attributes of the transactions rather than the firm as a whole that determine the most efficient mode of governance in alliance. However, because equity shares are specific to dyads, they can be analyzed only at the supplier-assembler dyad level.

[21] The bottom panel of Table 2.2 shows that the equity stakes of less than 5% account for 17% of the subsample. This statistics is lower than the corresponding data, i.e. 33%, in Ahmadjian and Oxley (2006). In their paper, the data in 1987 are utilized.

Table 2.1: Equity Stake Distribution of All Dyads

Size of Equity Stake	Percentages of dyads with equity stake in this range			
	1990	1997	2001	All
0	95.9	96.1	96.3	96.1
(0, 100%]	4.2	3.9	3.7	3.9
(0, 5%]	18.6	16.5	16.5	17.2
(5%, 10%]	16.2	17.0	19.9	17.6
(10%, 20%]	14.2	14.6	11.9	13.7
(20%, 33%]	19.1	20.3	21.0	20.1
(33%, 50%]	16.7	18.9	17.6	17.7
(50%, 100%]	15.2	12.7	13.1	13.7

Table 2.2: Equity Stake Distribution of the Dyads with Supply Relationships

Size of Equity Stake	Percentages of dyads with equity stake in this range			
	1990	1997	2001	All
0	86.4	88.4	89.5	88.2
(0, 100%]	13.6	11.6	10.5	11.9
(0, 5%]	18.0	16.3	16.7	17.0
(5%, 10%]	16.5	16.3	19.1	17.2
(10%, 20%]	14.4	14.9	12.5	14.0
(20%, 33%]	19.6	20.8	20.8	20.4
(33%, 50%]	16.5	18.8	17.9	17.7
(50%, 100%]	15.0	12.9	13.1	13.7

as an assembler buys something from a supplier, this assembler is counted as one of the supplier's customers. As discussed the previous section, the number of customers that a supplier serves is expected to have a negative impact on the size of equity stakes. The second key variable of this paper is variety$_{ast}$, which is the scope of auto parts purchased from a supplier by an assembler. It is defined for each supplier-assembler pair and predicted to have a positive effect on the size of equity stakes. The third key variable is the geographic distance between a supplier and an assembler denoted by distance$_{ast}$. All assemblers in this study have multiple factories, and these factories are in most cases located in different Japanese prefectures.[22] Distance$_{ast}$ is the distance from the headquarter of the supplier to the nearest factory of a certain assembler. The logic of this measure is that just-in-time delivery is important in the Japanese auto industry.[23]

[22]The sub-contract assembly factories of the assemblers are disregarded in calculating the distances.

[23]Another measure of distance has also been constructed. It is a weighted average of distance from a supplier to all the factories of an assembler, with the number of employees of each factory as weights. The estimated results are very similar to the estimates shown later other than with lower R^2 and higher RMSE in fitting Specification 2.3.1.

This dyadic measure of distances varies over time because several assemblers shut down or set up factories over the sample periods.

Table 2.3: The Definitions of Variables

Variable	Definition	Sign
$\text{Own\%}_{ast}$	Percentages of equity stake owned by the assembler	
Customer_{st}	Number of customers the supplier serves	negative
Variety_{ast}	Scope of auto parts transacted between a supplier and an assembler	positive
Distance_{ast}	Minimal distance between a supplier and an assembler's all factories[a]	zero or positive[b]
Ln L_{st}	Natural logarithm of a supplier's number of employees	
Ln (K_{st}/L_{st})	Natural logarithm of a supplier's capital intensity defined as the ratio of paid-up-capital over number of employees	
Age_{st}	Age of the supplier which is equal to current year minus the year when the supplier was established	
CustIV_{st}	Instrument for Customer_{st}, the maximal number of potential customers over all parts produced by the supplier	
VarIV_{st}	Instrument for Variety_{ast}, total number of related parts produced by the supplier	
DistIV_{ast}	Instrument for Distance_{ast}, the minimal transported distance over all the auto parts transacted in the dyad	

Note: All variables are defined in each year. (a) Each assembler usually has multiple plants in Japan. The distance from the supplier to the assembler's nearest plant is applied in the data set.(b) It depends on the underlying assumption about distance costs. Hypothesis 3A predicts no effect; hypothesis 3B predicts a positive effect.

Control Variables

Table 2.3 also presents control variables which describes the characteristics of suppliers. They are number of employees, age, and capital intensity, of which capital intensity is defined as the ratio of paid-up capital relative to number of employees. These variables help alleviate omitted variable bias.

Instrumental Variables

A main contributions of this paper is testing the *causal* effects of the three key variables on the size of equity stakes. To address endogeneity, two–stage least squares estimations

Table 2.4: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
$\text{Own\%}_{ast}$	1.08	7.37	0	100	15080
Customer_{st}	3.49	3.09	1	11	15081
Variety_{ast}	0.75	2.10	0	58	15125
Distance_{ast}	127.19	118.52	0	991.13	15125
L_{st}	2069.23	5852.06	12	76479	14905
$\text{K}_{st}/\text{L}_{st}$	2.43	4.21	.01	74.29	14905
Age_{st}	47.20	14.83	4	112	14696
CustIV_{st}	5.17	2.20	0	10	15125
VarIV_{st}	20.93	19.31	1	162	15059
DistIV_{ast}	83.37	28.84	13.07	213.95	4785

with three instrumental variables (IV) are applied. The following part of this section discusses the three instrumental variables. For each instrument, three issues are considered: construction methods, justification as a valid IV, and identification conditions.

A. Instrumental Variable for Customer$_{st}$

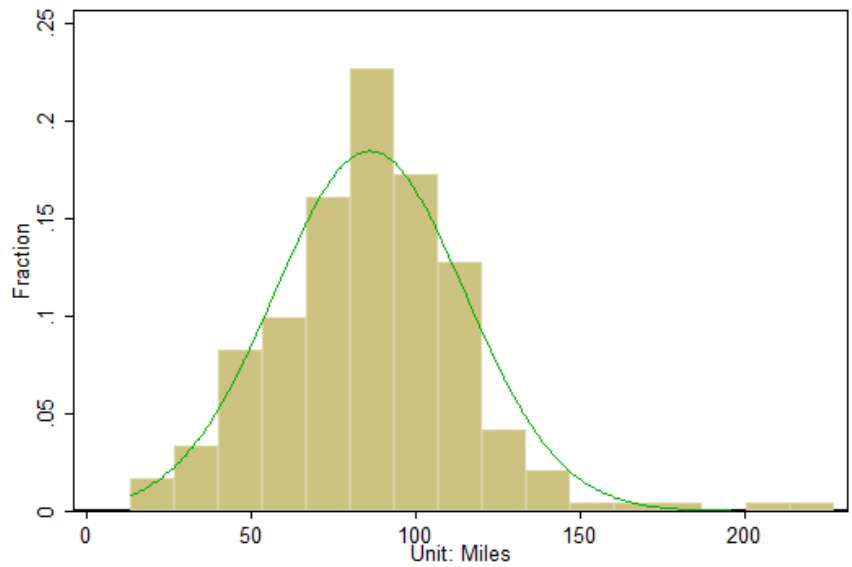
CustIV_{st} is the instrumental variable for customer_{st}. It is constructed as follows.[24] In step 1, the average transportation distance of each part is calculated. It is equal to the mean of the distances of all assembler–supplier pairs who have transactions in that part. This average distance reveals the transportability of each part. Figure 2.1 depicts the distribution of these distances over 243 auto parts.[25] The average transportation distance varies a lot, which is indispensable in the identification. In step 2, I count, for each auto part produced by the supplier, the number of assemblers who have at least one factory within that part's average transportation distance from the supplier. This number shows how many assemblers could potentially buy that auto part from the supplier simply in terms of the feasibility of transportation. In step 3, the maximal number from step 2 among all parts produced by the supplier is taken as the supplier's instrumental variable for customer_{st}.

These three steps could be illustrated intuitively by an example described in Table

[24] It is inspired by Holmes (1999).

[25] The in-house production observations are not included in the sample generating Figure 2.1.

Figure 2.1: Average Transportation Distance of 243 Auto Parts



2.5. Suppose there is a supplier who produces both air conditioners and air bag sensors. Its minimal distances to the 11 assemblers are shown in the last two rows of Table 2.5. For example, the distance from the supplier to Toyota is 10 kilometers; the distance to Nissan is 140 kilometers. The calculation in step 1 shows that the average transportation distances for air conditioners and air bag sensors are 80 and 15 kilometers, respectively.

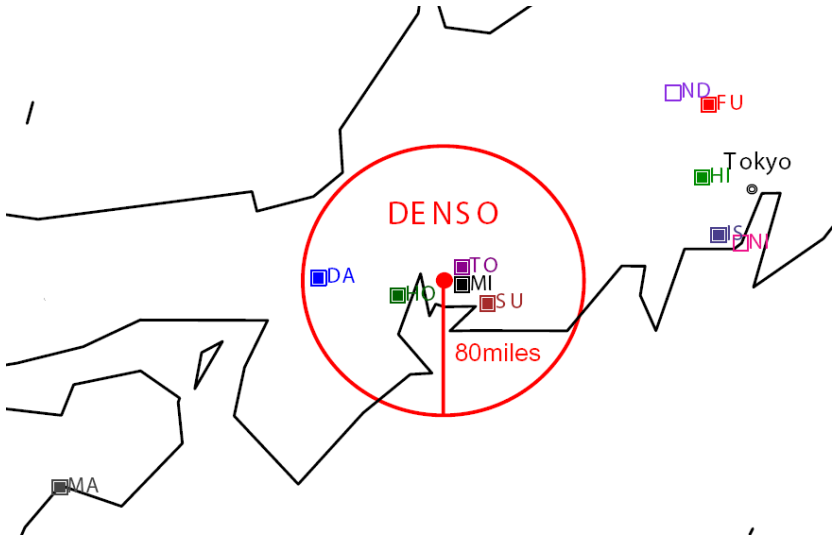
Table 2.5: An Example of the Construction Method of Customer's Instrument

Assembler	TO	NI	HO	MI	MA	ND	IS	SU	FU	HI	DA	Cu
Air conditioners	10	140	25	9	230	150	140	25	160	140	65	5
Air bag sensors	10	140	25	9	230	150	140	25	160	140	65	2

All assemblers are represented by their first two letters' abbreviation. The only exception is "ND" which refers to Nissan Diesel. All numbers are in kilometers.

In step 2, in order to know the number of potential customers of this supplier in air conditioners, each assembler–supplier distance is compared with 80. 5 assemblers are within this radius which is shown in Figure 2.2, and therefore the number 5 appears in the last column of the second row. Similarly, on air bag sensors, the supplier has 2 customers. Step 3 picks up the largest one of the numbers obtained from step 2. For this

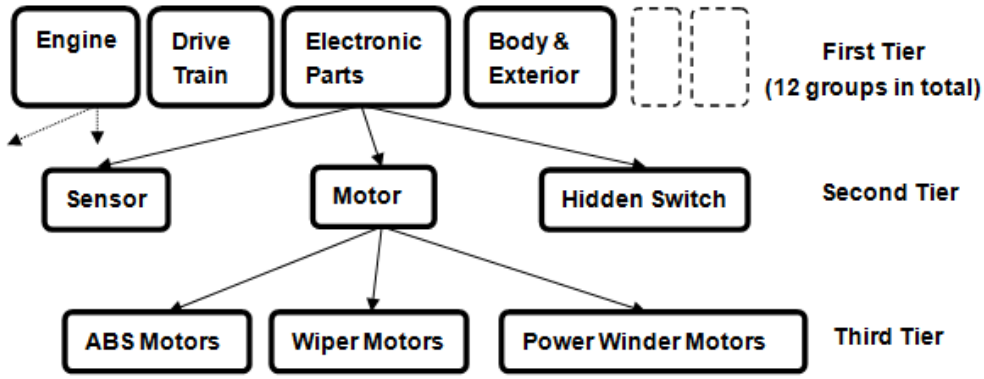
Figure 2.2: The Nearest Plants to Denso in 2001



supplier, the custIV_{st} equals to 5.

In order to be a good IV, custIV_{st} should predict customer_{st} and be unrelated to equity stakes. The economic justification of custIV_{st} as an IV are as follows. First, custIV_{st} predicts the variable it instruments for, i.e. customer_{st}. It is reasonable that the more assemblers are located within the convenient transportation distance to a supplier, the more customers that this supplier could actually have. Customer_{st} should be positively correlated with custIV_{st}. Their correlation is 0.303, as shown in the first panel of Figure 2.4. Second, there is little reason to believe this IV is endogenous. Specifically in terms of simultaneity bias, custIV_{st} is unlikely to be affected by equity stakes. It is true that a supplier's share–holding assembler could make use of its control rights and ask the supplier to locate nearby, they cannot forbid other assemblers from setting up factories close to their suppliers. Also, the suppliers who produce parts that are easily transported could have large numbers in custIV_{st}. Besides the channel through the control rights, reverse causation could also act through another channel. When the equity stake held by a buyer is zero or very low, the supplier thinks that the buyer has no tie with it and worries that the buyer may hold it up after the investment is made. In order to minimize this

Figure 2.3: Variety Instrumental Variable Construction Method Illustration



risk, the supplier tries to have more buyers. The instrument for the number of customers is the number of potential customers instead of the number of actual customers. As long as the location of the supplier is chosen before the actual customers are determined, the instrumental variable applied could still alleviate the endogeneity concern through this channel.

In addition to these basic requirements of valid instrumental variables, in order to be identified in specifications with supplier fixed effect, the IV must vary over time within suppliers. Because of the evolution of auto parts, the addition and cancellation of parts produced by suppliers, and the shutdown and opening of assemblers' factories, custIV_{st} varies within suppliers over time.

B. Instrumental Variable for Variety$_{ast}$

The second key variable is the scope of auto parts transacted in each supplier-assembler pair, i.e. variety_{ast}. Similar to customer_{st}, variety_{ast} is subject to simultaneity bias. In order to test whether variety_{ast} has a direct impact on $\text{own}\%_{ast}$, an instrumental variable (varIV_{st}) is constructed. Its construction is based on economies of scope in production.

The variable is constructed as follows. In step 1, auto parts are classified using a list of

auto parts classification from *Automotive Information Platform.*[26] Figure 2.3 illustrates the list's structure. In this classification, all auto parts are classified into three hierarchy tiers. At the highest tier, the auto parts are classified into 12 large groups, such as engine, drive train, body and exterior, clean energy system, electronics and electronic parts, etc. Each such group comprises different categories, i.e. the second tiers. Specific parts studied in this paper are in most cases listed in the third tier. Take the first-tier category "electronics and electronic parts" for instance. Within this tier, there are sensors, motors, hidden switches categories, etc. Under the motor category in tier three, there are ABS motors, wiper motors, power window motors, etc. Step 2 calculates the total number of related parts for each second-tier category. For example, there are three types of motors under the motor category. Hence, the number of related parts of the motor category is three. Step 3 identifies all the categories at the second tier that each supplier produces in each year, matches those categories with their corresponding numbers of related parts compiled in step 2, and then calculates the "total" number of related parts for supplier s in year t. The word "total" implies that when a supplier works in more than one second-tier category, the total numbers of related parts of the supplier is the sum of numbers of related parts over all categories involved.

The procedures to obtain varIV_{st} can be demonstrated by an example. Suppose there is a supplier who produces three parts, i.e. differential gears, differential castings, and drive position memory systems. According to the parts classification list, it is known that these three parts are involved in two second-tier categories: differentials and vehicle dynamic control systems. In the category of differentials, besides differential castings and differential gears, there are differential carriers. The number of related parts in the differential category is then equal to three. In the vehicle dynamic control system category, aside from drive position memory systems, there are two other memory systems. The number of related parts in the category of vehicle dynamic control systems thus also

[26] `http://www.marklines.com`, accessed in December 2006.

equals three. The total number of related parts for this supplier is therefore equal to six, i.e. 3+3=6. Each "3" corresponds to each category. What is noteworthy in this example is that this variable is unchanged as long as the supplier keeps producing parts in the categories of differentials and vehicle dynamic control systems, regardless of how many specific parts within each category are produced.

From this construction, we expect that varIV_{st} predicts the key variable variety_{ast} positively. This is based on economies of scope in production. Because related parts share some common fixed production costs, it is efficient to produce them together. In other words, related parts are likely to be produced together.

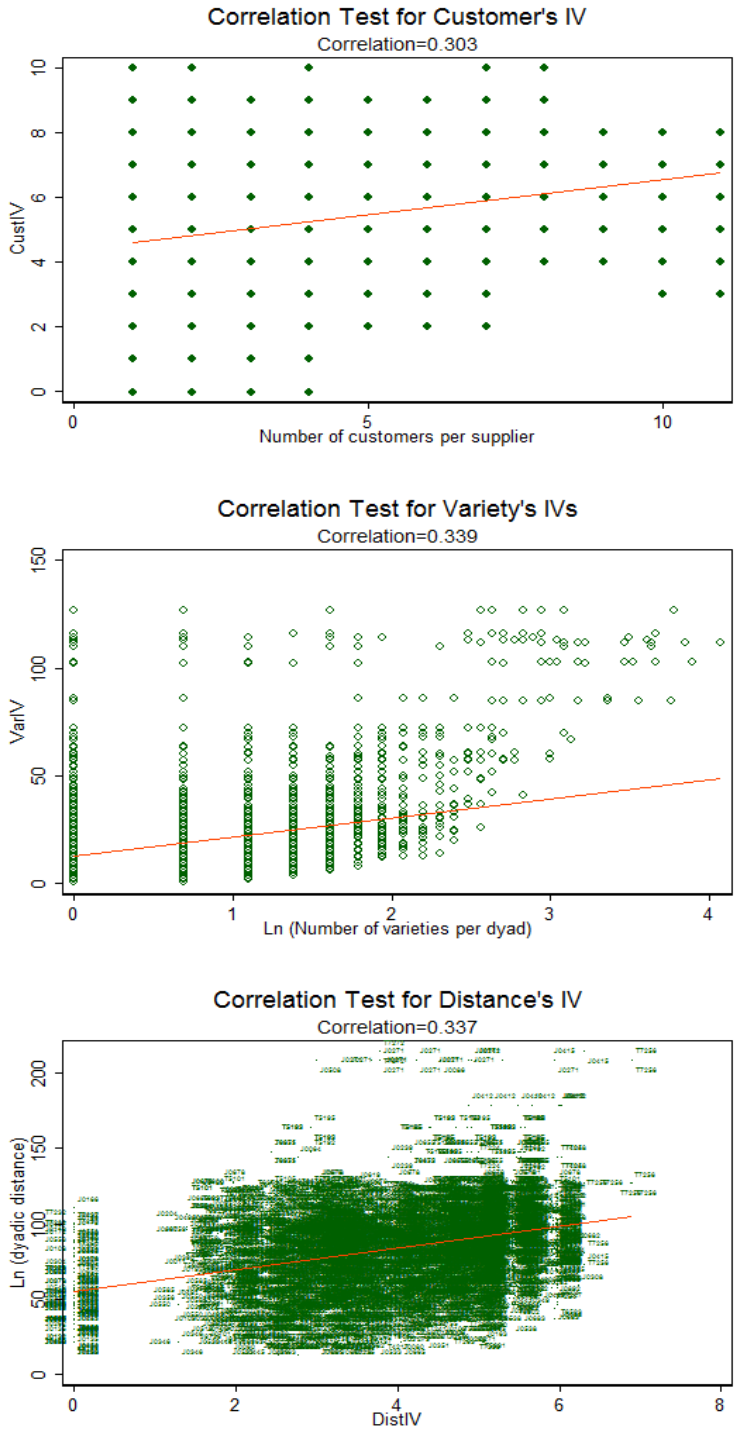
There is little reason to believe that the total number of related parts of suppliers are affected by equity stakes. The classification of auto parts is completely based on production technology. For suppliers or assemblers at a given time, the current technology is usually taken as given. Hence, this instrumental variable is exogenous.

However, varIV_{st} has limited time-series variation. As shown in the previous example, if a supplier adds or reduces parts only within its current second-tier categories, varIV_{st} has no time-series variation. Time-series variation exists only when the second tier categories vary over time for a supplier. This feature implies that varIV_{st} cannot be identified in specifications with supplier fixed effects. In other words, variety_{ast} instrumented by varIV_{st} cannot be identified in specifications with supplier fixed effects.

C. Instrumental Variable for Distance$_{ast}$

The third key variable is geographic distance between suppliers and downstream assemblers. Simple regression results are subject to serious simultaneity bias: high equity stakes endow share-holding assemblers with rights to order suppliers to locate nearby, or these suppliers' locations are determined when assemblers consider their equity stakes in those suppliers. To address this concern, the minimal transportation distance over all the parts transacted in a supplier–assembler relationship (distIV_{ast}) is taken as the instrumental variable for dyadic distance. The transportation distance is the average

Figure 2.4: Instrumental Variables Correlation Tests



transportation distance of each part calculated in the first step in constructing custIV_{st}. Its distribution is shown in Figure 2.1. Over all parts transacted within a dyad, the minimal distance is taken as the measure of distIV_{ast}.

This instrumental variable is positively correlated with dyadic distance and does not seem to be affected by equity stakes. The correlation between distIV_{ast} and distance_{ast} equals 0.337, and their scatter plot is shown in the last panel of Figure 2.4. This instrumental variable is constructed based on the transportability of auto parts. It is very unlikely that equity stakes affect this instrumental variable.

However, this distance IV has a drawback. It can be defined only for dyads with transaction links. Since dyads without transaction links constitute a large part of the sample, sample size decreases when distIV_{ast} is applied in the 2SLS estimations. The estimation results are not comparable with the other ones.

2.4 Results

This section demonstrates the empirical results. The estimates of Specification 2.3.1 *without* supplier fixed effects are described first, followed by the set of results that control for supplier fixed effects. The first set of results are explained in detail. This is because the instrumental variable for variety_{ast} has limited time-series variation, which makes variety_{ast} instrumented hard be identified precisely in all 2SLS with supplier fixed effects. Specifications without supplier fixed effects are necessary and are applied first.

2.4.1 Results Without Supplier Fixed Effects

This subsection shows the estimation results of Specification 2.3.1 without supplier fixed effects. The analysis starts with simple regressions which portray the partial correlations between key variables and $\text{own}\%_{ast}$. Then in order to test whether or not the three key variables have causal effects, two-stage least squares estimations are carried out using

the instruments for the key variables.

Table 2.6: Linear Models of Equity Stakes, Without Supplier Fixed Effects

Y: Own$\%_{ast}$	(1)	(2)	(3)	(4)	(5)
Sample	All	All	All	with links	with links
Instruments	No IV	No IV	Two IVs	No IV	Three IVs
Customer$_{st}$	-0.594^a	-0.602^a	-1.243^a	-0.959^a	-2.054^a
(supplier customers)	(0.065)	(0.068)	(0.272)	(0.122)	(0.434)
Ln (variety$_{ast}$ + 1)	4.998^a	4.946^a	3.997^a	4.706^a	1.594
(log dyad varieties)	(0.518)	(0.522)	(1.110)	(0.772)	(1.264)
Ln (distance$_{ast}$+1)	-0.522^a	-0.561^a	-0.744^a	-0.956^a	0.022
(log minimal dyad distance)	(0.158)	(0.163)	(0.194)	(0.367)	(1.345)
Ln L$_{st}$		0.063	0.745^a	0.011	1.461^a
(log supplier employees)		(0.087)	(0.213)	(0.293)	(0.523)
Ln (K_{st}/L_{st})		0.133^c	0.465^a	0.265	0.432
(log supplier capital intensity)		(0.071)	(0.154)	(0.275)	(0.362)
Age$_{st}$		-0.025^a	-0.015	-0.060^b	-0.040
(supplier's age)		(0.008)	(0.011)	(0.025)	(0.029)
Constant	3.861^a	4.762^a	3.238^a	10.871^a	5.733
	(0.845)	(1.071)	(1.101)	(2.959)	(5.503)
N	15036	14399	14388	4648	4647
R^2	0.126	0.132	0.062	0.160	0.084
RMSE	6.902	6.871	7.140	11.425	11.908

Note: Dependent variable is the equity stake of a supplier held by an assembler in year t. All year dummies and assembler dummies are controlled for in all specifications. Robust standard errors in parentheses are clustered at suppliers. ^{c} p<0.1,^{b} p<0.05, ^{a} p<0.01

Table 2.6 begins with the specification employing only the three variables of interest along with assembler and year fixed effects. In column 1, it shows that the coefficient on customer$_{st}$ is significantly negative, and the estimate of variety$_{ast}$ is significantly positive. Opposite to the two hypotheses about distance, distance$_{ast}$ appears to be negatively associated with own$\%_{ast}$. In column 2, in addition to those three key variables, I also control for the number of employees (L), capital intensity ($\frac{K}{L}$), and the age of suppliers (age). The inclusion of these variables helps alleviate omitted variable bias. Large suppliers

or suppliers with high capital intensity are able to produce a wide scope of auto parts. If assemblers tend to control influential suppliers with high equity stakes, the effect of variety could be overestimated when the number of employees and capital intensity of suppliers are not controlled for. In the second column, the coefficient of variety is slightly smaller in magnitude than the corresponding coefficient in column 1, and the coefficient of the number of employees is also significantly positive. These two points corroborate the concern about omitted variable bias. Besides L and $\frac{K}{L}$, Ahmadjian and Oxley (2006) suggest that supplier age could also be an important factor affecting equity stakes.[27] The coefficient of age in column 2 is significantly negative. Comparing the results in columns 1 and 2, it can be seen that the estimates of key variables do not change substantially either in magnitudes or signs as more control variables are included in the specification.

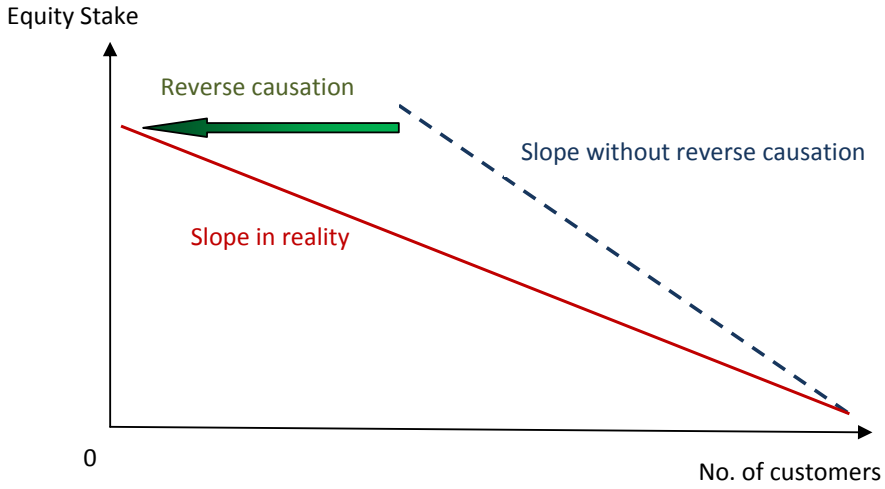
Turning to the variables of interest, the full specification in column 2 shows that equity stakes decrease by 1.86 percentage points when the number of customers increases by one standard deviation, ceteris paribus. Keep other things constant, equity stakes increase by 2.77 percentage points when $\ln(\text{variety}_{ast}+1)$ increases by one standard deviation. Although the magnitude sounds limited, in comparison to the average equity stakes of 1.1 percentage points, the effect is large. These results give strong support for the first two hypotheses: (1) equity stakes are reduced if suppliers serve multiple customers; (2) the scope of auto parts purchased from a supplier increases the producer's equity stake in that supplier.

The following columns of Table 2.6 focus on causality tests. In column 3, two key variables are instrumented—customer and variety.[28] Distance may be subject to simultaneity bias as well, so it is important to see the results when all three key variables are instrumented, because as long as one variable in the specification is endogenous, the

[27] Younger suppliers are more vulnerable to assembler opportunism as they are less able to absorb losses due to order cancellation or the like, so assemblers need to hold equity stakes to alleviate younger suppliers' concerns.

[28] The estimation results when customer and variety are instrumented individually are similar to the ones shown in column 3 of Table 2.6. They are available from the author.

Figure 2.5: Difference in Magnitudes



estimates of all variables are inconsistent. However, as mentioned in the discussion of distIV$_{ast}$, distIV is available only for supplier-assembler dyads with positive transaction links. In other words, once distance is instrumented by distIV, the sample size decreases by a large amount, and the estimation results become not comparable with previous ones. Therefore, before turning to the regressions with all three endogenous variables instrumented, it is good to check the results based on the full sample but only with customer and variety instrumented.

In column 3 of Table 2.6, the coefficient of customer is significantly negative, which shows that the number of customers that a supplier serves has a direct and negative impact on the size of equity stakes. The absolute value of its coefficient doubles in comparison to the one in column 2. Figure 2.5 illustrates why the influence of customer on equity stakes is lower without instrumenting. The dashed line indicates the "true" relationship, where the negative relationship portrays that the assembler needs to exercise less control when the number of customers increases. As mentioned in the hypotheses section, when the number of customers grows, the supplier is willing to exert high effort because her outside option is better. However, when reverse causation is present, an assembler with high equity stakes can compel the supplier to trade with fewer other

assemblers. The ability to distort the number of customers declines as the equity stake falls.

In column 3 of Table 2.6, the coefficient of variety is still significantly positive, which implies that variety has a positive causal effect on the size of equity stakes. In this column, distance has not been instrumented by its instrumental variable, and it is still negatively associated with own%.

Now turning to last two columns of Table 2.6, only the supplier-assembler dyads with positive transaction links are utilized in these two columns. In order to provide a baseline for the results when the three key variables are all instrumented, column 4 of Table 2.6 reports the simple regression results obtained from this small sample. The coefficients of the three key variables are similar to the estimates in columns 2 where a full sample is used. Customer and distance are negatively associated with own%, and the coefficient of variety is significantly positive. These results imply that, the partial correlations between the three key variables and own% conditional on the existence of transaction links are similar to the ones in the full sample.

The last column of Table 2.6 presents the estimation results when all three key variables have been instrumented. This result not only sheds light on the test of distance's causal effect, but also helps check the robustness of the results as to customer and variety. In column 5, the coefficient of customer is significantly negative at the 1% level, which strongly supports the hypothesis that equity stakes are reduced if suppliers serve multiple auto producers. As for the scope of auto parts, it remains positive and significant at 20% level. This significance level is usually not acceptable; however it still seems reasonable to believe the positive effect of variety since three variables have all been instrumented and efficiency is substantially lost.

In the last column of Table 2.6, the coefficient of distance changes substantially after it has been instrumented by distIV.[29] It becomes statistically insignificant. This result

[29]The estimation results when distance_{ast} is instrumented alone are similar to the ones shown in column 5 of Table 2.6. They are available from the author.

shows that the significantly negative coefficients shown previously are driven by reverse causation. After removing this reverse causation, distance does not appear to have a direct impact on the size of equity stakes.

As mentioned in the hypotheses section, the estimation result of distance is used to differentiate the underlying assumptions about distance costs. Under the assumption that distance has a negative effect only on the *marginal effect of the supplier's effort*, it is predicted that distance increases the size of equity stakes. Under the assumption that distance decreases only the *level of quasi-rents* rather than the *marginal effect* of the supplier's effort, the model predicts that distance has no effect on the size of equity stakes. The insignificant coefficient of distance shown in the last column of Table 2.6 appears to support the latter assumption.

2.4.2 Results With Supplier Fixed Effects

Table 2.7 presents the estimation results of specifications with supplier fixed effects. The first two columns give qualitatively same results as the first two columns of Table 2.6. Customer and distance are negatively associated with own%, and the coefficient of variety is significantly positive. In the third column of Table 2.7, both customer and variety are instrumented. Even though both customer and variety have the right signs as predicted, none of the key variables are statistically significant. The insignificant results in column 3 are not surprising to us. VarIV has limited time-series variation and cannot be identified precisely in models with supplier fixed effects. This insignificant result is driven simply by the identification problem related to varIV. When only customer is instrumented by its instrumental variable, customer's coefficient is -1.559, and it is statistically significant at the 5% level. In column 4 of Table 2.7, only the supplier-assembler dyads with positive transaction links are used. The simple regression gives the result quite similar to the estimates in column 2 of the same table. The last column of Table 2.7 reports the estimation results after instrumenting for three key variables.

Table 2.7: Linear Models of Equity Stakes, With Supplier Fixed Effects

Y: Own%$_{ast}$	(1)	(2)	(3)	(4)	(5)
Sample	All	All	All	with links	with links
Instruments	No IV	No IV	Two IVs	No IV	Three IVs
Customer$_{st}$	-0.386^a	-0.383^a	-1.680	-0.313^b	-8.201
(supplier customers)	(0.050)	(0.050)	(1.098)	(0.129)	(6.984)
Ln (variety$_{ast}$ + 1)	5.462^a	5.424^a	6.845	5.351^a	7.531
(log dyad varieties)	(0.529)	(0.538)	(4.883)	(0.635)	(5.884)
Ln (distance$_{ast}$+1)	-0.656^a	-0.695^a	-0.513	-0.742^a	3.796
(log minimal dyad distance)	(0.173)	(0.179)	(0.640)	(0.217)	(6.934)
Ln L$_{st}$		0.523^b	0.691^b	2.091^b	5.528
(log supplier employees)		(0.222)	(0.297)	(0.846)	(3.765)
Ln (K_{st}/L_{st})		-0.046	0.165	-0.386	1.008
(log supplier capital intensity)		(0.078)	(0.170)	(0.289)	(1.481)
Age$_{st}$		-0.021^a	0.009	-0.008	0.268
(supplier's age)		(0.006)	(0.022)	(0.026)	(0.258)
Constant	3.727^a	1.349		-13.145^b	
	(0.876)	(1.661)		(6.518)	
N	15036	14399	14388	4648	4593
R^2	0.144	0.148	0.132	0.201	-0.658
groups	552	520	519	517	462
S.E of u_i	2.221	2.302		18.038	
S.E.of e_{it}	6.712	6.688	6.747	6.308	9.071
RMSE	6.588	6.566	6.747	5.946	9.071

Note: Dependent variable is the equity stake of a supplier held by an assembler in year t. Observations are grouped for each supplier. All year dummies and assembler dummies are controlled for in all specifications. Robust standard errors in parentheses are clustered at suppliers. ^{c} p<0.1,^{b} p<0.05, ^{a} p<0.01. Columns 3 and 5 are estimated by xtivreg2 in STATA.

Similar to column 3, no key variable is statistically significant although the signs are consistent with the hypotheses.

In summary, two-stage least squares estimations with key variables instrumented give good evidence for the first two hypotheses. They show that the number of customers that a supplier serves and the scope of auto parts transacted between suppliers and assemblers both have direct causal effects on the size of equity stakes, and the direction of the effects are also consistent with the predictions of the analytic model. Distance appears to have no causal effect on the size of equity stakes. This empirical result implies that the assumption that distance reduces only the *level* of quasi-rents is consistent with the data in this study.

2.4.3 Results from Panel Data Fractional Response Variable

The results discussed above are obtained from models assuming linear functional forms that neglect the fact that equity stakes are bounded between 0 and 100%. This nature determines that the marginal effects of key variables are not likely to be constant. For example, when the equity stakes are very low, it is unlikely that the negative effect of the number of customers on equity stakes is as strong as that when the equity stakes are high. The previous linear estimate could then give misleading results since zero equity stakes represent a large proportion of the sample. In order to deal with this bounded nature, the panel data fractional probit proposed by Papke and Wooldridge (2008) is implemented. This method takes into account not only the bounded nature of equity stakes but also unobserved firm heterogeneity.[30]

[30]Papke and Wooldridge (1996) propose a direct model for the conditional mean of fractional response variable that keep the predicted values in the unit interval. It solves the problems associated with the log-odds ratio method. The responses at the corners, zero and one, are properly dealt with. Papke and Wooldridge (2008) study how to estimate fractional response models for panel data with a large cross-sectional dimension and relatively few time periods. In this context, incidental parameter problem will be incurred if fixed effects are estimated. The time-constant unobserved is allowed to be arbitrarily correlated with explanatory variables. Papke and Wooldridge (2008) extend their previous estimation method with two building blocks–Chamberlain device and control function approach. Chamberlain (1980) device deals with unobserved heterogeneities. The distribution of heterogeneities given

Table 2.8: Panel Data Fractional Probit Models

Y: Own_{ast}	(1)		(2)		(3)	
Sample	All		All		with links	
Instruments	No IV		Customer		Distance	
	Coef.	APE	Coef.	APE	Coef.	APE
Customer_{st}	-0.072^b	-0.138^a	-0.916^a	-1.764^a	-0.110^a	-0.603^a
(supplier customers)	(0.035)	(0.033)	(0.199)	(0.411)	(0.031)	(0.186)
Ln ($\text{variety}_{ast}+1$)	1.314^a	2.531^a	1.444^a	2.780^a	0.889^a	4.868^a
(log dyad varieties)	(0.046)	(0.212)	(0.082)	(0.243)	(0.198)	(1.136)
Ln (distance_{ast}+1)	-0.141^a	-0.272^a	-0.121^a	-0.233^a	-0.061	-0.337
(log minimal dyad distance)	(0.025)	(0.078)	(0.037)	(0.074)	(0.307)	(1.679)
Ln L_{st}	0.114	0.220	0.277	0.533	0.202	1.104
(log supplier employees)	(0.183)	(0.220)	(0.127)	(0.236)	(0.127)	(0.726)
Ln (K_{st}/L_{st})	-0.006	0.012	0.153	0.295	-0.020	-0.112
(log supplier capital intensity)	(0.092)	(0.074)	(0.053)	(0.096)	(0.045)	(0.258)
Age_{st}	-0.017	-0.033	-0.009	0.017	-0.019	-0.106
(supplier's age)	(0.020)	(0.042)	(0.024)	(0.044)	(0.026)	(0.178)
$v^2_{customer}$			0.919^a			
			(0.202)			
$v^2_{distance}$					0.021	
					(0.305)	
N	14399		14399		4648	
Scale Factor	0.019		0.019		0.055	

Note: Dependent variable is the equity stake of a supplier held by an assembler in year t. Observations are grouped for each supplier. All year dummies and assembler dummies are controlled for in all specifications. The standard errors in parentheses are fully robust. Standard errors in the last two models are obtained by bootstrapping the suppliers using 120 replications. ^{c} p<0.1,^{b} p<0.05, ^{a} p<0.01. In column 1, the time averages of all suppliers' characteristic variables (customer, variety, distance, number of employees, capital intensity, and age) are included. In column 2, the time averages of customer's instrument and all suppliers' characteristic variables except for customer are included. In column 3, the time averages of distance's instruments, customer, variety, number of employees, capital intensity, and age variables are included. The standard errors for the APEs are obtained from bootstrapping the suppliers using 120 replications.

Table 2.8 reports the estimation results. It starts with a scenario assuming all variables are exogenous. Column 2 and 3 then show the estimates when the variables customer and distance are individually instrumented. No result with the variable variety instrumented is reported because variety's instrumental variable has too little time-variation to be identified. In column 1, besides the variables in equation 2.3.1, time averaged key variables and suppliers' characteristics (employees, capital intensity, and age) are also included to capture firm heterogeneity when a pooled QML model is used to estimate equity stakes measured as fractions. In column 2, the variable customer is treated as endogenous and instrumented.[31] Variety, distance, suppliers' characteristics, and the instrument of the number of customers are all averaged over time and properly included in each stage of the control function approach. In the last column of Table 2.8, a similar method is applied to the variable distance. Over the three models, their estimated coefficients are not average partial effects (APEs), and they cannot be directly compared with the estimates from linear models. In Table 2.8, APEs are also calculated and shown.[32]

Results in Table 2.8 confirm the conclusions drawn from linear models. Column 1 shows that when all three key variables are treated as exogenous, the coefficients for the variables customers and distance are significantly negative and variety is positive. In column 2, the estimates give strong support for the statements that the variable customer is endogenous. The coefficient for customer's residual is significant at the 1% level. This endogeneity test dates back to Hausman (1978). The results in column 2 also confirm that the variable customer reduces the size of equity stakes even after controlling for its

explanatory variables is assumed to be a normal distribution with a function of averaged explanatory variables as the conditional mean. Control function approach is called upon when there is an endogenous explanatory variable. Intuitively, the endogeneity problem disappears after a generalized residual term from the first stage is controlled for. Key specifications in Papke and Wooldridge (2008) are equation 2.8 for exogenous cases and equations 4.4 and 4.7 for models with endogenous explanatory variables.

[31]The panel data fractional probit model can only be used when the endogenous variable is roughly continuous. The variable customer ranges from 1 to 11. It roughly satisfies this requirement.

[32]The formula for scale factors in Papke and Wooldridge (2008) are equations 2.11 and 4.10 respectively for exogenous and endogenous cases.

residual, i.e. the part of the variable customer which is correlated with the error term. In column 3, the variable distance is individually instrumented. Similar to the results from the linear models, distance does not appear to have a causal effect on the size of equity stakes after distance's endogeneity issue is taken into account. The results in column 3 do not provide strong evidence for the concern about the endogeneity of distance. The coefficient for distance's residual is not significant, which might be attributed to the small sample size involved in the regression in column 3.

The magnitude of APEs from the panel data fractional probit model in which all variables are assumed to be exogenous are smaller than the estimates from the linear models. Each APE is calculated as the product of the coefficient and scale factor.[33] In order to compare the results from linear models, I multiply the product of these two items by 100 because equity stakes are measured as fractions in Table 2.8 and percentages in previous linear models. In column 1, the scale factor equals 0.019. Customer's APE equals -0.138 with a standard error at 0.033.[34] It is significantly negative. This APE is less than a half of the estimate from the linear model (-0.383 from column 2 in Table 2.7). The coefficients for variety and distance are also smaller than their estimates from the linear model. This difference could be attributed to the large proportion of extremely low equity stakes.

As mentioned in the beginning of this section, it is hard for linear models to account for nonlinear marginal effects due to the bounded nature of fractional response variables. In order to demonstrate a whole picture, I graph customer's APEs for the full range of the number of customers.[35] Figure 2.6 presents the APEs and their corresponding 95% confidential intervals. The estimates of two models are drawn. One model assumes all variables are exogenous; the other model takes the variable customer as endogenous. In

[33]The single scale factor here is calculated as $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(\widehat{\psi}_a + x_{it}\widehat{\beta}_a + \bar{x}_i\widehat{\xi}_a)$.

[34]Its standard error is obtained by bootstrapping the suppliers using 120 replications.

[35]At each fixed value of customer, all the other explanatory variables in the model are averaged across.

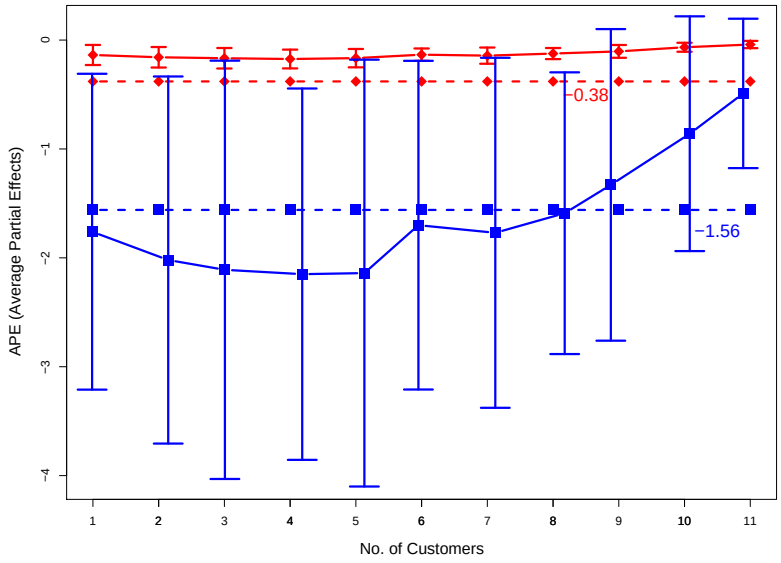


Figure 2.6: APEs, Exogenous and Endogenous

figure 2.6, the short red lines at the top demonstrate the APEs and confidential intervals for the exogenous case; the other set of long blue lines are drawn for the endogenous model. The APEs are denoted by the solid squares in the middle of intervals. In the endogenous case, the figure shows that customer's negative impact keeps strengthening when the number of customers increases from 1 to 5 and then diminishes as the number of customers continues increasing. When the number of customers is large enough, the APE becomes insignificantly different from zero. As to the exogenous model, its APEs actually share the same pattern, even though it is hardly seen given the metric in figure 2.6. The magnitudes of APEs increases first and then decreases as the number of customers increases.

The two dotted lines in figure 2.6 present the coefficients of the number of customers estimated from linear models. In the exogenous case, the coefficient from the linear model is much larger in magnitude than the APEs from the nonlinear model. This difference could be driven by the fact that extremely low equity stakes account for a large proportion of the whole data set. As to the endogenous case, the standard errors increase substantially. The coefficient from a linear model is located within the confidential in-

tervals of the APEs. Overall, both the linear and nonlinear estimates support that the variable customer has a negative and significant impact on the size of equity stakes.

In summary, the bounded nature of equity stakes implies that the marginal effects of key variables are not likely to be linear, and it is hard for linear models to capture the whole story. The panel data fractional probit model proposed by Papke and Wooldridge (2008) address this concern. Its results confirm the conclusions drawn from linear models. Meanwhile, it also demonstrates the average partial effects vary across the range of the explanatory variables.

2.5 Conclusions

This paper studies the determinants of the size of equity stakes that downstream producers hold in their upstream input suppliers. It extends a property-rights model and formally predicts three characteristics' casual effects on the likelihood of backward integration. In its analytic part, this paper introduces geographic distance into the study of vertical integration and predicts the effect of distance on the likelihood of backward integration. Since different assumptions about distance costs give different predictions, empirical analysis is used to determine which one of the assumptions is consistent with the data. In the empirical part, this paper contributes to the literature by focusing on the *causal* effects of key variables and addressing the bounded nature of the fractional dependent variable. Testing the causal effects of key variables presents an econometric challenge—reverse causation. The control rights associated with equity stakes allow the downstream producers who hold the equity stakes of suppliers to manipulate the suppliers' operating decisions and then change the characteristics studied in the hypotheses. To address this concern, three instrumental variables—one for each key variable—are utilized. Panel data fractional probit estimator is implemented to deal with the bounded nature of the fractional dependent variable.

This paper draws the following conclusions: (a) the number of customers that a supplier serves has a significant negative causal effect on the size of equity stake that downstream producers hold in the supplier; (b) the scope of auto parts involved in a vertical supply relationship has a significantly positive effect on the size of equity stakes; and (c) geographic distance between a supplier and a customer does not appear to have a causal effect on the size of equity stakes. This last result implies that distance decreases the total output of the relationship by decreasing the part of the output which is unrelated to the marginal effect of the supplier's effort. Last, the firm-level data from the Japanese auto parts industry fit the model built by Acemoglu, Aghion, Griffith, and Zilibotti (forthcoming) very well, which suggests that this model framework could be useful in the study of vertical integration in other contexts.

Bibliography

Acemoglu, D., P. Aghion, R. Griffith, and F. Zilibotti. Forthcoming. Vertical Integration and Technology Theory and Evidence. *Journal of European Economic Association*.

Ahmadjian, C. and J. Oxley (2006). Using Hostages to Support Exchange: Dependence Balancing and Partial Equity Stakes in Japanese Automotive Supply Relationships. *Journal of Law, Economics, and Organization 22*(1), 213–233.

Che, Y.-K. and J. Sákovics (2008). *The New Palgrave Dictionary of Economics, hold-up problem*, Chapter hold-up problem. DOI:10.1057/9780230226203.0743: Palgrave Macmillan.

Grossman, S. and O. Hart (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Economy 94*, 691–719.

Hart, O. and J. Moore (1990). Property Rights and the Nature of the Firm. *Journal of Political Economy 98*, 1119–1158.

Head, K. and J. Ries (2008). Fdi as an Outcome of the Market for Corporate Control: Theory and Evidence. *Journal of International Economics 74*(1), 2–20.

Holmes, T. (1999). Localization of Industry and Vertical Disintegration. *The Review of Economics and Statistics 81*(2), 314–325.

Lafontaine, F. (1992). Agency Theory and Franchising: Some Empirical Results. *Rand Journal of Economics 23*, 263–283.

McLaren, J. (2000). Globalization and Vertical Structure. *American Economic Review 90*(5), 1239–1254.

Minkler, A. (1990). An Empirical Analysis of a Firm's Decision to Franchise. *Economics Letters 34*, 217–223.

Oxley, J. (1997). Appropriability Hazards and Governance in Strategic Alliances: A Transaction Cost Approach. *Journal of Law, Economics, and Organization 13*, 387–409.

Papke, L. and J. Wooldridge (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics 11*, 619–632.

Papke, L. and J. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*, 121–133.

Pisano, G. (1989). Using Equity Participation to Support Exchange: Evidence from the Biotechnology Industry. *Journal of Law, Economics, and Organization 5*, 109–126.

Whinston, M. (2001). Assessing the Property Rights and Transaction Cost Theoryies of Firm Scope. *American Economic Review, Papers and Proceedings 91*, 184–188.

Whinston, M. (2003). On the Transacton Cost Determinants of Vertical Integration. *Journal of Law, Economics, & Organization 19*(1), 1–23.

Williamson, O. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications.* Free Press, New York.

Williamson, O. (1985). *The Economic Institutions of Capitalism.* Free Press, New York.

Chapter 3

Multinational Retailers

3.1 Introduction

Since the 1990s, large multinational retailers have increasingly entered the Chinese market. They not only sell products to Chinese consumers, but also use China as an important global purchasing base.[36] Wal-Mart (US-based, the largest retailer in the world), Carrefour (France-based, the second largest retailer worldwide and the largest in Europe), Metro AG (German-based, the fourth largest retailer) and Tesco (UK-based, the fifth largest retailer) are among these multinational retailers.[37] Wal-Mart and its suppliers "in 2004, landed Chinese-made goods in the United States with a wholesale value of $18 billion."[38] In 2001, 61% of Carrefour's purchases in Asia were Chinese products. In the same year, Chinese products accounted for 65% of all Asian products in Metro's global purchasing network.[39] Products labelled "made in China" appear to obtain more opportunities in overseas markets by entering multinational retailers' procurement systems. In 2002, *People's Daily Online*,[40] an official newspaper in China, stated that: "the direct supply of 'made in China' products to foreign chain groups had become a key channel for their entry into the global market."

More recently, these large multinational retailers have shown an interest in entering

[36] A version of this chapter will be submitted for publication. Head, K., Jing, R., and Swenson, D. From Beijing to Bentonville: Do Multinational Retailers Link Markets?

[37] Deloitte 2006 Global Retailing Powers Study. The fourth largest retailer Home Depot, is classified as a specialist retailer.

[38] *The Wal-Mart Effect* (p. 102)

[39] *People's Daily Online* in April, 2002

[40] April 22, 2002

the Indian retail market. In its negotiations with the Indian government, Wal-Mart argued that "at present it has 120 people in Bangalore, buying about ten categories of Indian products for its shops (clothing, textiles, shoes, leather wallets, stationery and so on). Were it to have outlets in India, its procurement would naturally increase. (The Metro operation in Bangalore buys 98% of its stock locally.) Suppliers would become familiar with its requirements, and exports would also climb. In China, Wal-Mart accounts for nearly 10% of all exports to America."[41]

This paper subjects Wal-Mart's argument to stringent econometric tests as we investigate whether the presence of multinational retailers in China increases China's exports; and if so, what mechanisms tie retailer presence to the promotion of China's exports? Two potential mechanisms are proposed and tested. First, it is possible that multinational retailers increase Chinese exports through the creation of linkage connections with distant markets. We call this mechanism the multinational retailers' "linkage effect." It predicts that multinational retailers increase exports from a city in China to *specific* countries where the multinational retailers have stores. Alternatively, the presence of retailers may improve local Chinese suppliers' *general* export capabilities by stimulating their productivity growth. We call this mechanism the multinational retailers' "capability effect." This effect predicts that Chinese cities in which multinational retailers are located benefit from an enhanced ability to export to *all* destination countries, no matter whether the destination countries host retail outlets of the multinational or not.

This paper tests these hypotheses using export data from 35 Chinese major cities to 50 countries over the period 1997–2005. We expect the pro-trade effect of multinational retailers to be confined to the final consumer goods that are sold by stores such as Wal-Mart. We use HS4 (harmonized classification of goods by 4-digit codes) to classify goods as belonging to one of two categories—retail and non-retail goods. The retail goods category includes all HS4 product categories that are sold in multinational retail-

[41] Economist (2006), April 15.

ers' stores; all remaining HS4 product categories are assigned to the non-retail goods group. Within each group, we generate the export value for the group by summing the export value of all HS4 belonging to the group. The export value of retail goods is taken as the dependent variable in this paper. The measure of linkages due to exposure to the world's four largest non-specialist multinational retailers—Wal-Mart, Carrefour, Metro and Tesco—represents the density of linkages at the city-country level formed by multinational retailers. In addition to retail stores in China, each of the multinational retailers has set up global procurement centers in China. Consequently, we form linkage variables which reflect both types of retailer activity. Only exports of retail goods are considered.

We use standard gravity models and models with origin–destination dyadic fixed effects to test whether Chinese exports are positively related to the presence of multinational retailers. We find that the presence of multinational retailers is associated with bilateral trade above the level predicted by gravity models. Furthermore, this effect remains even after we control for dyadic fixed effects, which means that the identification of retailers' pro-trade effects comes solely from the variation of key variables within dyads over years. With regard to the multinational retailers' linkage effects, we use the stringent specification for bilateral trade suggested by Baldwin and Taglioni (2006) which includes dyadic, origin–year, and destination–year fixed effects. Notably the positive and significant association between retailer presence and city exports disappears when we control for city–year fixed effects. To understand why the stringent specification that includes origin–year fixed effects, absorbs the apparent linkage effects of multinational retailers, we conduct further tests to learn whether retailer presence create benefits that are manifested in city export capabilities. Therefore, we conduct a second set of exercises which test whether the estimated origin–year fixed effects are positively related to the origin city's proximity to global procurement centers and to the number of retailer outlets. Controlling for the current development of a city, with origin and year fixed effects, we find that as a city's distance to global procurement centers decreases by 10%, the

city's general export capability increases by 2.4%. Similarly, as the number of outlets in the city increases by 10%, the origin city's export capability increases by 2.36%.

Our paper contributes to three strands in the literature. The first is the effect of networks on trade. Ethnic groups and immigrants have been shown to have positive and significant effects on bilateral trade. Gould (1994) and Head and Ries (1998) for US and Canadian trade, respectively, find immigrants promote trade with the immigrants' countries of origin. Similarly, Rauch and Trindade (2002) uncover a positive relationship between trade and overseas Chinese populations.[42] Our paper extends the literature to the networks generated by multinational retailers.

Our paper is also linked to the large literature which examines the relationship between foreign direct investment and exports. Head and Ries (2004) survey the literature on this topic and identify the economic mechanisms linking foreign direct investment and exports through the theories of the multinational corporations. Swenson (2008) shows that growth in the presence of multinational firms is positively associated with the formation of new trade by local Chinese firms and suggests that information spillovers may drive this result. The multinational retailers studied in this paper create an infrastructure of investment as they set up their outlets and procurement centers in China. As a result, these retailers may ship "made in China" goods through their purchasing systems for sale in their worldwide outlets. Thus, this paper provides another example of the potential complementarity between foreign direct investment and trade.

Finally, although the retail industry and the evolution of international trade are both popular topics of current attention, research on the effects of multinational retailers on international trade is quite limited. Basker and Van (2007) derive a theoretical model and show that large retailers have stronger incentives than small retailers to import from distant, low-cost countries because large retailers enjoy economies of scale in both retailing and importing. Basker and Van (2008) find that large US retailers have a

[42] Rauch (2001) is a survey paper on this literature.

higher marginal propensity to import relative to small US retailers for every one dollar's increase in sales. In other words, these two papers provide evidence regarding the impact of multinational retailers on the imports of *retailers' home countries*. In contrast with this work on multinational retailers and international trade, our work takes the perspective of countries hosting multinational retailers and focuses on the retailers' impact on *exports* from retailers' host countries.

In related work Javorcik and Li (2007) examine on the impact of Wal-Mart's entry on host country firm *productivity* . Their work provides empirical evidence that Wal-Mart has a positive effect on Romanian supplier firm productivity. The mechanisms for such productivity improvements are illustrated in the Javorcik et al. (2006) case study discussion about the effects of Wal-Mart's entry on the soap, detergent and surfactant industry in Mexico. Since our work focuses on the impact of multinational retailers on host country *exports* our results provide insight to *host country* policy makers about the connection between multinational retailers and local exports. In essence, we test the argument that Wal-Mart has put forth in seeking to convince the Indian government that it should open its retail market.

The remainder of the paper is structured as follows. Section 3.2 describes the procedures followed by Metro and Wal-Mart when they export Chinese products to their outlets outside China. Section 3.3 discusses the hypotheses based on two potential mechanisms. Specifications and key variables are explained in section 3.4, while data are described in section 3.5. Sections 3.6 and 3.7 show and discuss the econometric results. We conclude in section 3.8.

3.2 Background

In this section, we provide an overview of the purchasing procedures followed by multinational retailers in China. This information not only provides a background for this

paper but also explains why we think that lagged rather than contemporaneous measures of retailer linkages are most appropriate for examining the effects of multinational retailers on China's exports. This section also highlights significant elements that influenced China's deregulation of its retail market, as these factors help to explain why the four multinational retailers chose distinct locations within China as the base of their operations, and the point from which they spread. This is one source of variation that we exploit in identifying how multinational retailer linkages affect international export development.[43]

3.2.1 Procurement Procedures

In this subsection, we discuss the export procedures of Metro, followed by those of Wal-Mart. While there is little public information on the other two retailers, there is no strong reason to believe that their procurement practices should differ dramatically from those of Metro and Wal-Mart.

Metro uses the Chinese domestic market as a testing ground for Chinese products to be exported. "Metro AG emphasizes the quality of products. The products which have entered Metro's procurement system will be sold in Chinese regional markets first. If they sell well, the products will be sold in all stores within China. At last they will be sold globally."[44] This statement suggests that the more stores Metro has in China, the more Chinese products are likely to be exported through Metro's procurement system.

Wal-Mart Global Procurement (WMGP)[45] facilitates the purchase of Chinese goods by Wal-Mart's outlets in the world. WMGP regularly finds new products and inquires about prices.[46] After making some simple classifications, WMGP sends the information

[43]The other source of variation that we use is based on the differential presence of the four retailers in different export destinations.

[44]Wu (2004), p. 74.

[45]"Wal-Mart began purchasing merchandize from China in the 1980s. Until the late 1990s, all its purchases were made through intermediary companies. After it moved its global purchasing center to Shenzhen in 2001, [...] both its direct and total purchases increased steadily." Wang and Zhang (2006)

[46]The following procedures are summarized from an article on www.jamoo.net.

by email to the buyers of all Wal-Mart outlets. The outlet buyers then decide what kind of products their stores may need and which should be explored in their "buying trips." This procedure is time consuming. Outlet buyers routinely meet two to three times each year in China. Before the buyers arrive in China, WMGP prepares the required samples. The staff at WMGP mark the product's price and features on the sample but cover the manufacturer's name. In other words, the outlet buyers do not know the manufacturer of samples when they make their initial purchasing decision. During the meeting, the buyers decide which products to buy. The staff at WMGP do not give the buyers much input before the buyers make decisions. Next, the buyers and the WMGP staff discuss privately the price and other details of orders. Afterwards, WMGP contacts manufacturers and starts negotiations. The outlets' buyers have little or no direct contact with manufacturers. Once the order is made, WMGP handles the logistics of the order. They check the factory (called "FC") to make sure there are no child workers, no excessive overtime, etc. They check product quality (called "QA") at least twice—once during production and once after production. In addition, WMGP contacts shipping companies such as the Maersk Line, and prepares export documents, for instance, letters of credit.

The anecdotal details from these procurement procedures suggest that Chinese local suppliers could benefit from the presence of global procurement centers of multinational retailers. Local Chinese suppliers could try out their products for multiple foreign markets without paying market-specific fixed entry costs in advance. In addition through the interaction with global procurement centers, Chinese suppliers would learn some international standards on quality or safety issues and the general preferences of foreign consumers. All these benefits suggests that the presence of global procurement centers in China may have a positive impact on China's exports.

3.2.2 Deregulation in the Chinese Retail Market

This subsection summarizes the history of deregulation in China's retail market, focusing on regulatory factors that caused different retailers to enter in different Chinese locations. Carrefour made the first entry of the retailers with a hypermarket[47] located in Beijing at the end of 1995. In 1996, Wal-Mart established its first super-center in Shenzhen, while Metro opened its first cash and carry in Shanghai. Finally, Tesco entered the Chinese retail market when it set up its first store in Shanghai in 1998. It is important to understand why the retailers chose distinct entry locations in China, since location variation is a key component that is used to identify the importance of retailer linkages.

Wang and Zhang (2006, p. 295) list important policy changes in the Chinese retail market, which we summarize in Table 3.1. Two points in the table are worth emphasizing. First, during the period from 1992–1995, foreign retailers were allowed to operate in only 11 designated areas, and each area was permitted to host only one or two foreign retailers (including the retailers with headquarters in Hong Kong, Taiwan, and Macau, which were treated as foreign by the government). This restriction helps explain why Wal-Mart, Tesco, and Carrefour chose different entry locations when they first entered China. The second important point highlighted by Table 3.1 is the fact that majority foreign-owned joint ventures which involved retailer chains between 1999–2004 were only conditionally permitted. The condition was that large quantities of domestically-made goods be exported through the retailer's distribution channels. This requirement meant that multinational retailers had another direct incentive to export Chinese products through their procurement systems—in order to obtain more freedom for their operations in the Chinese *retail* market. All these restrictions were lifted after 2005.

[47] A hypermarket or superstore is a retail self-service establishment offering a broad range of food and non-food products. (source: `http://stats.oecd.org/glossary/detail.asp?ID=6250`)

Table 3.1: Deregulation of Foreign Retailers in Chinese Retail Market

Restrictions	1992–1995	1995–1999	1999–2004	After 2005[3]
Geographic	6 cities and [1] 5 Special [2] Economic Zones	–	provincial capitals	lifted
Operational				
retail/wholesale	retail only	wholesale allowed	–	lifted
single/chain	single only	chain only allowed in Beijing	chain	lifted
Percentages of Goods purchased within China	$\geq 70\%$	–	–	lifted
Equity	joint venture minority stake	–	joint venture chains with majority stake as conditionally allowed	lifted
Permission	entry must be approved by both state and local governments	–	–	only local gov't permit is needed

"–" implies no regulation is changed. 1.Beijing, Shanghai, Tianjin, Guangzhou, Dalian, and Qingdao. 2. Shenzhen, Zhuhai, Shantou, Xiamen, and Hainan. 3. Since December 11, 2004.

3.3 Hypotheses

Two potential mechanisms may cause the presence of multinational retailers to increase China's exports. First, multinational retailers could increase China's exports by creating linkage connections with distant markets. The managers in multinational retailers' purchasing centers or retail stores learn about nearby Chinese suppliers and convey the information to their stores outside China when they discover attractive products. Because information related to suppliers or products is systematically shared only with other affiliates in the retailer's global operations, this mechanism suggests that the presence of multinational retailers will increase the exports *only* for city–country dyads which are populated by same-firm retailer affiliates at each end. In other words, if firm-specific network linkages are important, we predict that multinational retailers will facilitate exports from Chinese cities where these retailers have stores or nearby global procurement centers. The increased exports will only go to the countries where these retailers have established stores. Each multinational retailer creates a "strategic network" when it provides its branches or subsidiaries with access to information, and resources (Gulati and Zaheer, 2000). We call this mechanism the multinational retailers' "linkage effects."

Therefore, we propose

Hypothesis 1 (linkage effects): The presence of multinational retailers in a Chinese city increases the city's exports to the countries where the multinational retailers have stores.

A second potential mechanism by which multinational retailers increase China's exports will operate if multinational retailers stimulate the productivity growth of local Chinese suppliers thus improving the general export capabilities of the Chinese suppliers. There are many avenues by which multinational retailers could stimulate the productivity of local suppliers. As Javorcik et al. (2006) and Javorcik and Li (2007) emphasize, the entry of multinational retailers is likely to increase the competitive pressures facing suppliers in host countries. This is generally true, since multinational retailers often have more bargaining power relative to other retailers in host countries. When multinational retailers require suppliers to lower prices or/and improve products, high-cost suppliers are driven out of the market, while suppliers that remain in operation improve their productivity by labor-shedding and innovation. As suggested by Javorcik and Li (2007) the entry of multinational retailers may further increase the productivity of suppliers, if the entry of multinational retailers introduces advanced retail technologies and international management practices. Local firm productivity is enhanced by this mechanism if suppliers reallocate their savings in distribution costs to production. The third channel for local supplier productivity increases comes into play if the activities associated with the multinational retailer allows local suppliers to achieve economies of scale. Each of these points support the argument that the presence of multinational retailers increases the productivity of local Chinese suppliers. We call this pro-productivity mechanism the "capability effect." When the productivity of suppliers is improved, we anticipate that a city will increase its exports to all destination countries, independent of the number of stores of the multinational retailer that the destination country hosts. Hence, we propose

Hypothesis 2(capability effects): The presence of multinational retailers in a Chinese city increases the city's exports to *all* countries.

There is another economic mechanism through which the presence of multinational retailers could increase the exports of a city to *all* destination countries. As the large multinational retailers purchase goods from China, firms in other foreign countries become aware that Chinese suppliers could manufacture high quality products at reasonable prices and also start importing them from China. The awareness of Chinese firms' productivities are improved. For convenience, we regard this mechanism as a part of the "capability effect" in this paper.

3.4 Estimation Strategies

In this section, we discuss the empirical methods we use to test the linkage and capability effects of multinational retailers. For each of these effects two issues are discussed: model specification and the construction of key variables.

3.4.1 Method for Testing Linkage Effects

It is widely recognized that the estimation of the common gravity model suffers from omitted variable bias. In particular, Anderson and van Wincoop (2003) point out that gravity models omit "multilateral resistance terms" which are needed to capture the general equilibrium conditions on exporting countries' outputs and importing countries' price level and expenditure. Baldwin and Taglioni (2006) summarize the theory underlying the gravity model and show that the unidirectional trade value can be modelled by the following equation:

$$V_{od} = \tau_{od}^{1-\sigma} \left(\frac{Y_o \cdot E_d}{\Omega_o P_d^{1-\sigma}} \right). \tag{3.4.1}$$

V_{od} is the total value of exports from country o to country d. τ refers to trade costs; σ is the elasticity of substitution among varieties in the CES utility function. Y_o is the exporting country's production output of traded goods. E_d is the importing country's expenditure on traded goods. P_d is country d's CES price index. $\Omega_o = \Sigma\left(\tau_{oi}^{1-\sigma}\frac{E_i}{P_i^{1-\sigma}}\right)$. This formula suggests that the specification for bilateral trade panel data should be

$$\ln V_{od,t} = \ln Y_{o,t} - \ln \Omega_{o,t} + \ln E_{d,t} + (\sigma - 1)\ln P_{d,t} - (\sigma - 1)\ln \tau_{od,t} \tag{3.4.2}$$

This equation decomposes log bilateral exports into three components. The first component, $\ln Y_{o,t} - \ln \Omega_{o,t}$, is origin-year specific. The second component, $\ln E_{d,t} + (\sigma - 1)\ln P_{d,t}$, is destination-year specific. The last component, $(\sigma - 1)\ln \tau_{od,t}$, is a time-varying dyadic term. Standard gravity equations model the origin and destination effects with the GDPs and per capita incomes of the origin and destination countries. Thus, they omit the Ω and P terms. These terms, often referred to as "remoteness" before Anderson and van Wincoop (2003) labelled them "multilateral resistance", depend on the Y and E of *all other countries*, discounted by trade costs. The omission of these terms can lead to inconsistency in the estimates of the included right-hand side variables because it is likely that the multilateral resistance terms are correlated with GDP and trade cost measures.

The final term in equation 3.4.2 encompasses two distinct factors. The first is the time-invariant trade costs between the origin and destination. The other is the time-varying costs. In this paper, part of the *time-varying* cost is related to the linkage effect. The other time-varying cost will be absorbed by the error term. In the standard gravity equation, distance is always used as a proxy for trade costs. Most studies add other proxies such as dummies for common languages and a prior colonial relationship. Since the list of potential proxies is a long one and few are well-measured, a more appealing

approach is to allow for dyadic fixed effect. We embed our linkage hypothesis within the specification in terms of origin-year, destination-year, and dyadic fixed effects, as advocated by Baldwin and Taglioni (2006).

$$\ln X_{odt} = \alpha_{ot} + \beta_{dt} + \delta_{od} + \lambda L_{odt} + \varepsilon_{odt} \tag{3.4.3}$$

o : origin $\qquad$ d : destination $\qquad$ t : year

X_{odt} is the total exports value from origin o to destination d in year t. α_{ot} and β_{dt} represent origin-year and destination-year fixed effects, respectively. δ_{od} controls for dyadic fixed effects. L_{odt} is the linkage variable formed by multinational retailers. Its construction method will be explained later. The technical problems associated with this specification will be discussed in Appendix 2.2. Our unit of analysis is at the origin–destination–year level. The coefficient we are interested in is λ, i.e. the effect of multinational retailers on the total export value of retail goods. We expect λ to be significantly positive if multinational retailers increase Chinese retail goods exports via their linkage effects.

Two linkage variables are constructed to proxy for the linkages generated by the four multinational retailers. One linkage variable is primarily based on locations of the four retailers' global procurement centers in China; the other is based on retailers' store counts. As mentioned in section 3.2, we speculate that retailers' procuring and retailing systems may both have positive effects on Chinese regional exports. Leaving either stores count or the location of global procurement centers out of the specification might introduce omitted variable bias and makes the estimate inconsistent. However, we believe that the positive effects of global procurement centers are intended and stores' pro–trade effects are accidental, and hence the coefficients of global procurement centers for retail goods are expected to be larger in magnitude and more statistically significant than the coefficients of stores.

The linkage variable based on procurement centers is GP_{odt}, and it is constructed as

$$GP_{odt} = \sum_{r=1}^{4} \frac{n_{drt}}{D_{ort}}.$$

r denotes the retailer. n_{drt} refers to the number of outlets that retailer r has in country d in year t. D_{ort} is the geographic distance from the city to the nearest procurement center of retailer r in that year. The city's proximity to all other global procurement centers will be considered in a robustness test. GP_{odt} increases when the store count of any retailer in the destination country increases, or the distance from the city to the nearest global procurement center of any one of the four retailers decreases. We expect this linkage related to the procurement centers to have a significantly positive effect on exports.

The other linkage variable is constructed according to the number of stores that each retailer has in a Chinese city and a destination country. N_{ort} and n_{drt} respectively represents the number of outlets of retailer r in origin city o and destination country d in year t. ST_{odt} is the sum of products of the number of stores for each city–country dyad:

$$ST_{odt} = \sum_{r=1}^{4} n_{ort} n_{drt}.$$

In the empirical test, we apply lagged period measures of key variables to test linkage effects of retailers. There are two reasons to do so. One is driven by business practices; the other enhances the convenience of explaining the coefficients. In terms of business practices, it takes time for procurement centers and stores to show their full strength in the customs records after their establishment, and lagged measures are more appropriate for showing the effect of multinational retailers on exports. Take Wal-Mart's procurement centers as an example. Before signing contracts, Wal-Mart's procurement centers need to accumulate manufacturers' information, send the product list to the buyers in overseas stores, get their feedback, and arrange business trips. A spokeswoman with Carrefour

China revealed that "it usually takes us half a year, even a year, to clinch a deal."[48] After the contracts have been signed, manufacturers also need time to manufacture the products. There is a time gap between the date the global procurement center or store is set up and the time when the first shipment of goods arrives at customs. The lagged measures of key variables allow for such a time gap and are more effective for examining the effects. In addition, lagged measures of key variables allows us to circumvent the simultaneity and common cause problems to some extent, and simplify the explanation of the results. The set-up of procurement centers or stores is pre-determined with regard to the current exports level. The coefficient estimated thus cannot be explained as driven by contemporaneous reverse causation. This measure meets the challenge of common cause problems as well. The lagged measure of multinational retailers' linkage variable is pre-determined with respect to city–year dummies as well. The coefficients of the linkage variables thus cannot be explained as the current unobserved city heterogeneities attract entries of multinational retailers, and simultaneously increase the city's exports.

We construct lagged period measures in the following ways. The lagged i period GP link is calculated as

$$GP_{od(t-i)} = \sum_{r=1}^{4} \frac{n_{drt}}{D_{or(t-i)}}.$$

Only the denominator takes lagged value. The motive for such a construction is the trade-off between limitations of data and sample size. The worldwide store distribution of the four multinational retailers provides data only back to 1996. As we take the lagged measure of key variables, our sample size shrinks. One the one hand, it reduces the efficiency of the estimates. On the other hand, it makes the estimates not comparable over time. In the meantime, it seems to be common sense that stores need to purchase products before they start operation, so the number of stores set to open could also affect the link formed by multinational retailers. Following the same logic, the lagged i period

[48] www.chinadaily.com.cn, reported on October 14, 2006.

of ST_{odt} link is constructed as

$$ST_{od(t-i)} = \sum_{r=1}^{4} n_{drt} n_{or(t-i)}.$$

It is worth mentioning that when we take lagged value of key variables in a symmetric way (i.e. both the origin and destination parts are taken lagged value), the estimation results are quite similar.

3.4.2 Method for Testing Capability Effects

In the second half of this paper, we investigate the alternative mechanism through which multinational retailers increase China's exports. The specific mechanism tested is whether multinational retailers increase China's exports by increasing Chinese cities' *general* export capabilities.

We take the following procedures to test this effect. We first run specification 3.4.3. We extract the estimated coefficients for each city–year dummy ($\hat{\alpha}_{ot}$), and then take them as a measure of cities' time-varying export capabilities. Each city–year coefficient thus becomes a unit of observation in the second stage estimation. Following the suggestion of Saxonhouse (1976), estimated standard errors of the coefficients obtained in the first step are extracted as well, and their inverses are taken as the weights in the second stage estimation. The observations with higher variance are thus given lower weights. The second stage estimation tests the effects of multinational retailers on cities' general export capabilities, controlling for city dummies, year dummies, and other time-varying city–level variables, such as gross value of industrial output per capita and population.

The general export capability could be influenced by a city's access to global procurement centers and the number of stores in the city. A city's access to global procurement

centers (i.e. cityGP_{ot}) is defined as:

$$\text{cityGP}_{ot} = \sum_{r=1}^{4} \frac{1}{D_{otr}},$$

where D_{otr} refers to the distance from the city to the nearest procurement center of retailer r. Four retailer–specific measures are then summed to generate cityGP_{ot}. The number of multinational retailers' stores available in a city (i.e. cityST_{ot}) is the sum of stores over the four retailers in the city. It is defined as

$$\text{cityST}_{ot} = \sum_{r=1}^{4} n_{otr},$$

in which n_{otr} is the number of stores that retailer r has in city o in year t.

In order to test the capability effects of multinational retailers, we apply the following specification:

$$\begin{aligned}\hat{\alpha}_{ot} =& \theta + \gamma_{gp} \ln \text{cityGP}_{ot} + \gamma_{st} \ln \text{cityST}_{ot} + \lambda_{pop} \ln \text{pop}_{ot} \\ &+ \lambda_{gviopa} \ln \text{gviopa}_{ot} + \phi_o + \phi_t + \varepsilon_{ot} \end{aligned} \tag{3.4.4}$$

The dependent variable $\hat{\alpha}_{ot}$ is the estimated coefficients of city–year dummies from the first stage estimation. Pop_{ot} and gviopa_{ot} refer to the population and gross value of industrial output per capita of cities, respectively. ϕ_o is the origin fixed effect, and ϕ_t is the year fixed effect. ε_{ot} is the idiosyncratic error term.

3.5 Data Sources and Features

In order to generate the linkage variable created by the four multinational retailers, three data sets are combined: the four retailers' store distribution in China, stores' distribution over the world, and the distribution of global procurement centers in China.

The worldwide store distributions of the four multinational retailers are summarized from these retailers' annual financial reports. Each retailer usually has multiple formats of stores in operation. In addition to the main format, stores that carry the widest range of goods, the four retailers also have specialist stores, such as electronics stores, apparel stores, etc. This paper focuses on the stores of the main format. The goods sold in the specialist stores differ from those sold in the typical stores. Including specialist stores may introduce measurement errors. The second concern is that the sales areas in the main format stores are usually much larger than those of the other formats. This is especially true for a large number of convenience stores via which those retailers reach the consumers in catchment areas. The main formats for the retailers are as follows: Carrefour—hypermarket; Metro—cash and carries; and Wal-Mart—super-center and Sam's club. Tesco does not have a clear classification of the formats of stores. In their financial reports, only the total number of stores is consistently reported.

Four retailers' store distributions in China are collected from a Chinese web site—`http://www.linkshop.com.cn/index.htm`. It lists the opening date and location of each store of these four retailers. The number of stores of a specific retailer in a Chinese city in a year is then calculated by adding up all stores that have been opened until that time. This number neglects the fact that some stores may have been closed. Since the number of stores in a given year may be miscalculated, the estimation results are subject to attenuation bias.[49] Carrefour, Metro, and Wal-Mart usually operate only their main format of stores in China.[50] We also collected the information on the four retailers' global procurement centers in China. This data set is based on a comprehensive summary of reports in the media, whether in Chinese or English. The sources of this data will be

[49]This data contains measurement errors because some stores must have been closed; however, we do not know which specific stores have been closed. The difference between the store numbers in China reported in retailers' financial reports and the sum of all opened stores increases over time. The number of stores recorded in retailers' financial reports does not provide enough information given our identification strategy. We need to know the specific city in which the store is set up.

[50]There is an exception. Wal-Mart has opened a few neighborhood stores in Guangdong province in very recent years.

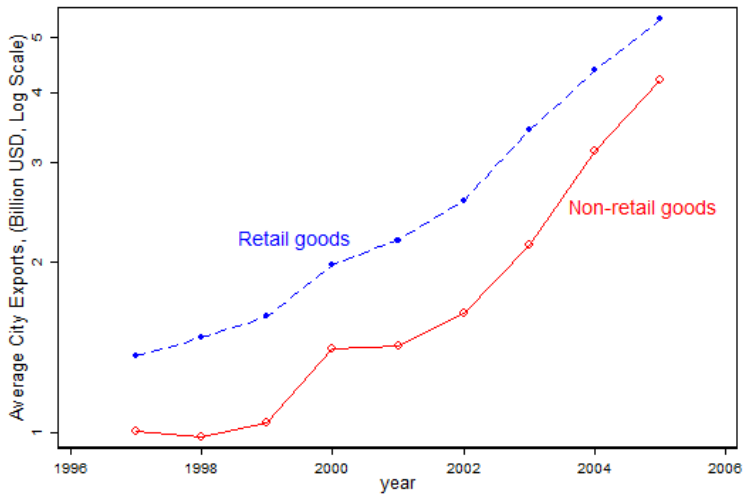


Figure 3.1: Average City–level Exports

detailed in the data appendix of the final version. We classify the global procurement centers established July or later as occurring in the subsequent year.

The exports data are based on Chinese exports of products at HS4 level of disaggregation, reported in the customs General Adminstration of the People's Republic of China over the period 1997–2005. This data set was used under the license to the Center for International Data at the University of California, Davis. It includes information on the city-district of origin and country destination of these exports. The data sources for all control variables used in the gravity models are listed in the Appendix 2.1.

Figure 3.1 displays the average city-level exports. It shows that over the period of 1995–2006, average city exports of both retail and non-retail goods keep increasing. In the meantime, over the last decade, multinational retailers increasingly expanded their operation in mainland China. The two panels in figure 3.2 illustrates the operation of four retailers in 1996 and 2005, respectively. In 1996, multinational retailers are concentrated in the three largest cities in China—Beijing, Shanghai, and Shenzhen. In 2005, multinational retailers have opened their stores in smaller cities, and their procurement centers are present in more inner land cities. These expansions greatly reduce the distances

from each major Chinese cities to global procurement centers of the four multinational retailers.

In figure 3.3, we show how Chinese cities' proximity to global procurement centers, the number of retailer stores, and the average value of city exports of retail goods have grown over time. Each series is expressed as an index relative to its 1997 value (set equal to 100). It shows that over the nine years, Chinese cities' exports are positively associated with the presence of multinational retailers in China.

Before turning to the empirical results, we also check the correlation between our two key variables. Our two linkage variables share a common part, which is the number of stores outside China. The simple correlation between GP_{odt} and ST_{odt} over all years is 0.563. The correlations between cityGP_{ot} (i.e. $\sum_{r=1}^{4} \frac{1}{D_{otr}}$) and cityST_{ot} (i.e. $\sum_{r=1}^{4} n_{otr}$) is -0.382. Their specific correlations in 2000 and 2005 respectively are scatter plotted in figure 3.4. The left panel is based on the city-level data, and the panel on the right-hand side is for the province-level data. Both of them show that these two key variables are not highly correlated.

3.6 Linkage Effects

In this section, we investigate whether multinational retailers increase Chinese retail goods exports by creating global linkages connecting distant markets. We test this effect on Chinese city-level exports and find that multinational retailers do *not* increase cities' exports to the countries where the retailers have outlets *in particular*. In order to check the robustness of this finding, we also test linkage effects of multinational retailers in province-level data and conduct the city-level analysis again using different measures of linkage variables.

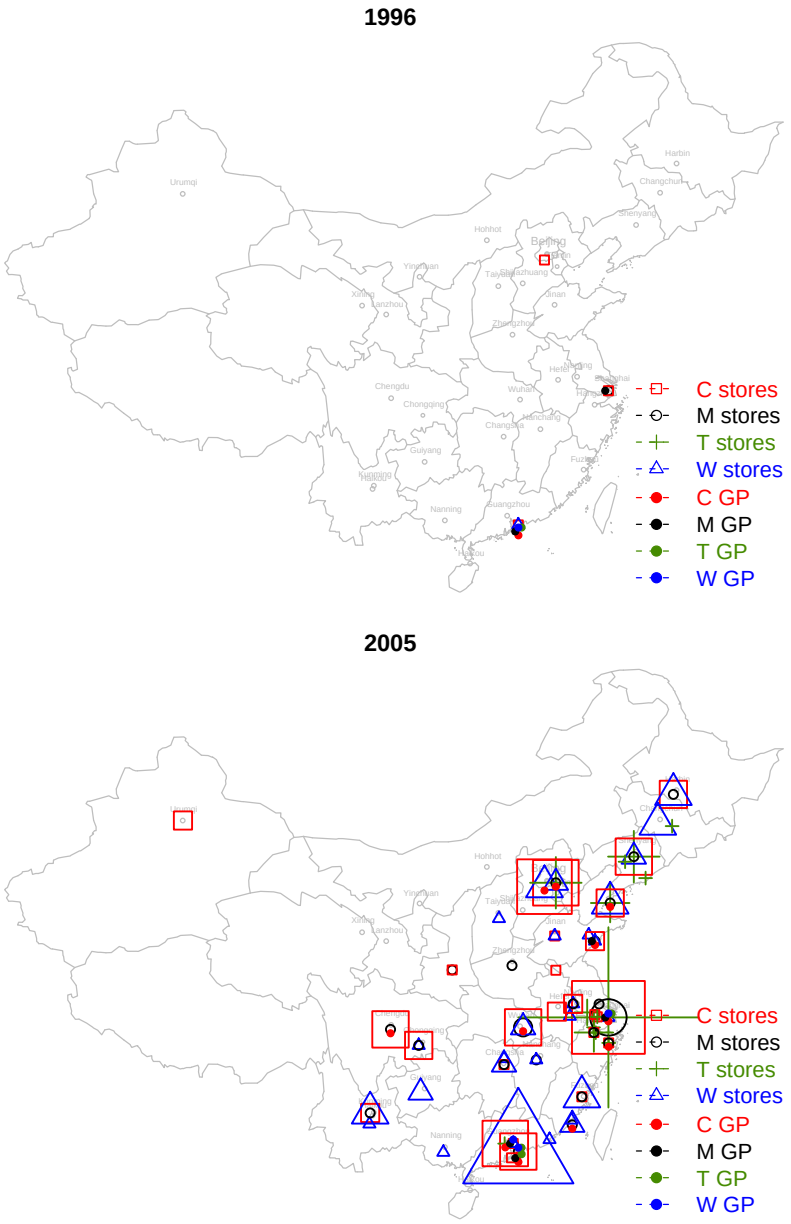


Figure 3.2: Operation of The Four Multinational Retailers in China

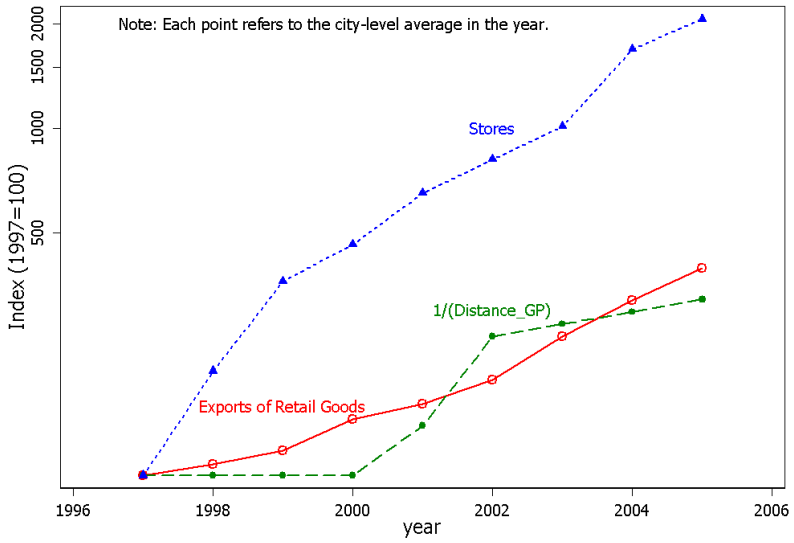


Figure 3.3: City Trend

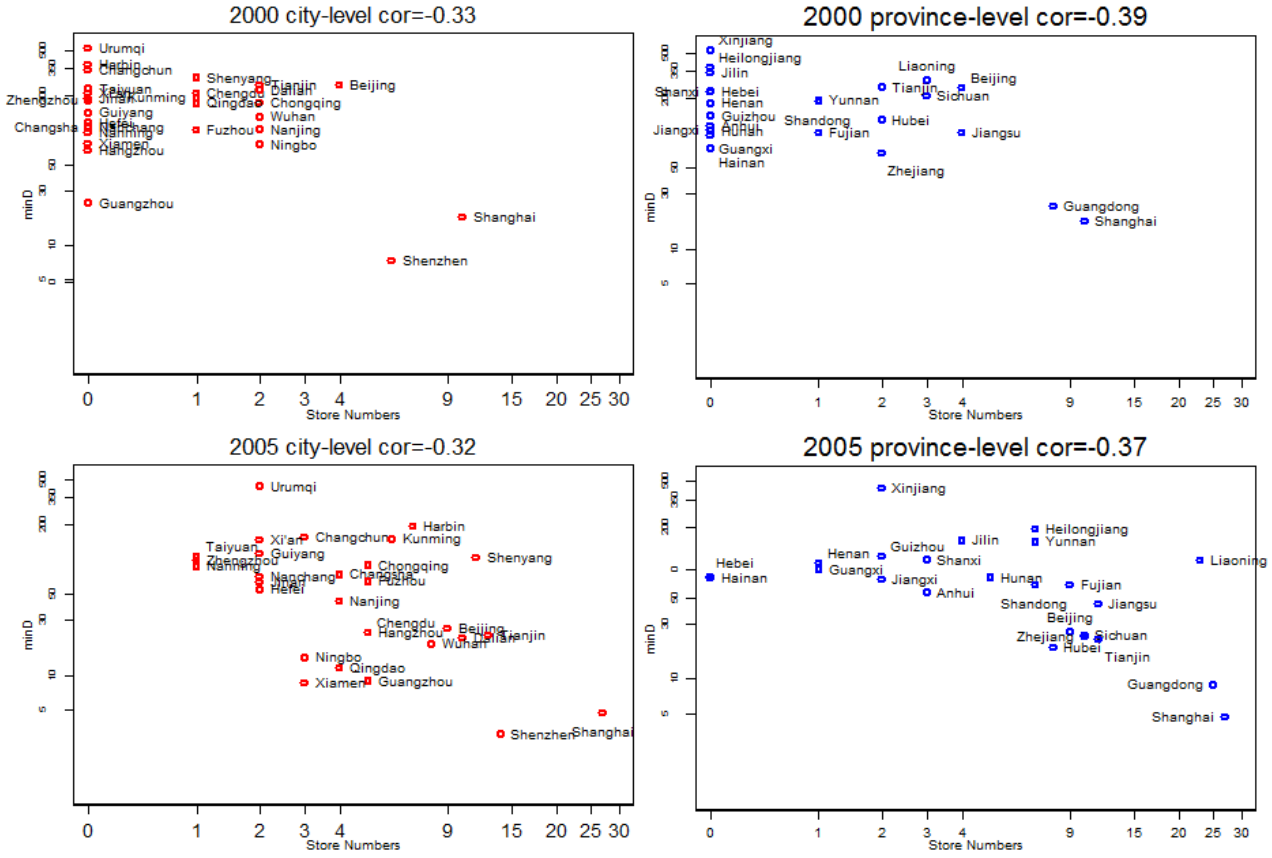


Figure 3.4: Correlations between Procurement Centers and Stores in China

3.6.1 City-level Results

Table 3.2 reports the results when the key independent variables are measured at the lagged 1 period. We start with the basic gravity model and then gradually relax the assumptions on the unobserved heterogeneities of origins, destinations, and dyads. Our preferred specification 3.4.3 is shown in the last column of Table 3.2.

Table 3.2 starts with the simple gravity model. All variables of standard gravity models enter with the expected signs and magnitudes. In this regression, two variables differ from a traditional gravity model—intD_{ot} and gviopa_{ot}, and they require some explanation. The Chinese "open door policy" mainly takes effect in coastal cities, so the distance to the nearest port is an important factor for a city's exports, and it should be included in the specification. Instead of using GDP per capita, we use the gross value of industrial output per capita (i.e. gviopa_{ot}) because of data availability. The gross value of industrial output, by definition, includes the value of intermediate goods. Since almost half of Chinese exports are processing trade,[51] Gvio_{ot} captures the effect of the origin city's economic mass on exports more accurately. Column 1 in Table 3.2 shows that GP_{odt} is positively associated with a city's exports of retail goods. The coefficient of GP_{odt} equals 0.250. This shows that exports of retail goods increase by 2.5% as GP_{odt} increases by 10%. In this specification, stores do not appear to have a significantly positive effect on the exports of retail goods.

In column 2 of Table 3.2, we add city and country fixed effects. These two sets of fixed effects control for the time-invariant heterogeneities in the city's and country's unobserved features. They also incorporate the permanent component of multilateral resistance terms. Column 2 of Table 3.2 shows that the coefficient of GP_{odt} falls very slightly to 0.229 and remains significant at the 1% level. Disdier and Head (2008) report the mean effect of distance on trade is -0.9. In this regression, the coefficient of extD_{od} equals

[51] Andreas Freytag states that "[...]the share of processing trade in China's export appears to have grown over the last decades from 47 percent in 1992 to 55 percent in 2005."(http://www.voxeu.org/index.php?q=node/1150, accessed in September 2008.)

Table 3.2: Linkage Effects Test, City-level

	(1)	(2)	(3)	(4)	(5)
Fixed effects controlled	t	t country city	t dyadic	dyadic country–t	dyadic country–t city–t
$L.1\ln(\mathrm{GP}_{odt}+1)$	0.250^a	0.229^a	0.139^a	0.400^a	-0.0705
	(0.0615)	(0.0489)	(0.0392)	(0.0526)	(0.0643)
$L.1\ln(\mathrm{ST}_{odt}+1)$	0.00115	0.0260^b	0.0277^a	0.0495^a	0.0121
	(0.0156)	(0.0113)	(0.00978)	(0.0100)	(0.0122)
$\ln(\mathrm{gviopa}_{ot})$	1.380^a	0.225^a	0.224^a	0.181^a	
	(0.0442)	(0.0408)	(0.0394)	(0.0391)	
$\ln(\mathrm{pop}_{ot})$	0.901^a	-0.384^a	-0.398^a	-0.321^a	
	(0.0425)	(0.125)	(0.125)	(0.117)	
$\ln(\mathrm{gdppa}_{dt})$	0.936^a	1.026^a	1.064^a		
	(0.0252)	(0.0881)	(0.0848)		
$\ln(\mathrm{pop}_{dt})$	0.918^a	2.479^a	2.977^a		
	(0.0293)	(0.734)	(0.731)		
$\ln(\mathrm{extD}_{od})$	-0.802^a	-0.916^a			
	(0.0463)	(0.119)			
$\ln(\mathrm{intD}_{od})$	-0.386^a				
	(0.0254)				
Constant	-31.00^a	1.158	-9.541^a	10.29^a	14.60^a
	(1.478)	(3.408)	(2.841)	(1.272)	(0.121)
N	14022	14022	14022	14141	14141
R^2	0.661	0.820	0.156	0.244	0.305
groups			1714	1716	1716
S.E of u_i			3.537	2.941	3.592
S.E.of e_{it}			0.915	0.880	0.854
RMSE	1.609	1.174	0.857	0.823	0.798

Note: Note: Robust standard errors in parentheses are clustered at the dyadic level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels.

-0.916, which is consistent with the previous study. In column 3 of Table 3.2, we apply dyadic fixed effects. This set of fixed effects controls for not only the features captured in column 2 but also the unobserved permanent dyadic features, such as geographic distance, common border, etc. Our procurement centers' link variable, GP_{odt}, must be correlated with these dyadic features. Not controlling for these features, they enter the error term and make the estimation inconsistent. After we control for the dyadic fixed effects, we identify the effects of GP_{odt} based only on GP_{odt}'s variation within the dyad over time. We find that a 10% increase in GP_{odt} leads to a 1.39% increase in the exports of retail goods. The first three columns of Table 3.2 show that global procurement centers are positively associated with bilateral trade above the level predicted by both the standard gravity and dyadic fixed effects models. Given the GP_{odt} is pre-determined, these results show that multinational retailers have a positive effect on cities' exports of retail goods.

In columns 4 and 5 of Table 3.2, we include the city–year and country–year fixed effects. Two concerns prompt us to set up these specifications. First, it helps to fully capture the multilateral resistance terms. The time-varying components of the multilateral resistance terms are taken into account. Second, this specification allows us to relax the assumption that cities have the same unobserved features *over time*. Column 4 gives the estimation result when only the country–year fixed effects are included. The coefficient of GP_{odt} is significant, and its magnitude is even much larger. The estimation result of our preferred specification 3.4.3 is shown in column 5, which controls for both city–year and country–year fixed effects. Once we add the city–year fixed effects, the coefficient of GP_{odt} is largely reduced in magnitude and becomes statistically insignificant. ST_{odt} exhibits the same pattern. It is significantly positive in columns 2 to 4 in Table 3.2, and it becomes insignificant in column 5. In summary, the results in Table 3.2 demonstrate that linkage effects are not the working mechanism through which multinational retailers increase China's exports.

The large decrease in the magnitudes of the coefficients of the two key variables and

the striking change of signs over the columns in Table 3.2 give a clear indication that city–level time-varying unobserved features create the significantly positive associations shown in all columns other than the last one. These city–level unobserved features could be explained as cities' recent development in their export capabilities related to retail goods. This unobserved feature increases the city's exports to *all* countries rather than *only* to the countries where the multinational retailers have stores.

3.6.2 Robustness Tests

In this section, we conduct three robustness checks of our results. First, we re-estimate the regressions of the previous section on province–level trade data. Second, we construct a new linkage variable formed by global procurement centers and re-run the regressions. Third, we report the estimation results lagging key variables by two and three periods.

The motive for analyzing province–level data is that province-level export data include the data of *all* Chinese provinces. These data are more comprehensive relative to the city–level data, which include only 35 major Chinese cities. We find that the provincial results are qualitatively similar to the city–level results. The presence of multinational retailers is positively associated with bilateral trade above the levels predicted by standard gravity and dyadic fixed effects models; however, these significant coefficients of GP_{odt} and ST_{odt} disappear once province–year fixed effects are controlled for. Table 3.3 gives the estimated results.

In column 1 of Table 3.3, a 10% increase of GP_{odt} is only associated with a 1.52% increase in exports, much smaller than the coefficient at the city level, i.e. 2.5%. GP_{odt} is also less significant at the province level. These differences hold true in the following four columns of Table 3.3 as well.

We also found that ST_{odt} performs better than GP_{odt} at the province level, which could be driven by the way we construct GP_{odt} for each province. Because of data limitations, only the location of a province's capital is used in calculating the *province*'s distance to

Table 3.3: Linkage Effects Robustness Test, Province-level

	(1)	(2)	(3)	(4)	(5)
Fixed effects controlled	t	t nation province	t dyadic	dyadic nation–t	dyadic nation–t province–t
$L.1$ ln(GP$_{odt}$+1)	0.152^b	0.111^c	0.0369	0.231^a	-0.0181
	(0.0744)	(0.0587)	(0.0530)	(0.0619)	(0.0770)
$L.1$ ln(ST$_{odt}$+1)	-0.0157	0.0267^b	0.0351^a	0.0469^a	-0.00989
	(0.0144)	(0.0108)	(0.00939)	(0.00897)	(0.0113)
ln(gdppa$_{ot}$)	2.565^a	0.581^a	0.519^a	0.482^a	
	(0.0675)	(0.194)	(0.191)	(0.173)	
ln(pop$_{ot}$)	1.138^a	-0.240	-0.0787	-0.411	
	(0.0359)	(0.460)	(0.451)	(0.411)	
ln(gdppa$_{dt}$)	0.941^a	1.102^a	1.130^a		
	(0.0239)	(0.0841)	(0.0832)		
ln(pop$_{dt}$)	0.935^a	2.285^a	2.711^a		
	(0.0294)	(0.685)	(0.667)		
ln(extD$_{od}$)	-0.881^a	-0.958^a			
	(0.0532)	(0.164)			
ln(intD$_{od}$)	-0.401^a				
	(0.0250)				
Constant	-63.49^a	-8.667	-17.70^a	3.772	14.11^a
	(2.069)	(6.741)	(6.451)	(5.503)	(0.113)
N	13445	13445	13445	13554	13554
R^2	0.737	0.851	0.271	0.357	0.427
groups			1470	1472	1472
S.E of u_i			3.289	3.114	3.180
S.E.of e_{it}			0.882	0.847	0.809
RMSE	1.470	1.110	0.833	0.798	0.761

Note: Robust standard errors in parentheses are clustered at the dyadic level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels.

its nearest procurement center. This method neglects all the other areas in the province which might be either closer or further away from procurement centers. GP_{odt} at the province level thus measures a province's proximity to retailers' procurement centers less precisely. However, stores are relatively dispersed over a large geographic area in a province. The presence of multinational retailers in a province is better represented by store numbers.

In the previous analysis, we always use the GP_{odt} as our measure of global procurement centers. It is calculated using the formula $GP_{odt} = \sum_{r=1}^{4} \frac{n_{drt}}{D_{ort}}$, where D_{ort} is the geographic distance from city o to the nearest procurement center of retailer r in year t. In the construction of this variable, we neglect all the other global procurement centers that a city could reach at that time even though those procurement centers are further away. In order to ensure our insignificant linkage effect is not driven by this specific measure, we construct another measure. The new measure is defined as $SGP_{odt} = \sum_{r=1}^{4} n_{drt} \sum_{c=1}^{M_{crt}} \frac{1}{D_{ortc}}$, where c refers to a specific procurement center of retailer r in year t. M_{crt} refers to the total number of global procurement centers that retailer r has in China in year t. The measure increases whenever a new procurement center is set up, whether this new center is close to or far away from the city. We run the same regressions as before, but we replace GP_{odt} with SGP_{odt}. The new measure does not affect the results qualitatively. Once we control for city–year fixed effects, the significantly positive coefficient associated with SGP_{odt} disappears. The estimation results are shown in Table 3.4.

To this point, we have focused on the lagged 1 period measures of the key variables. However, it might take procurement centers or stores longer to affect the data in the customs figures. In order to preclude this concern, we run the regressions using the lagged 2 and 3 period measures of key variables. When we construct the lagged key variables, we follow the rule mentioned in the data section. The contemporaneous measure of worldwide store counts is used so that the sample does not reduce as earlier measures are utilized. This feature guarantees that the change of the estimates is not driven by

Table 3.4: Linkage Effects Robustness Test, An Alternative Measure

	(1)	(2)	(3)	(4)	(5)
Fixed effects controlled	t	t country city	t dyadic	dyadic country–t	dyadic country–t city–t
$L.1$ ln(SGP$_{odt}$+1)	0.150^a	0.119^a	0.00602	0.472^a	-0.0770
	(0.0559)	(0.0438)	(0.0368)	(0.0661)	(0.0824)
$L.1$ ln(ST$_{odt}$+1)	0.00513	0.0323^a	0.0348^a	0.0489^a	0.0122
	(0.0159)	(0.0116)	(0.01000)	(0.01000)	(0.0122)
ln(gviopa$_{ot}$)	1.392^a	0.239^a	0.240^a	0.175^a	
	(0.0437)	(0.0406)	(0.0392)	(0.0391)	
ln(pop$_{ot}$)	0.899^a	-0.376^a	-0.386^a	-0.324^a	
	(0.0426)	(0.124)	(0.125)	(0.117)	
ln(gdppa$_{dt}$)	0.943^a	1.031^a	1.055^a		
	(0.0253)	(0.0885)	(0.0849)		
ln(pop$_{dt}$)	0.925^a	2.473^a	2.951^a		
	(0.0297)	(0.734)	(0.730)		
ln(extD$_{od}$)	-0.797^a	-0.910^a			
	(0.0464)	(0.119)			
ln(intD$_{od}$)	-0.389^a				
	(0.0254)				
Constant	-31.46^a	0.628	-9.862^a	10.44^a	14.24^a
	(1.460)	(3.414)	(2.846)	(1.270)	(0.0970)
N	14022	14022	14022	14141	14141
R^2	0.660	0.820	0.155	0.245	0.305
groups			1714	1716	1716
S.E of u_i			3.496	2.920	3.251
S.E.of e_{it}			0.915	0.880	0.854
RMSE	1.610	1.175	0.857	0.823	0.798

Note: Robust standard errors in parentheses are clustered at the dyadic level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels.

the change of sample, but rather comes from the timing of the effect. The key variable's estimates become larger in magnitude than the corresponding coefficients in Table 3.2, before we control for city–year fixed effects. However, as before, the inclusion of city–year fixed effects results in the disappearance of significantly positive coefficients. Table 3.5 and 3.6 show the results.

Table 3.5: Linkage Effects Robustness Test, Lagged 2 Period

	(1)	(2)	(3)	(4)	(5)
Fixed effects controlled	t	t country city	t dyadic	dyadic country–t	dyadic country–t city–t
$L.2$ ln(GP$_{odt}$+1)	0.267^a	0.237^a	0.152^a	0.452^a	-0.0477
	(0.0652)	(0.0515)	(0.0417)	(0.0548)	(0.0699)
$L.2$ ln(ST$_{odt}$+1)	0.0020	0.0383^a	0.0462^a	0.0735^a	0.0194
	(0.0164)	(0.0122)	(0.0102)	(0.0104)	(0.0125)
N	14022	14022	14022	14141	14141
R^2	0.661	0.821	0.157	0.247	0.305
groups			1714	1716	1716
S.E of u_i			3.604	2.971	3.324
S.E.of e_{it}			0.914	0.879	0.854
RMSE	1.609	1.174	0.856	0.822	0.798

Note: Note: Robust standard errors in parentheses are clustered at the dyadic level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels.

In conclusion, this section shows that the four largest multinational retailers do not create a significant linkage effect by connecting distant markets. They do not increase a city's exports particularly to the countries in which the multinational retailers have outlets, even though there is good evidence showing that the presence of multinational retailers is positively associated with the exports of retail goods.

3.7 Capability Effects

In the previous section, we demonstrate that linkage effects do *not* appear to be the influential working mechanism. All sets of regressions discussed show that once we control for the regional time-varying unobserved features, the significantly positive coefficients of

Table 3.6: Linkage Effects Robustness Test, Lagged 3 Period

	(1)	(2)	(3)	(4)	(5)
Fixed effects controlled	t	t country city	t dyadic	dyadic country–t	dyadic country–t city–t
$L.3$ ln(GP$_{odt}$+1)	0.272^a	0.251^a	0.153^a	0.490^a	-0.0153
	(0.0716)	(0.0587)	(0.0484)	(0.0658)	(0.0786)
$L.3$ ln(ST$_{odt}$+1)	-0.0109	0.0402^a	0.0480^a	0.0748^a	0.00453
	(0.0173)	(0.0129)	(0.0104)	(0.0110)	(0.0129)
N	14022	14022	14022	14141	14141
R^2	0.660	0.820	0.157	0.246	0.305
groups			1714	1716	1716
S.E of u_i			3.574	3.005	3.579
S.E.of e_{it}			0.914	0.879	0.854
RMSE	1.610	1.174	0.856	0.823	0.798

Note: Robust standard errors in parentheses are clustered at the dyadic level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels.

GP$_{odt}$ and ST$_{odt}$ disappear. This *recurring* evolution of the coefficients gives us a strong hint, which is that the regional time-varying unobserved features play a key role in the positive associations shown above.

3.7.1 City-level Results

In this subsection, we test whether the presence of multinational retailers in China *increases* Chinese cities' time-varying unobserved features. In other words, we test multinational retailers' capability effects—whether the presence of multinational retailers in a Chinese city improves that city's *general* export capabilities.

As discussed in section 3.4.3, specification 3.4.4 is applied to test whether being close to multinational retailers' global procurement centers or having a large number of stores in the city, increases the city's general export capabilities. The general export capability is estimated with the city–year dummies from specification 3.4.3. The closeness of a city is measured by cityGP$_{ot}$, and it is defined as $\sum_{r=1}^{4} \frac{1}{D_{otr}}$, in which D_{otr} refers to the distance from the city to the nearest procurement center of retailer r in year t. The concentration

of stores in a city is measured by cityST_{ot}. It is defined as $\sum_{r=1}^{4} n_{otr}$, where n_{otr} denotes the number of stores.[52]

In testing the capability effects, GDP per capita, population, city fixed effects, and year effects are controlled for in order to alleviate the endogeneity concern associated with location choice decisions. When retailers choose cities to set up stores, they usually look for places with high disposable income and large population of middle class. Meanwhile, GDP per capita and population also have strong impacts on cities' general export capabilities. In the absence of these two variables, even if we find a significantly positive coefficient for cityST_{ot}, the estimate cannot be taken as evidence supporting the statement that retailers' stores have a positive effect on cities' general export capabilities. As to global procurement centers, when multinational retailers search for locations, they look for places with strong export capabilities. City fixed effects capture all city level time-invariant unobserved heterogeneities that may promote trade. They may involve local transportation and logistic systems, government preferential policies towards foreign direct investments and exports, consumption habits of local consumers, etc. The policy changes affecting all Chinese cities are absorbed by year effects.

Because the dependent variable in testing the capability effects takes the estimated parameters of city–year dummies from the previous linkage effects analysis, the data set used here has a unique structure. In specification 3.4.3, the total number of city–year dummies that can be estimated is $N_o \cdot (t-1)$ rather than $N_o \cdot t$. This difference is resulted from a perfect collinearity problem between city–year and dyadic fixed effects. Appendix 2.2 demonstrates this issue with a simple example.[53] For each city, after demeaning city–year dummies on each dyad, the sum of those demeaned city–year dummies *for each city over all years* are zero vectors. This explains why STATA normally drops *some* city–year

[52]In robustness tests, we run the same regressions on the data at the province level. Under that circumstance, these two variables are named as provGP_{ot} and provST_{ot} and constructed correspondingly.

[53]There are another two sets of perfect collinearity problems embedded in specification 3.4.3 suggested by Baldwin and Taglioni (2006).

dummies automatically when we follow the traditional method and set only the first city–year dummy as default. In order to replicate the results obtained by allowing STATA to decide which dummies to drop, and fully control the city–year dummies estimated, we must drop one city–*year* dummy for each *city*. In the following regressions, we set the earliest city–year dummy of each city as default. Therefore, in total, 280 (35*(9-1)=280) city–year dummies are estimated in specification 3.4.3. In the following analysis, zeros are plugged in as the estimated parameters of the 35 *first year* of city–year dummies since they are taken as the default default group.[54]

Table 3.7 contains the estimation results of capability effects at the city level. The first three columns show the impacts of lagged one, two, and three period measures respectively on the general export capabilities. Column 4 shows the estimates after controlling for the contemporaneous measures of cityGP_{ot} and cityST_{ot}. In the last column, future cityGP_{ot} and cityST_{ot} are putted in to test whether the significantly positive coefficients shown in the first four columns are driven by retailers locating their stores or procurement centers in the cities with high export *potentials*. Over the five columns, we focus on the coefficients of lagged cityGP_{ot} and cityST_{ot} because lagged measures are pre-determined with regard to general export capabilities, which simplifies the explanation of coefficients. It is hard to find a time-varying factor which attracts entries of multinational retailers and simultaneously increases general export capabilities. It is also hard to imagine that current export capabilities could attract retailers' entries two or three years ago unless the entry decision is based on the predictions of each city's export potentials.

The estimation results in the first three columns of Table 3.7 show that proximity to global procurement centers of multinational retailers has a significantly positive effect on cities' export capabilities, and this effect becomes stronger as time goes by. The same arguments also apply to stores. In column 1, cityGP_{ot} is significantly positive,

[54]In the robustness test, when the estimated standard errors are needed as weights, only the 280 observations estimated are utilized.

Table 3.7: Capability Effect Robustness Tests, City, Unweighted

Lagged t	(1) t=1	(2) t=2	(3) t=3	(4) All	(5) With F.1
L.1 ln(cityGP$_{ot}$)	0.253^a			-0.026	-0.032
	(0.083)			(0.058)	(0.061)
L.1 ln(cityST$_{ot}$+1)	0.102			-0.113^b	-0.154^c
	(0.074)			(0.056)	(0.079)
L.2 ln(cityGP$_{ot}$)		0.268^a		0.085	0.147^b
		(0.091)		(0.064)	(0.062)
L.2 ln(cityST$_{ot}$+1)		0.203^b		0.048	0.119
		(0.081)		(0.071)	(0.072)
L.3 ln(cityGP$_{ot}$)			0.270^a	0.179^a	0.294^b
			(0.081)	(0.060)	(0.130)
L.3 ln(cityST$_{ot}$+1)			0.255^a	0.184^b	0.090
			(0.081)	(0.069)	(0.064)
ln(cityGP$_{ot}$)				0.173^a	0.097
				(0.057)	(0.075)
ln(cityST$_{ot}$+1)				0.118	0.051
				(0.070)	(0.070)
F.1 ln(cityGP$_{ot}$)					0.085
					(0.060)
F.1 ln(cityST$_{ot}$+1)					0.030
					(0.068)
ln(pop$_{ot}$)	-0.292	-0.271	-0.309^c	-0.130	-0.135
	(0.218)	(0.204)	(0.178)	(0.205)	(0.202)
ln(gviopa$_{ot}$)	0.154^b	0.136^c	0.073	0.058	0.057
	(0.072)	(0.074)	(0.099)	(0.080)	(0.073)
Constant	0.593	0.818	2.125	1.943	2.912^b
	(1.224)	(1.175)	(1.495)	(1.430)	(1.406)
N	315	315	315	315	280
$Within R^2$	0.870	0.876	0.880	0.890	0.875
RMSE	0.241	0.236	0.231	0.224	0.207

Note: Robust standard errors in parentheses are clustered at the city level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. City and year fixed effects are both controlled.

which provides direct evidence for the statement that city time-varying effects absorb the positive impacts of multinational retailers on exports. The results in column 3 show that after the global procurement centers have been operating for three years, as cityGP_{ot} increases by 10%, the general export capability increases by 2.7%; as cityST_{ot} increases by 10%, the general export capability increases by 2.55%.

In column 4 of Table 3.7, cityGP_{ot} and cityST_{ot} are included in order to make sure previous results are not driven by reverse causation and time-series correlations. Cities with high general export capabilities could attract multinational retailers to set up global procurement centers and stores there. Meanwhile, the lagged measures of key variables are highly correlated with their contemporaneous measures. Without controlling for contemporaneous variables, the significantly positive results shown in the first three columns could just pick up the strong correlation between the contemporaneous measures and the dependant variable, which can be driven by reverse causation. The results in column 4 show that this argument is not a big concern. The lagged 3 period cityGP_{ot} and cityST_{ot} are significant even though cityGP_{ot} are significant as well. Since the changes in retailer presence occurring prior to changes in city export capability, this result gives some reason to believe the effect could be causal.

Before claiming multinational retailers have a directly causal effect on cities' general export capabilities, it is important to show that the previous results are not driven by retailers locating their stores or procurement centers according to cities' export *potentials*.[55] Large multinational retailers are sophisticated and experienced. They are able to predict which Chinese cities will probably have big jumps in their export capabilities in the next a few years. When they choose cities, their decisions are based on not only a city's present features, such as GDP per capita, but also a city's export potentials. In other words, since the variations of cityGP_{ot} and cityST_{ot} heavily depend on retail-

[55]This forward-looking story should not be a serious concern in our analysis. If it plays an important role, multinational retailers are expected to locate their procurement centers or stores in same cities. However in fact, the stores and procurement centers are widely distributed in China.

ers' locations, the key variables could be positively correlated with error terms in future periods. Under this circumstance, fixed effect estimates are inconsistent.

In the last column of Table 3.7, we implement a "strict exogeneity test", which is discussed in Wooldridge (2002, p. 285). To do it, we add in future measures of cityGP$_{ot}$ and cityST$_{ot}$. If cityGP$_{ot}$ and cityST$_{ot}$ are strictly exogenous to export capabilities, $F.1$ cityGP$_{ot}$ and $F.1$ cityST$_{ot}$ should be insignificant. In column 5 of Table 3.7, neither the coefficient of $F.1$ cityGP$_{ot}$ or $F.1$ cityST$_{ot}$ is statistically significant, and the p–value of the F–test for these two variables is 0.339. These results suggest that the forward-looking story is not supported by the data. The total average treatment effect equals the sum of cityGP$_{ot}$ and cityST$_{ot}$'s coefficients in all periods other than the future measures. Its magnitude is 0.612, and its p–value of the F–test is 0.011.

3.7.2 Robustness Tests

In order to check the robustness of the findings, we conduct the following two tests. First, we take $1/s.e.(\widehat{\alpha_{ot}})^2$ as the weight and re-run the analysis at the city level. We then conduct the same analysis at the province level.

Table 3.8 presents the estimation results when the inverses of the estimated standard errors are taken as weights. Saxonhouse (1976) shows that when estimated parameters are utilized as dependent variables, the estimates from ordinary least square estimations are unbiased; however, they are inefficient because of heteroscedasticity. The standard errors calculated from the usual formula could be underestimated, which implies that variables reported as significant might be not at all significant. In order to deal with this problem, Saxonhouse (1976) suggests that weighting each observation on all variables used in the second stage equation by the inverse of the estimated standard error of the dependent variable. Table 3.8 shows this GLS estimation results and confirms our previous findings. The magnitudes of the coefficients are close to the estimates in Table 3.7, and lagged cityGP$_{ot}$ and cityST$_{ot}$ are still significant even though the significant

Table 3.8: Capability Effect Tests, City, Weighted

Lagged t	(1) t=1	(2) t=2	(3) t=3	(4) All
L.1 ln(cityGP$_{ot}$)	0.214^a			-0.009
	(0.078)			(0.055)
L.1 ln(cityST$_{ot}$+1)	0.070			-0.090
	(0.067)			(0.058)
L.2 ln(cityGP$_{ot}$)		0.221^b		0.073
		(0.088)		(0.054)
L.2 ln(cityST$_{ot}$+1)		0.170^b		0.030
		(0.078)		(0.096)
L.3 ln(cityGP$_{ot}$)			0.240^b	0.126^b
			(0.089)	(0.054)
L.3 ln(cityST$_{ot}$+1)			0.236^a	0.192^b
			(0.079)	(0.086)
ln(cityGP$_{ot}$)				0.135^b
				(0.059)
ln(cityST$_{ot}$+1)				0.095
				(0.073)
ln(pop$_{ot}$)	-0.287	-0.242	-0.302^b	-0.154
	(0.174)	(0.156)	(0.127)	(0.163)
ln(gviopa$_{ot}$)	0.294^a	0.291^a	0.274^b	0.227^b
	(0.086)	(0.088)	(0.102)	(0.098)
Constant	-0.987	-1.326	-0.345	-0.976
	(1.429)	(1.410)	(1.727)	(1.563)
N	280	280	280	280
R^2	0.906	0.910	0.913	0.919
RMSE	0.230	0.226	0.222	0.217

Note: Robust standard errors in parentheses are clustered at the city level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. Each observation is weighted by its inverse of standard error. City and year fixed effects are both controlled.

levels of some coefficients decrease from 1% to 5%. This table does not contain the results of the strict exogeneity test because the test has to be implemented in fixed effect models in which the analytic weight option is not allowed in STATA.

Province–level results are shown in Table 3.9. The dependent variable takes the estimated parameters of province–year dummies in our preferred specification by using the province–level data, i.e. column 5 of Table 3.3. The results corroborate the findings at the city–level. After global procurement centers have been operating for three years, as provGP$_{ot}$ increases by 10%, the general export capability increases by 1.41%. The number of stores in a province consistently has a positive effect on the general export capability of the province, and it becomes stronger as time elapses. After the stores have been operating for three years, a 10% increase in provST$_{ot}$ induces a 1.91% increase in general export capabilities. In column 5, the p–value of a F–test for $F.1$ provGP$_{ot}$ and $F.1$ provST$_{ot}$ is 0.47, and the p–value of the F–test for all the other periods' provGP$_{ot}$ and provST$_{ot}$ is smaller than 0.01. This result again confirms that the presence of multinational retailers have a directly positive effect on general export capabilities. At the province level, store variable performs better than the linkage variable based on global procurement centers in comparison to the results at the city level. This is similar to the results in the discussion of linkage effects. As explained before, it is driven by the fact that we use only the locations of provincial capitals in constructing provGP$_{ot}$. It does not precisely capture the proximity of a province to the nearest global procurement centers.

In summary, this section provides good evidence for multinational retailers' capability effects. It supports the hypothesis that the presence of multinational retailers increase China's exports via improving the general export capabilities of Chinese regions.

Table 3.9: Capability Effect Robustness Tests, Province, Unweighted

	(1)	(2)	(3)	(4)	(5)
Lagged t	t=1	t=2	t=3	All	With F.1
L.1 ln(provGP$_{ot}$)	-0.009			-0.195^c	-0.192^c
	(0.107)			(0.098)	(0.098)
L.1 ln(provST$_{ot}$+1)	0.151^c			-0.059	-0.180^c
	(0.081)			(0.059)	(0.088)
L.2 ln(provGP$_{ot}$)		0.089		0.061	0.121^c
		(0.084)		(0.069)	(0.066)
L.2 ln(provST$_{ot}$+1)		0.163^c		-0.047	0.101
		(0.083)		(0.074)	(0.094)
L.3 ln(provGP$_{ot}$)			0.141	0.201^b	0.277^c
			(0.087)	(0.075)	(0.162)
L.3 ln(provST$_{ot}$+1)			0.191^c	0.277^b	0.177
			(0.095)	(0.121)	(0.146)
ln(provGP$_{ot}$)				-0.001	-0.058
				(0.093)	(0.121)
ln(provST$_{ot}$+1)				0.050	0.111
				(0.070)	(0.078)
F.1 ln(provGP$_{ot}$)					0.078
					(0.103)
F.1 ln(provST$_{ot}$+1)					-0.099
					(0.094)
ln(pop$_{ot}$)	-0.102	-0.549	-0.923	-0.961	-0.829
	(1.310)	(1.271)	(1.389)	(1.295)	(1.449)
ln(gviopa$_{ot}$)	0.722	0.617	0.679	0.802	0.618
	(0.485)	(0.498)	(0.576)	(0.547)	(0.599)
Constant	-18.958	-13.051	-12.378	-15.879	-10.791
	(14.860)	(15.732)	(19.294)	(17.169)	(19.328)
N	300	300	270	270	240
R^2	0.753	0.759	0.769	0.774	0.714
RMSE	0.298	0.294	0.293	0.293	0.294

Note: Robust standard errors in parentheses are clustered at the province level with ^{a}, ^{b}, and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. Each observation is weighted by its inverse of estimated standard error. Province and year fixed effects are both controlled.

3.8 Concluding Remarks

We motivated this paper with the argument Wal-Mart offered to the government of India: "Were it to have outlets in India, its procurement would naturally increase. (The Metro operation in Bangalore buys 98% of its stock locally.) Suppliers would become familiar with its requirements, and exports would also climb."[56] What do our results imply with regard to Wal-Mart's claim? First, estimates from the standard gravity model show that cities near purchasing centers export significantly more to countries with retail stores than do other cities of similar size and distance. Second, this positive effect does not seem to arise directly from procurement by the retailer in the Chinese city to serve its overseas stores. Rather, retailer presence (proximity to purchasing centers and the placement of stores in a city) is correlated with the city's export fixed effect. The evidence that changes in retailer presence occur prior to changes in city export capability gives some reason to believe the effect could be causal. Therefore, a city with Wal-Mart presence is indeed likely to increase its exports, but those exports will not be biased in the direction of countries with large Wal-Mart presences.

Two lessons could be learned from this econometric exercise. First, it is important to have a good econometric identification strategy to test linkage effects. In this exercise, the estimates of naïve gravity and dyadic fixed effects specifications all suggest that linkages have a significant effect on bilateral exports. However, the estimated linkage effects disappear once we control for the general export capabilities of origins. In other words, without taking into account endogeneity of linkages, we would have ended up with a misleading result. Second, by applying the second stage estimation, we find that the presence of multinational retailers improves the general export capabilities of origins, which is in line with Javorcik and Li (2007)'s findings of the retailer-induced productivity improvement for Romanian food suppliers. The next step in this research agenda would be to devise methods to discriminate between different economic mechanisms through

[56] Economist (2006), April 15.

which proximity to procurement centers enhances export capabilities of cities.

Bibliography

Anderson, J. E. and E. van Wincoop (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *The American Economic Review 93*, 170–192.

Baldwin, R. and D. Taglioni (2006). Gravity for Dummies and Dummies for Gravity Equations. Working Paper 12516, National Bureau of Economic Research.

Basker, E. and P. H. Van (2007). Wal-Mart as Catalyst to U.S.–China Trade. Working Papers 0710, Department of Economics, University of Missouri. Available at http://ideas.repec.org/p/umc/wpaper/0710.html.

Basker, E. and P. H. Van (2008). Imports "R" US: Retail Chains as Platforms for Developing-Country Imports. available at: http://ssrn.com/abstract=1123282.

Disdier, A.-C. and K. Head (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. *The Review of Economics and Statistics 90*(1), 37–48.

Economist (2006). Coming to Market—Retailing in India. *The Economist, U.S. Edition*.

Gould, D. M. (1994). Immigration links to the home country: empirical implications for U.S. bilateral trade flows. *Review of Economics and Statistics 76*, 302–16.

Gulati, N. and A. Zaheer (2000). Strategic Networks. *Strategic Managment Journal 21*, 203–215.

Head, K. and J. Ries (1998). Immigration and trade creation: Econometric evidence from Canada. *Canadian Journal of Economics 31*(1), 47–62.

Head, K. and J. Ries (2004). Exporting and FDI as Alternative Strategies. *Oxford Review of Economic Policy 20*(3), 409–423.

Javorcik, B., W. Keller, and J. R. Tybout (2006). Openness and Industrial Responses in a Wal-Mart World: A Case Study of Mexican Soaps, Detergents and Surfactant Producers. Working Paper 12457, National Bureau of Economic Research.

Javorcik, B. S. and Y. Li (2007). Do the Biggest Aisles Serve Brighter Future? Implications of Global Retail Chains' Presence for Romania. Working paper.

Rauch, J. E. (2001). Business and social networks in international trade. *Journal of Economic Literature 39*(4), 1177–1203.

Rauch, J. E. and V. Trindade (2002). Ethnic Chinese Networks in International Trade. *The Review of Economics and Statistics 84*(1), 116–130.

Saxonhouse, G. R. (1976). Estimated parameters as dependent variables. *American Economic Review 66*(1), 178–83.

Swenson, D. L. (2008). Multinationals and the Creation of Chinese Trade Linkages. *Canadian Journal of Economics 41*(2).

Wang, S. and Y. Zhang (2006). Penetrating the Great Wall, Conquering the Middle Kingdom: Wal-Mart in China. *Wal-mart World—The World Biggest Corporation in the Global Economy (New York, NY: Routledge,Taylor & Francis Group, 2006)*, 293–313.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data.*

Wu, K. (2004). *The Happiness of Carrefour in China.* (translated), China Economy Publishing House.

Chapter 4

Exclusionary Contracts

4.1 Introduction

The debate over whether contracts can have anti-competitive exclusionary effects has been central to competition policy.[57] In cases involving tying,[58] exclusive dealing,[59] and long-term contracts[60], courts have struck down contracts as anti-competitive. The early Chicago school of law and economics argued that buyers would not enter into an exclusionary contract with a dominant firm without sufficient compensation to overcome the costs to them of exclusion. (Director and Levi, 1956; Bork, 1978). Commercial contracts must maximize the combined wealth of parties to the contracts and according to this view contracts that might appear to exclude competitors must nonetheless be efficient.

Aghion and Bolton (1987) showed this claim to be false. Contracts need not be efficient where there are externalities imposed on agents outside the contracts. Aghion and Bolton show that even a simple long-term contract can be anti-competitive in acting as a barrier to the entry. A substantial literature has since developed investigating the conditions under which contracts can profitably be used to exclude rivals.

This paper synthesizes and extends the theory of exclusionary contracts, then applies

[57] A version of this chapter has been submitted for publication. Jing. R., and Winter, R. Exclusionary Contracts.

[58] E.g., *Northern Pacific Railway v. United States*; and the *U.S. v. Microsoft* cases

[59] E.g., *United States v. United Shoe Machinery Corporation* (1922); *Milkovich v. Lorain Journal Co.*, 497 U.S. 1 (1990)

[60] *Director of Investigation and Research v. D&B Companies of Canada Ltd., CT-1994-01 (Canada), ("Nielsen").* This case is analyzed in section 5 of this paper.

the theory to a case, Director of Investigation and Research v. D&B Companies of Canada Ltd., CT-1994-01 (Canada) (*Nielsen*). *Nielsen* reveals a wider range of incentives for exclusionary contracts than has been developed in the literature as well as a rich set of complementary strategies.

In 1986, Nielsen, an incumbent monopolist in the market for scanner-based marketing information in Canada, faced the threat of entry by a second firm, IRI. Nielsen's first response to this threat was to offer selected buyers long-term contracts, with stipulated damages, in exchange for price concessions. Long-term contracts between buyers and an incumbent makes entry more difficult because to attract buyers an entrant must compensate them for the liquidated damages required to exit the contracts. The probability of entry is reduced. Why would buyers enter into contracts that would render entry, and beneficial competition, less likely? Part of the explanation is that each buyer recognizes that *if* entry is successful the entrant (having very low marginal cost and the ability to price discriminate in this market) is then likely to make a contract offer attractive enough to induce the buyer to exit the long-term contract. The stipulated damages established in the ex ante, long-term contract, in other words, serve to elicit a lower price from the entrant. The ex ante contract can thus be designed to implement a contingent transfer from the entrant to one member of the buyer-incumbent contracting pair, increasing their combined wealth. This is the central theory of Aghion-Bolton (1987). Long-term contracts can be designed to extract a contingent transfer from a successful entrant, with the effect of decreasing the probability of entry.

A second source of incentive for long-term contracts with each buyer in Nielsen is also set out in Aghion-Bolton. The entrant, IRI, faced substantial fixed costs in the purchase of raw data. The entrant therefore needed sufficient buyers free of long-term commitments (or with sufficiently low damage payments in long-term contracts) to make the investment in entry worthwhile. It is a Nash equilibrium for all buyers to enter long-term commitments with an incumbent in exchange for a very small price concession,

however: entry would be unaffected by a decision by a single buyer to decline the offer (Rasmussen et al 1991). While it would be in the collective interest of all buyers to decline the offer, a best response by any single buyer to "decline" by all other buyers is to decline the offer. The offer of long-term contracts to all buyers thus deters entry by exploiting a collective action problem among buyers. An important role is played by price discrimination in the theory, in that one set of terms can be offered to a set of buyers sufficient to deter entry, and monopolistic terms offered to the remaining buyers.[61] The ability to price discriminate strengthened the entry deterrence role of contracts in Nielsen, as predicted by Segal and Whinston (2000).

Entry-deterring contracts emerged in this case not only in the downstream market with buyers, but in the upstream market with input suppliers. The upstream suppliers in this case were 10 regional grocery chains providing raw scanner data. The case illustrates the basic principle that in establishing contracts with upstream input suppliers, incumbency does not yield a first-mover advantage (as it does in setting contracts with downstream buyers). The entrant and the incumbent competed simultaneously, over a relatively short period of time, for the rights to the raw data from each of the upstream suppliers.

The following model allows us to understand the outcome of the competition for upstream rights. Two firms bid simultaneously for the rights to each of n upstream inputs. Each firm i submits a two-part bid (b_{ij} , e_{ij}) to each upstream supplier j. The first element is a bid for the right to the input j; the second element is a bid for the exclusive right to the input j. Then each upstream firm j accepts the highest among: e_{1j}, e_{2j}, and $(b_{1j} + b_{2j})$. The decisions of the n input suppliers thus determine an allocation of rights $a = (a_1 ... \ a_n)$ with each $a_j = 1$, 2 or B (with $a_j =$ B representing an allocation of the jth input to both firms). Given the allocation, the two downstream firms compete in the downstream market with products whose value to buyers depends on the set of

[61] Without price discrimination, a prediction of entry deterrence relies on the selection of a Pareto-inferior equilibrium ("accept" by all buyers) in the acceptance subgame.

inputs incorporated in each product.

The bids by each firm for input rights reflect, of course, the profits anticipated in the downstream competition conditional upon different allocations. We show that when a pure strategy equilibrium exists in this game, the equilibrium maximizes aggregate profits of the two downstream firms in the second-stage of the game. For example, if the presence of both firms in the market is necessary for aggregate profit maximization, and a particular input is essential to each product, then an equilibrium of the bidding game will allocate the input to both firms.

Profits are maximized at an allocation of all inputs to one firm, and the equilibrium of the game is for one firm to establish exclusive contracts with all input suppliers, under two conditions: a high degree of complementarity of upstream inputs, and high inherent substitutability of the products downstream. (By inherent substitutability we mean the substitutability of the two products conditional upon the same set of inputs.) High complementarity of inputs restricts the equilibrium to three possibilities: all inputs allocated to one firm; to the other firm; or all inputs to both firms. Strong substitutability of the products downstream means that the last of these three allocations is ruled out: the prospect of strong price competition between two close substitutes means that profits are higher with a monopolist.

This model is in the spirit of Bernheim and Whinston (1998) with two differences. First, the downstream firms here compete over a number of inputs as opposed to competing for rights to representation in a single firm as in Bernheim and Whinston.[62] Second, informational constraints here restrict strategies to bids expressed in dollar values (and exclusivity-or-not) rather than bids in the form of more elaborate contracts as in Bernheim and Whinston (1998). Bernheim and Whinston (1998) is in a sense the formalization of the Bork position that exclusivity restraints will not emerge as a constraint in equilibrium contracts when there are no inherent efficiencies. We reach the opposite

[62]The competition in Bernheim and Whinston (1998) is for representation in a downstream firm rather than the rights to an upstream input, but this distinction is inessential.

conclusion under three conditions that in fact characterize our case application: strong complementarity of upstream inputs; strong inherent substitutability downstream; and bids restricted by informational conditions to a simple form. All three conditions are necessary to understand the emergence of exclusive, and exclusionary, contracts with upstream suppliers.

In addition to the three theories for exclusionary contract incentives outlined above, we argue that Nielsen involved a fourth incentive channel. This theory involves an interaction between the upstream game and the downstream long-term contract strategy. In the upstream game, if the incumbent and the entrant are nearly equal in terms of their ability to earn profits as a subsequent monopolist, then almost all prospective rents are transferred upstream in the form of bids. This is because equilibrium bid in any auction under certainty is the value of the item (here, the rights) to the agent with the second highest value. This means that any strategy that the incumbent can implement to reduce the value of the rights to the entrant leads to a lower bid for the rights–and therefore a reduction in the transfer of monopoly rents upstream. A natural such strategy is the adoption of long term contracts downstream. Hence we have an additional incentive channel for explaining the long-term contract with each downstream buyer, beyond the extraction of transfers from the entrant or from other buyers: by adding asymmetry to the upstream bidding game, the contract implements a transfer of rents from the upstream suppliers to the contracting pair.

In the remaining sections, we begin by synthesizing the theories of incentives for long-term contracts with buyers by an incumbent with a first-mover advantage. We set out a "Chicago benchmark" set of conditions under which no incentive for long-term exclusionary contracts exists: no market power outside the contracting pair and no fixed costs on the part of entrants. Entrants, in particular, are perfectly competitive and share the same realized cost. In this benchmark, maximum profits are extracted with a long-term contract in the form of a simple call option, with the exercise price

equal to the incumbent's cost of production; this contract ensures that a buyer faces a price equal to the minimum marginal cost of production among producers. We then introduce, separately, the three simplest possible deviations from this benchmark that predict exclusionary contracts: market power on the part of entrants; market power on the part of an input supplier; and fixed costs (without the possibility of profit, i.e. contestability) on the part of entrants. The idea is not that we expect to find real-world cases in which any one of these incentives is the only force at work; the point is that an understanding of the full set of incentives for exclusionary contracts is best accomplished by isolating the simplest conditions sufficient for each individual theory.

We then move up the supply chain to contracts with input suppliers: the simultaneous offer by the incumbent and entrant of contracts for the exclusive or nonexclusive rights to inputs. This section develops the bidding game outlined above. We apply the theory to Nielsen, discuss the set of strategic interactions in the market even beyond the theory developed in this paper, and conclude with a summary.

4.2 Existing Theories of Exclusionary Contracts

Introduction

Two kinds of allegedly exclusionary contracts have been considered in antitrust law and discussed in the literature. Practices such as exclusive dealing or requirements tying explicitly condition a sale on whether the customer purchases from another seller.[63] Exclusive dealing prohibits a buyer from purchasing from another seller and requirements tying similarly prohibits the buyer of a good A from purchasing any of product B from another supplier.[64] The traditional legal view of these contracts is that when imposed by a dominant firm on a substantial share of a market, the contracts are anti-competitive. If

[63] We suppose in the following discussion that it is the buyers who are subject to the exclusivity restrictions; similar logic applies when it is sellers who are so restricted and in our model we consider both simultaneously.

[64] Exclusive dealing and requirements tying contracts generalize to contracts that offer price inducements rather than absolute prohibitions against dealing with alternative suppliers.

an incumbent monopolist imposes contracts on buyers that prevent rivals from entering the market, then competition by the rivals is foreclosed. Part of the intuition underlying this view is the (incorrect) idea that a monopolist has the power to impose contract clauses that it wants, irrespective of the impact on buyers. The early Chicago School did not accept this traditional view of exclusive contracts as anti-competitive. It reasoned that if the incumbent firm is imposing an exclusivity restriction then it must compensate the buyer to accept the contract. A monopolist cannot extract maximum rents by charging the monopoly price and then extract additional rents by imposing a restraint that forecloses competition in the market. Monopoly rents can be collected only once. If we observe an exclusive contract then according to this argument it must be that efficiencies justify the lower price required to compensate the buyer for the exclusivity restraint.[65]

The second class of allegedly exclusionary contracts are particular long-term contracts. These contracts contain no reference to dealing with other suppliers. Obviously long-term contracts are ubiquitous and almost always efficient. But a number of articles (in what is sometimes called the post-Chicago school) demonstrate that even these contracts, with no explicit exclusivity restraints, can be exclusionary.

In the following, we present the simplest framework for assessing exclusionary contracts. We then provide within this framework a set of assumptions on market structure underlies that Chicago position. Departures from the benchmark structure support the post-Chicago position that even long-term contracts alone can be exclusionary.

The Economic Framework

Figure 4.1 presents the canonical structure of an incumbent facing the threat of entry. A number n suppliers upstream, producing at zero cost, provide an essential input to the

[65] Judge Robert Bork is often cited for this view. He states "The truth appears to be that there has never been a case in which exclusive dealing or requirements contracts were shown to injure competition. A seller who wants exclusivity must give the buyer something for it. If he gives a lower price, the reason must be that the seller expected the arrangement to create efficiencies that justify the lower price." (Bork (1978): 309).

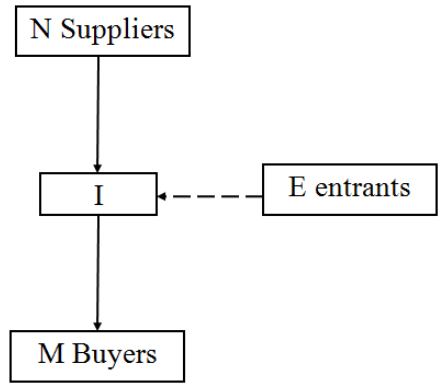


Figure 4.1: Basic Structure

incumbent, I, who supplies value-added (a second essential input) at constant marginal cost c_I and sells a product to m buyers downstream. The inputs and output are conventional private goods. To set aside any price discrimination incentives, all buyers are identical. Each buyer has a quasi-linear utility $u(q) + x$, where x is the expenditure on other products. Corresponding to this utility is an indirect utility function $v(p)$ and downward-sloping demand $d(p)$ for the product. Each unit of the final product requires one unit of the primary input and one unit of input by the incumbent. The incumbent faces the threat of potential entry into the market by a set of E entrants. The cost of production c_E on the part of the entrants is uncertain but common across entrants. The distribution $G(\cdot)$ of c_E is smooth with continuous density $g(\cdot)$ and has support given by $[0, \overline{c}]$ with $\overline{c} > v$. The density $g(\cdot)$ is bounded away from 0 on its support. Contracts can be struck between the incumbent and the buyers and/or suppliers *ex ante* (prior to the realization of the uncertainty in entrants' costs) or *ex post.* Even a buyer that has entered an ex ante, or "long-term," contract may choose to purchase from an entrant ex post; we let q_I and q_E refer to the quantities that the buyer obtains from the incumbent and entrant respectively. In the long-term contract both q_I and q_E are contractible. Finally, no resale of the product is possible among the buyers. We refer to the above assumptions as "the general framework."

4.2.1 The Chicago Benchmark

In the simple framework that we have set out the only possible incentive for exclusivity would be foreclosure, and the prediction of the Chicago School is therefore that exclusivity should not be observed. A set of assumptions within our framework supports this prediction: (A1) perfect competition in the supply of the primary input; and (A2) perfect competition among entrants (with common realized cost) in the supply of the alternative. A long-term contract between the incumbent and a buyer is a payment $R(q_I, q_E)$ from the buyer to the incumbent. For example, a contract that satisfies $R(q_I, q_E) = R(q_I, 0) + x$ contains an exclusive dealing clause with a penalty x for violating the clause.

If she turns down the offer of a long-term (i.e. ex ante) contract, a buyer will purchase from the entrants if the realized cost c_E is less than c_I. If $c_I < c_E$ ex post then given a history of no long-term contract, then following standard reasoning the ex post contract with the incumbent will be a two-part price: a variable price equal to c_I and a fixed fee that leaves the buyer with total surplus that she could obtain by purchasing from entrants at a price c_E. In short, the buyer's surplus if she declines a long-term contract is $v(c_E)$, whether she purchases ex post from the entrants or the incumbent. If a long-term contract is to be accepted, it must provide the buyer with $Ev(c_E)$; this is the individual rationality constraint on the incumbent's offer of a long-term contract. The incentive compatibility constraint requires that ex post, at each realization of c_E, the buyer chose q_I and q_E optimally given the contract $R(q_I, q_E)$ and an entrants' price equal to c_E. In short, the incumbent's optimal contract solves

$$\max_{R(\cdot,\cdot);q_I(c_E),q_E(c_E)} E[R(q_I(c_E), q_E(c_E)) - c_I q_I(c_E)]$$

subject to

$$E\{u\left(q_I(c_E)+q_E(c_E)\right)-R(q_I(c_E),q_E(c_E))-c_E q_E(c_E)\}\geq Ev(c_E) \qquad \text{(IR)}$$

$$\forall c_E \qquad [q_I(c_E),q_E(c_E)] = \arg\max_{\widetilde{q}_I,\widetilde{q}_E} u\left(\widetilde{q}_I+\widetilde{q}_E\right)-R(\widetilde{q}_I,\widetilde{q}_E)-c_E\widetilde{q}_E \quad \text{(IC)}$$

It is straightforward to show the following:[66]

Proposition 1. *("Chicago Benchmark") If entrants and upstream suppliers are perfectly competitive in the general framework, then the optimal contract is a two-part price with variable price c_I. (If demand is zero-one, the optimal contract is a call option with exercise price c_I.) The long term contract achieves the same profit as no long-term contract.*

The basis for this Chicago result is clear. No parties other than the incumbent and buyer are affected by the contract and since transfers between these two parties are possible, the contract must maximize total expected surplus. This can be achieved only by allowing the buyer complete freedom to purchase at the entrant's price, c_E, if this price is less than c_I. Under the Chicago benchmark assumptions of perfect competition, apart from the monopoly incumbent facing the threat of entry, there is no incentive for exclusionary contracts.

4.2.2 Aghion-Bolton

In discussing the post-Chicago theories, we follow Aghion and Bolton (1987) in assuming that each buyer purchases 0 or 1 units of the product and values the unit at v. Any explicit exclusivity would be superfluous under this assumption, so a contract contains no reference to purchasing from the entrant. The question is whether simple long-term contracts can be anti-competitive. A general long-term contract between the incumbent

[66] Maximization of the objective function subject to the ICC alone yields a constant variable price equal to c_I. A fixed fee can then be chosen to satisfy the IR constraint.

and the buyer contains a price p and a liquidated damage (stipulated damage) d that is paid by the buyer should she decide not to purchase from the incumbent ex post. This contract can be equally interpreted as a *call option*: the buyer pays a price d up front for the option to buy a unit ex post at an exercise price $p - d$.

Proposition 2. *(Aghion-Bolton) With a single entrant and perfectly competitive upstream suppliers, the optimal contract is a call option with exercise price, $p - d < c_I$.*

It is useful to offer a short proof of this proposition. Note that if the buyer turns down the offer of a long-term contract she will benefit from Bertrand competition between the incumbent and the entrant ex post. The ex post price following a realization c_E is therefore given by $\min(v, \max(c_I, c_E))$: if the realized $c_E \in (c_I, v)$ then the incumbent sells to the buyer at the limit price c_E; if $c_E < c_I$ then the entrant sets a take it or leave it price c_I and if $c_E > v$ then the incumbent sells at a price v. Ex ante, the buyer's expected surplus from rejecting a long-term contract offer, and relying simply upon the ex post market, is therefore given by $G(c_I) \cdot (v - c_I) + \int_{c_I}^{v} (v - c_E) dG(c_E)$. The buyer realizes surplus $v - p$ from accepting the long-term contract whether breaching or not, since the entrant (making a take it or leave it offer) extracts any surplus from the breach decision. The individual rationality constraint on the contract offer is therefore

$$v - p \geq G(c_I) \cdot (v - c_I) + \int_{c_I}^{v} (v - c_E) dG(c_E) \tag{1}$$

If a long-term contract (p, d) is signed, the buyer will exit (or breach) the contract, paying d in exchange for the right to buy from the entrant, if $c_E < p - d$ since then the entrant can offer a price that will not leave the buyer worse off. The incumbent's expected profit is therefore given by

$$G(p - d) \cdot d + [1 - G(p - d)](p - c_I) \tag{4.2.1}$$

The optimal contract maximizes (4.2.1) subject to (1). The first-order condition for the

optimal d in this optimization is

$$G(p-d) + g(p-d)[d-(p-c)] = 0 \tag{4.2.2}$$

Any contract with $p < d$ is equivalent to a contract with $p = d$,[67] which from (4.2.2) is dominated by a contract with $p > d$. Therefore $p > d$, which implies that $G(p-d) > 0$ and $g(p-d) > 0$. It is immediate from (4.2.2) that $p - d < c$. ■

The impact of adding market power on the part of the entrants to the Chicago benchmark is thus to *reduce the optimal exercise price of the call option below* c_I. The optimal contract leads to exclusion of the more efficient firm, the entrant, whenever the entrant's cost is above the optimal exercise price of the call option but below c_I, i.e. whenever $c_E \in (p-d, c_I)$. Long-term contracts are anti-competitive in this sense. The source of the inefficiency is in the incentive for the incumbent and the buyer, as a contracting pair, to extract rents from the entrant. For each dollar that d is raised, the price charged by the entrant must fall in those states in which it enters the market. The contracting pair trades off the creation of surplus in the market and the extraction of a higher share of this surplus in the same way as would a monopsonistic purchaser of the entrant's input facing the random supply at c_E. Entry is deterred in the Aghion-Bolton I model but not completely: $G(p-d) > 0$. Finally, the incumbent makes more profit in the event of breach, under the optimal contract, than in the event of production: $d > p - c_I$. (This follows, under the call option interpretation, from the fact that the exercise price is less than the incumbent's cost.)

In the Chicago theory all parties excluded from the contract compete as perfectly competitive agents in markets with free entry, and therefore bear no externalities from the contract. The set of post-Chicago theories can be organized in terms of who bears the externality from a long-term contract. In this first Aghion-Bolton theory, the source

[67] The density $g(\cdot)$ is bounded away from 0 on its support.

of the inefficiency is the externality imposed on the entrant.[68]

4.2.3 Horizontal Externalities

The second theory of exclusionary contracts turns on the exploitation of a collective action problem among the buyers. Suppose that entrant's costs include a fixed cost. In this case, the event of entry depends upon whether enough buyers are free of long-term contract obligations to justify entry. As Aghion and Bolton explain "... when one buyer signs a long-term contract with the incumbent, he imposes a negative externality on all other buyers. By locking himself into a long-run relationship with the seller, he reduces the size of the entrant's potential market so that, *ceterius paribus,* the probability of entry will be smaller. As a result, the other buyers will accept higher prices" (Aghion-Bolton 1987: 397.)[69]

The Aghion-Bolton framing of this theory, in section 3 of their paper, allows for contracts with buyers to depend on acceptances by other buyers. Rasmusen et al. (1991) (RRW) analysis of essentially the same exclusionary theory (restricted to complete exclusion) does not consistently incorporate the strategy of discriminating among buyers in contract offers (Segal and Whinston (2000)). Segal and Whinston's analysis does solve the relatively complex RRW model. The natural and simplest model, we suggest, is one that isolates the horizontal externality in the minimal departure from the Chicago benchmark. Incorporating a set of *contestable* entrants rather than a single entrant ensures

[68]The first Aghion-Bolton theory has been criticized on several grounds. As Masten and Snyder (1989) first pointed out, the Aghion-Bolton entry deterrence effect requires the following: (1) that courts force stipulated damage greater than last profit $(p - c_I)$. In fact, under the "penalty doctrine", US courts have refused to do so. (It is also an unrealistic implication of the model that the incumbent hopes that entry is *not* deterred.); and (2) no ex post renegotiation. The Aghion-Bolton effect, however, can be predicted without these conditions by recognizing specific investments on the part of the incumbent (Chung (1995), Spier and Whinston (1995)). The argument is that if the incumbent's cost c_I is reduced by specific investment v, via $c_I(v)$, then a marginal increase in v results in a reduction in the price that the entrant must change its price to attract the buy. This transfer from the entrant induces excessive specific investment on the part of the incumbent and entry deterrence. The most recent extensions of the Aghion-Bolton theory involve contracts with downstream buyers that compete with one another (Fumagalli and Motta (2006), and Simpson and Wickelgren (2007)

[69]In our model in this section, there are strictly speaking no externalities in equilibrium; the problem is one of collective action. But this is merely an artifact of our simplifying assumption of perfect certainty.

that entrants earn zero profits whatever the outcome, and isolates the horizontal externality as the source of inefficient exclusionary contracts. In the model below, entrants have fixed costs but contestability is ensured by allowing the fixed cost to be incurred after contracts are signed. The one departure from the Chicago benchmark is thus to assume contestable entrants instead of competitive entrants.

Within the general framework, we refer to the following set of assumptions as "the contestable entrants framework." Upstream supply is perfectly competitive. Downstream buyers each purchase 0 or 1 units of the product and have a common value v for the product. At least two entrants share identical, and <u>known</u>, costs consisting of a constant variable cost c_e and a fixed cost F. We assume that $c_e + F/n < v$ so that there are potential gains to trade with an entrant (otherwise the presence of entrants is irrelevant) and that $c_e + F > v$ so that entrants need more than one buyer to generate gains from trade. The ex ante contract consists of an agreement to transact at a price p and is enforceable. The incumbent may or may not be able to price discriminate in its contract offers to buyers. The ex post pricing game is Bertrand, with price offers being made to free buyers by entrants and the incumbent, and yields a Bertrand price of x to free buyers. The fixed cost need not be incurred until after buyers have accepted a price offer from an entrant (hence the phrase contestable entrants). Contracts are publicly observable, whether price discrimination is possible or not.

Within this framework, the final stage of the game is simple. If there are m free buyers (buyers who have not accepted an ex ante contract offer), the price is $x = \min\{c_e + F/m, v\}$ with the incumbent supplying the free buyers providing that $c \leq c_e + F/m$ and a single entrant supplying the buyers otherwise. In the following, $s(x)$ refers to the maximum number of acceptances that will lead to an ex post market price not greater than x. Using the notation $\lceil z \rceil$ for the minimum integer equal to or greater than z, $s(x) \equiv \min\{\text{integer } s \mid c_e + F/(n-s) \leq x\} = n - \lceil F/(x - c_e) \rceil$.

The contestable entrants framework yields multiple equilibria. In the following propo-

sition, we say that the incumbent can "assure" a profit level $\overline{\pi}$ if it can offer a contract for which any equilibrium in the subgame of acceptance decisions yields a payoff to the incumbent of $\overline{\pi}$.

Proposition 3. *Within the contestable entrants framework,*

(1) If price discrimination by the incumbent is not possible, then

(a) there exists a subgame perfect equilibrium (SPNE) with no contract offer followed by an ex post price $x = c_e + F/n$*;*

(b) a contract offer $p = v$ *and acceptance by* 0 *or by any* $s \in (s(v), n]$ *buyers is an SPNE;*

(c) for any $p \in [c_e + F/n, v)$*, there is an SPNE with a contract offer of* p *and acceptance by all buyers.*

(2) If price discrimination is possible, then

(a) an offer of $p = v$ *followed by acceptance by all buyers is again an SPNE;*

(b) by offering long-term exclusionary contracts to s(v) buyers, the incumbent can assure profit arbitrarily close to

$$\widetilde{\pi} \equiv s(v)(c_e - c + F/n) + [n - s(v)](v - c)$$

if $\widetilde{\pi} > 0$*, where* $s(v) = n - [F/(v - c_e)]$*;*

(c) Given v*,* n*,* F*, there exists* $\underline{c}_e$ *such that* $\widetilde{\pi} > 0$ *if and only if* $c_e > \underline{c}_e$*. Exclusion of efficient entrants is profitable:* $\underline{c}_e + F/n < c$*.*

Proof: We consider first the case of no price discrimination. The strategy by all buyers of rejecting any offer at or above $c_e + F/n$, combined with no offer by the incumbent is an SPNE since no buyer gains from accepting alone and the incumbent's strategy is also a best response to the buyers' strategies. This proves 1 (a). Following a price offer $p = v$ the best response by a particular buyer to rejection by all other buyers is rejection since rejection yields a price of $c_e + F/n < v$ for the buyer. Acceptance is a best response to accept by s other buyers, for $s \in (s(v) - 1, n]$, since then even rejecting yields too

few free buyers to elicit a price offer less than v by the potential entrants in the ex post pricing game. This proves 1 (b). Re (c), consider the subgame in acceptance decisions following an incumbent offer of $p \in [c_e + F/n, v)$. (A) The optimal response by any buyer to a number $\widetilde{s}$ of acceptances among *other* buyers, for $\widetilde{s} \in [s(p), n]$, is "accept" since acceptance achieves a price p for the buyer whereas rejecting leaves so few free buyers that $x > p$ and (B) the optimal response to $\widetilde{s} \in [0, s(p) - 1]$ acceptances among other buyers is to reject, since doing so yields an ex post price $x = c_e + F/(n - \widetilde{s})$, whereas accepting yields a price p; and[70]

$$x = c_e + \frac{F}{n - \widetilde{s}} \leq c_e + \frac{F}{n - s(p)} = c_e + \frac{F}{\lceil F/(p - c_e) \rceil} < c_e + \frac{F}{F/(p - c_e)} = p \qquad (4.2.3)$$

It is immediate from (A) and (B) that the two equilibria in acceptance decisions following a price offer $p < v$ are $s = 0$ and $s = n$. An example of a common buyer strategy that would support an offer of $p \in [c_e + F/n, v)$ as an SPNE is "reject any offer above p; accept any offer at p or below." This proves 1 (c).

Next we turn to the case where price discrimination is possible. The incumbent's possible actions in the first stage of the game consist of sets of long-term contract offers $(p_1, p_2, ...p_n)$ to the n buyers. (We can let $p_i = \infty$ represent "no contract offer to buyer i".) Part 2 (a) of the proposition follows from the previous argument. Regarding 2 (b), consider first the "bribe-s"strategy of offering $c_e + F/n - \varepsilon$ for small $\varepsilon > 0$ to the first $s \leq s(v)$ buyers (and $p_i = \infty$ to the remaining buyers). The following buyers' responses are all dominant strategies in the acceptance subgame: accept for $i \leq s$ and reject for $i > s$. The ex post price under this set of price offers is $x = c_e + F/(n - s)$ and the profit from the strategy, labelled $\pi(s)$, is therefore given by $\pi(s) = s(c_e + F/n - \varepsilon - c) + (n - s)[c_e + F/(n - s) - c]$. It follows (extending the domain of π to $\mathbb{R}$, with s a continuous variable) that (over the range $s \leq s(v)$)

[70] We set aside the possibility that $F/(p - c_e)$ is an integer. (The proof is easily extended to account for this case.) The strict inequality in (4.2.3) then follows, recalling the notation $\lceil x \rceil$ as the smallest integer greater than x, from this assumption.

$$\frac{\partial \pi}{\partial s} = \frac{\partial}{\partial s} s(c_e - c + \frac{F}{n} - \varepsilon) + (n-s)[c_e - c + \frac{F}{n-s}] = \frac{F}{n} - \varepsilon > 0$$

This implies that the optimal s, within the class of bribe-s strategies, is $s = s(v)$ or $s(v) - 1$. (If $s(v) - 1$ is "almost enough" acceptances to ensure $p \geq v$ then the optimal s will be $s(v) - 1$.) Taking the case where the optimum is $s(v)$ (the other case being similar), the profit from the optimal bribe-s strategy is arbitrarily close to $\widetilde{\pi}$. This proves 2 (b) of the proposition. The proof of 2(c) is straightforward ■

Do fixed costs on the part of entrants lead to exclusionary contracts? In the no price discrimination case, they may or may not. Exclusionary contracts and no contract are both equilibria. The prediction of exclusionary contracts requires an assumption that buyers cannot solve a coordination problem: the exclusionary contract equilibrium requires the selection of a Pareto-inferior equilibrium in contract acceptance decisions among buyers. As Segal and Whinston (2000) pointed out, price discrimination is the key to a prediction of exclusionary contracts. Price discrimination in the contestable entrants framework *always* yields exclusionary contracts even when the entrants are more efficient than the incumbent (providing they are not so much more efficient that $c_e + F < c$). Where $c_e + F/n < c$, ex ante contracts exclude more efficient rivals.[71]

Where $c_e + F/n > c$ there is no exclusion because the incumbent would produce alone with or without the contracts. But there is an impact of the right to offer ex ante (long-term) contracts on the disciplining power on the market of the contestable entrants even in the equilibrium most favorable to buyers. The effect is to transfer the disciplining power of the entrants entirely to the prices offered to a subset of buyers in the ex ante contract; the surplus of free buyers is extracted completely in the ex post market.

[71]The following is a numerical example of the parameters under which exclusionary contracts that assure positive profits are offered. Let $v = 10$; $n = 50$; $c = 5$ and $F = 30$. Then (using the definition of $\widetilde{\pi}$), exclusionary contracts are profitable when $c_e > 3.4$ Production by an entrant is efficient, on the other hand, when $c_e < 4.4$.

4.3 Exclusionary Contracts and Vertical Externalities

The two theories of exclusionary contracts that we have reviewed and simplified can be supplemented by a third well-established theory. If there are exogenous barriers to entry into the upstream market, and a second firm or potential competitor about to enter the downstream market, then the purchase of exclusive rights to all of the upstream inputs guarantees a monopoly downstream. The gains to trade, which are the monopoly profits protected, are shared with the upstream firms via the purchase price of the exclusive rights.[72]

Do these three theories exhaust the channels for the incentive for exclusionary contracts by an incumbent monopolist? This section considers a setting with a single supplier upstream from the incumbent, a single buyer downstream, and perfectly competitive potential entrants facing identical, random costs. The Aghion-Bolton theory does not apply in this setting because of the assumption of competitive entrants. The horizontal externalities does not apply because of the assumption of a single player at each stage of the supply chain. And the theory of monopoly through acquisition of all exclusive rights does not apply since there is already a single player upstream.

The setting isolates a fourth incentive: the exploitation of vertical externalities. If a long-term contract is possible only with the buyer, for example, then ex post a low realized cost by the competitive entrants means that—if the upstream monopolist's input is priced

[72] A case which is often associated with this theory is *Alcoa* (*United States v. Aluminum Co. of America*, 44 F . Supp. 97 (S.D.N.Y. 1941). Lopatka and Godek (1992), however, suggest a different view. The theory is incomplete as outlined, since a "hold-out" problem arises among upstream suppliers in that some may choose to wait to sell to a rival entrant. (An analogous hold-out problem is analyzed in section 4 of this paper.) Note that the theory has many variants. For example, if input suppliers vary in cost, it may be optimal for an incumbent to contract with only the major, lowest-cost input suppliers. And the profitability of exclusivity in this case may derive not from the complete monopolization through erection of barriers to entry but through raising the costs of existing rivals. This theory of vertical foreclosure is a central theme of the raising-rivals'-costs literature (Krattenmaker and Salop (1986); Salop and Scheffman (1983) Salop and Scheffman (1983), Salop and Scheffman (1987); Ordover et al. (1990)).

low enough—purchasing from the entrants justifies breaching the long-term contract and paying the stipulated damage. The buyer and input supplier share in the gains from the buyer's exiting the long run contract. A dollar increase in the stipulated damages extracts rents from the input supplier (the agent excluded from the contract) in proportion to the supplier's share of the prospective rents from purchasing from an entrant under low cost realizations. The ex-post-rent-extraction incentive for the long-term contract thus operates (once again) on the excluded party, here the upstream supplier.

In Appendix 3.1, we consider the case where the incumbent can offer long-term contracts both upstream and downstream and show that there is an additional source of gains from exclusionary contracting. Each contract serves to relax the participation constraint in the other contract by reducing the probability of successful entry in the absence of the other contract.

The vertical externality that is revealed in our case study is more involved than in the model below: the effect of long-term contracts agreed to by downstream buyers and sellers is to render the entrant disadvantaged in a bidding game for upstream inputs, with the effect that the price paid for the upstream inputs was reduced to the detriment of the upstream suppliers. Our aim in this section is not to model the vertical externality as manifest in the case, but to isolate the minimal conditions giving rise to the externality.

To develop the vertical-externality theory of exclusionary contracts, we retain much of the notation introduced: c is the incumbent's cost; c_e is the random cost shared by competitive entrants; $f(\cdot)$ and $F(\cdot)$ are the density and cdf of the random cost c_e; v is the common value for a unit of the product shared by all buyers (who purchase 0 units or 1 unit). The incumbent can strike a contract (p, d) with the buyer ex ante. Ex post (after the entrants' cost c_e is realized), the incumbent makes take-it-or-leave it offers to the input supplier and—if the buyer is not committed to a long-term contract—to the buyer as well. After the offer(s) by the incumbent, the buyer and supplier may avail themselves of the competitive alternative to the intermediate input supplied by the incumbent,

providing that the buyer exits the long-term contract, paying d. We assume that any surplus achieved from leaving the contract is shared between the input supplier and the buyer in proportions λ and $(1-\lambda)$. (To maintain the non-cooperative structure of the game, we imagine that the input supplier discovers the competitive entrants first, with probability λ, and then makes a take-it-or-leave-it offer to the buyer; with probability $(1-\lambda)$ it is the buyer who discovers the entrants.)

Starting at the last stage of this game: the buyer will breach if $p-d<c_e$, i.e. if the exercise price of the call option is less than the entrant's realized cost. We claim that the optimal contract satisfies

$$p-d<c \tag{4.3.1}$$

To see this, suppose to the contrary that $p-d>c$. (We consider the case $p-d=c$ below.) Given the contract (p,d) and an offer w by the incumbent, the buyer will breach (or be induced by the supplier to breach) if $c_e+d<p-w$. If the incumbent is to offer a non-breach-inducing input price w to the upstream input supplier, this w will be the minimum such price: $w=p-d-c_e$; if the incumbent chooses to offer a breach-inducing price, we assume that this price is $w=0$. The incumbent's payoff in the event of breach is d; the incumbent's payoff in the event of non-breach is $p-c-w=p-c-(p-d-c_e)=d+(c_e-c)$ which exceeds its breach payoff if $c_e>c$. Thus, under the hypothesis $p-d>c$, the incumbent would induce breach (by offering $w=0$) if $c_e<c$ and otherwise would offer $w=p-d-c_e$. Under this hypothesis, the expected payoff to the incumbent is therefore $E\pi_I=d\cdot F(c)+\int_c^{\bar{c}}(p-c-w)dF(c_e)=d+\int_c^{\bar{c}}(c_e-c)dF(c_e)$. The expected payoff to the buyer is $(v-p)+(1-\lambda)\int_0^c(p-c_e-d)dF(c_e)$. Note that an equal change in p and d is a means of implementing a lump-sum transfer between the incumbent and the buyer; hence the objective in designing the contract (p,d) is to maximize the sum of expected

payoffs to the two parties. This sum is

$$B = d + (v - p) + (1 - \lambda) \int_0^c (p - c_e - d) dF(c_e) + \int_c^{\bar{c}} (c_e - c) dF(c_e).$$

Under the hypothesis $p - d > c$, the marginal effect of an increase in d on combined payoffs (at any level d under the hypothesis) is then $\partial B/\partial d = 1 - (1 - \lambda)F(c) > 0$. In short, the hypothesis is inconsistent with the necessary first-order condition, $\partial B/\partial d = 0$, for an optimal contract.

This shows that $p - d \leq c$. This, in turn, implies that $d \geq p - c$, so that the incumbent's payoff under breach is not less than the payoff under non-breach. The incumbent will therefore not offer $w > 0$ ex post. The payoff to the incumbent under the established inequality, $p - d \leq c$ is therefore $E\pi_I = d \cdot F(p - d) + (p - c)[1 - F(p - d)]$ and the payoff to the buyer is $(v - p) + (1 - \lambda) \int_0^{p-d} (p - c_e - d) dF(c_e)$. Defining the sum of these benefits again as B and evaluating the marginal value of d, at $(p - d) = c$, yields $\partial B/\partial d = \lambda F(p - d) > 0$. This expression reflects the fact that in each 1 dollar increase in d extracts λ dollars in rent from the input supplier in all breach states, $c_e < (p - d)$.[73] This demonstrates the following

Proposition 4. *In the vertical externalities framework, with one buyer and one input supplier, the optimal contract satisfies $p - d < c$. Entry would be efficient, but is deterred, whenever $c_e \in (p - d, c)$.*

The parallel with Aghion-Bolton is clear. An increase in the damage measure d above the point where (given p) efficient breach would be induced is profitable because of the extraction of rents from the party outside the contract that is earning rents – in this case an upstream supplier located along the same supply chain.

All three of the models of exclusionary contracts developed to this point rely on

[73] The second effect of an increase in d, the inefficiency in the induced breach decisions, is only a second-order effect.

the assumption that the incumbent has a first-mover advantage in offering contracts. Incumbency naturally leads to a first-mover advantage in contracting with downstream buyers: a potential entrant cannot sell a unit of output until it is produced, which takes time, and complete futures contracts are impossible to write. Nothing, however, prevents an entrant from purchasing upstream inputs at the same time as an incumbent. A potential entrant cannot immediately sell a finished automobile, but it can purchase steel.

We turn next to a model where the bidding for exclusionary rights is simultaneous.

4.4 Bidding for Exclusionary Rights

Consider two firms that are supplied by n upstream suppliers, and seller to downstream buyers. For simplicity (and to match the facts of the case studied in the next section), the n inputs supplied are, *rights* such as patent rights or the rights to the use of particular information or other property. That is, the goods are nonrivalrous. Each downstream firm acquires a subset of the rights from the upstream suppliers in a bidding game described below, and the two firms then compete in the downstream market, earning profits that depend on the allocation of rights to the two firms. If the downstream output were observable, it would in general be optimal to submit contracts such as non-linear royalty contracts for the upstream inputs.[74] Competition would take the form of contract offers, as in Bernheim and Whinston (1998). We assume that outputs are not observable, so that the only feasible bids for upstream inputs are dollar amounts. Whether or not the rival has access to an input is observable, so that bids can be conditioned upon that event. The questions we ask are how the payoffs in the non-cooperative bidding game compared to the total profits that could possibly be achieved in the market; whether there is incentive to enter into exclusive bids; and when the equilibrium in the bidding

[74]For example, if $n = 1$, then the single upstream input supplier by accepting the appropriate royalty contracts from downstream firms could elicit the prices downstream that maximized total industry profits.

game for upstream rights will assign all rights to one firm, so that the outcome is an exclusionary *set* of contracts that ensures a monopoly for one downstream firm.

We consider in particular the following game. First, the downstream firms $i = 1, 2$ simultaneously submit bids (b_j^i, e_j^i) to each of the n upstream suppliers; b_j^i is a bid by i for the (shared) right to j's input; e_j^i is a bid for the exclusive right. Next, each upstream supplier j accepts bid(s), choosing the maximum from $\{b_j^1 + b_j^1,\ e_j^1,\ e_j^2\}$. The result is an allocation $a \equiv \{a_1, ...a_n\}$ with $a_j \in \{1, 2, B\}$ where $a_j = B$ indicates that the input has been allocated to both firms. That is, each input j is allocated to $1, 2$ or both. The two downstream firms earn profits $\pi_1(a)$ and $\pi_2(a)$. These profit functions are an exogenous reduced form summary of the payoffs from downstream competition. We have in mind that the profit functions represent the payoffs from a differentiated Bertrand competition subgame, in which the value of either good, 1 or 2, to buyers depends on the set of upstream inputs incorporated in the downstream product. (In the case study following, the downstream product is an aggregation with value added, of upstream, geographically-differentiated, raw information inputs.) $\pi_i(B, ...B) > 0$ because of, for example, *inherent* product differentiation in a Bertrand pricing game, as opposed to product differentiation induced by the assignment by a of different inputs to the two firms.[75]

Define $a^* = \arg\max_a \pi_1(a) + \pi_2(a)$. This is the allocation that would be chosen by the entire set of firms, upstream and downstream, if lump sum transfers were possible. It is the maximum industry profits subject to the constraint that given a, firms 1 and 2 compete downstream. The optimum a^* can in principle take on any one of seven configurations: all $a_i = 1$; all $a_i = 2$; all $a_i = B$; all $a_i = 1$ or 2; all $a_i = 1$ or B; all $a_i = 2$ or B; and some $a_i =$ each of $1, 2$ and B. For example, if inherent product differentiation is high enough that both firms produce in the optimum; if a subset of inputs, M, is critical to the production of either product; and if simultaneous purchase of any input

[75] Inputs could in principle be cost-reducing rather than value-adding in the downstream market.

outside of M would greatly reduce product differentiation (thereby making downstream price competition more intense) then a^* would allocate the inputs in M to both firms and the remaining inputs exclusively to one firm or the other.

We assume that each profit function is monotonically increasing in the set of inputs allocated to the firm and decreasing in the set of inputs allocated to the rival firm. We also impose the following concavity-type condition on the profit functions π_i: Define a local optimum of π_i as any allocation a for which changing any a_i individually does not increase π_i. We impose the condition of a unique local optimum, i.e. that the global optimum for π_i is the unique local optimum.[76]

It is useful to provide some intuition by previewing the results before characterizing the equilibrium. Note that an equilibrium may not exist: suppose for example that $n = 10$, and a^* assigns all inputs to the firm 1 which then earns monopoly profits of 100 downstream. The most that firm 1 could pay for each input on average is 10. But firm 2, a close inherent substitute, may respond by paying a total of 30 for 3 inputs: duopoly profits are less than half monopoly profits, but they may exceed 30%. In other words, 3 input suppliers may "hold out" for a higher bid from 2. So a^* is not an equilibrium. But then any allocation other than a^* is also not an equilibrium, since in any duopoly between close substitutes, the total monopoly profits exceed the duopoly profits; therefore 1 will out-bid 2 for any particular subset of inputs. An equilibrium does not exist in this case. On the other hand, there may be multiple equilibria. The collective profit-maximizing allocation may be B, but this requires the two bidding firms to coordinate on bids b_1 and b_2; $b_1 = b_2 = 0$ for example is always an equilibrium. We are able to show that where an equilibrium $\widehat{a}$ does exist, a^* is also an equilibrium. If we take as an equilibrium concept the Pareto optimal selection among the equilibrium bids, then total-profit-maximization

[76]The following is a sufficient condition for the the "unique local optimum" condition. Change the notation slightly to denote $a_j = (1,0), (0,1)$ or $(1,1)$ depending on whether input j is allocated to 1, 2, or both. Then define the profit functions on the resulting lattice. The condition is that the profit function π_i can be extended from $\{0,1\}^{6n}$ to a strictly concave function on $[0,1]^{6n}$.[no: check discrete convexity literature]

is a necessary but not sufficient condition for equilibrium.

Proposition 5. *Under the unique local optimum condition, either a^* is an equilibrium, or no pure strategy equilibrium exists.*

Proof: Suppose that a pure strategy equilibrium $\widehat{a}$ exists. We must show that a^* is an equilibrium. One of the following conditions must hold:

(1) $\widehat{a}$ is a local optimum for $\pi_1(a)+\ \pi_2(a)$.

(2) changing $\widehat{a}_i$ from 1 to 2 or from 2 to 1, for some i, will increase $\pi_1(a)+\ \pi_2(a)$.

(3) changing $\widehat{a}_i$ from B to 1 or from B to 2, for some i, will increase $\pi_1(a)+\ \pi_2(a)$

(4) changing $\widehat{a}_i$ from 1 to B or from 2 to B, for some i, will increase $\pi_1(a)+\ \pi_2(a)$.

If (1) holds then by the unique local optimum condition, $\widehat{a} = a^*$, and the proposition is proved. We show that conditions (2) and (3) lead to contradictions and thus cannot hold. If condition (4) holds, we find an equilibrium that implements the change indicated, increasing $\pi_1(a)+\ \pi_2(a)$. Applying the procedure described repeatedly yields a sequence of Nash equilibria along which $\pi_1(a)+\ \pi_2(a)$ is strictly increasing. Since the number of allocations is finite, any such sequence must end with a^*, thus proving that a^* is an equilibrium.

To begin, suppose that condition (2) holds and that $\pi_1(a)+\ \pi_2(a)$ is increased when $\widehat{a}_i$ is changed from (say) 1 to 2. Let $\widetilde{a}$ be the resulting allocation, and adopt the notation of $\pi_j(...1...)$ as the profit function for firm j with 1 as the ith element, all other elements fixed at their common values under $\widehat{a}$ and $\widetilde{a}$, and similarly for $\pi_j(...2...)$. Then

$$\pi_1(...1...) + \pi_2(...1...) < \pi_1(...2...) + \pi_2(...2...) \tag{4.4.1}$$

But the fact that $\widehat{a}$ is an equilibrium, with $\widehat{a}_i = 1$ implies that the value by 1 for exclusive use of asset i (as compared to use by 2 only) is no less than the corresponding value on the part of 2. That is, $\pi_1(...1...) - \pi_1(...2...) \geq \pi_2(...2...) - \pi_2(...1...)$. This contradicts (4.4.1). (2) cannot hold.

Suppose that (3) holds. Then (using analogous notation)

$$\pi_1(...B...) + \pi_2(...B...) < \pi_1(...1...) + \pi_2(...1...) \tag{4.4.2}$$

The fact that $\widehat{a}_i = B$ means that in the equilibrium $\widehat{a}$

$$b_i^1 + b_i^2 \geq e_i^1 \tag{4.4.3}$$

The best-response requirement that b_i^1 be no greater than 1's profit from the right to input i (and similarly for 2) imply:

$$b_i^1 \leq \pi_1(...B...) - \pi_1(...2...) \tag{4.4.4}$$

$$b_i^2 \leq \pi_2(...B...) - \pi_2(...1...) \tag{4.4.5}$$

Recalling that that $\widehat{a}_i = B$, and that $\widehat{a}$ is an equilibrium, not that 1 had the option of raising e_i^1 to achieve $e_i^1 = b_i^1 + b_i^2$, i.e. to achieve exclusive use of input i. This strategy would have involved $e_i^1 = b_i^1 + b_i^2$. Since this option was not chosen, we know that the gain in profits from moving to this strategy was less than the increase in payment, which would have been $e_i^1 = b_i^1 + b_i^2$. That is,

$$\pi_1(...1...) - \pi_1(...B...) \leq b_i^2 \tag{4.4.6}$$

(4.4.5) and (4.4.6) imply that $\pi_1(...1...) - \pi_1(...B...) \leq \pi_2(...B...) - \pi_2(...1...)$, i.e. that $\pi_1(...B...) + \pi_2(...B...) \geq \pi_1(...1...) + \pi_2(...1...)$, contradicting (4.4.2). Hence (3) cannot hold.

It remains to show that in case (4), a equilibrium set of strategies exists that implements the change from 1 to B (or analogously, from 2 to B). Consider the following pair of strategies (labelled s_1 and s_2). Let s_1 be $e_i^1 = \pi_2(...2...)$ and $b_i^1 = \pi_2(...2...)-$

$\pi_2(...B...)$; let s_2 be $e_i^2 = \pi_2(...2...)$ and $b_i^2 = \pi_2(...B...)$; and let B be the choice for outlet i. For firm 2, s_2 earns zero profits but is a best response to s_1 and B because b_i^2 cannot be lowered in an attempt to gain non-exclusive rights more cheaply and raising e_i^2 to gain exclusive rights would leave 2 with negative profits. Consider firm 1's strategy s_1. Firm 1 cannot lower b_i^1 to gain non-exclusive rights more cheaply. If firm 1 raised e_i^1 to gain exclusive rights, its net profits Π would be less than $\pi_1(...1...)$ minus its current bid $e_i^1 = \pi_2(...2...)$, i.e. $\Pi < \pi_1(...1...) - \pi_2(...2...)$. Its profits under s_1 given s_2 and B are $\pi_1(...B...) - [\pi_2(...2...) - \pi_2(...B...)] = [\pi_1(...B...) + \pi_2(...B...)] - \pi_2(...2...)$ which, by hypothesis in case (4), is greater than $\pi_1(...1...) - \pi_2(...2...)$ which exceeds Π. Finally, B is clearly a best response by outlet i to s_1 and s_2. ■

To address our central question—when will the equilibrium set of contracts result in exclusion of one firm from the market?—we rely on the case of symmetric inputs, in assuming that $\pi_i(a)$ depends only on the numbers of inputs assigned to each firm. (This will be the case if the values of the two products to final buyers depend only on the numbers of inputs incorporated in each product.) Define the profit functions in this case as $\widehat{\pi}(n_1; n_2)$, i = 1; 2, where (n_1; n_2) are the numbers of inputs allocated to the two firms. Then the equilibrium $\widehat{a}$ will be $(1, ..., 1)$, i.e. will assign all inputs to firm 1, if $a^* = (1, ...1)$ and the following condition holds ("no holdout"): there is no $m < n$ such that $\widehat{\pi}_1(n, 0)/n < \widehat{\pi}_2(n - m, m)/m$. If $a^* = (1, ..., 1)$ and the no-holdout condition fails, then there is no equilibrium.

The central question is thus reduced to a necessary and sufficient set of two conditions: a*=(1...1) and the no-holdout condition. When will these be satisfied? To answer this question, we must get underneath the exogenous profit functions and into the conditions in the downstream market. To do this, we restrict the model still further by relying on a parameterizations of the symmetric case. Consumers are uniformly distributed along a unit line segment between two downstream firms and have a common transportation

cost, t. (Higher t will represent lower inherent downstream substitutability of products.)[77] The two downstream firms have zero production cost, so once we have gross values v_1 and v_2 of consumers for the two products, it is straightforward to compute profits. The willingness of a consumer to pay for a particular downstream product, i = 1 or 2, depends on the number of inputs m_i embodied in good i, and is given by m_i^θ. The parameter θ measures the value and complementarity of the upstream inputs.[78] Costs upstream are zero and the only costs downstream are the fixed costs of purchasing the rights to inputs.

The parameters of the model are t, θ; and n. Figure 4.2 below illustrates, for n = 10, the sets of parameters t and θ for which (1) the privately efficient allocation a* assigns all inputs to the same firm; and (2) the parameters for which this allocation is implemented in the bidding game. When upstream complementarity is sufficiently high, the only possible values for a* assign all rights to one firm or all rights to both firms, i.e. a* = (B,B,...B), since the inputs must be used together. If we add the condition of sufficiently high inherent substitutability (low t), then the allocation (B,B,...B) is ruled out by the intensity of competition that would drive down profits were both firms to acquire the inputs. This leaves exclusivity as the privately efficient outcome with low t and high θ. For this exclusivity outcome to be implemented by the auction, however, the hold-out problem must be overcome. This requires even stronger upstream complementarity and/or downstream inherent substitutability because when the profitability of exclusivity at one firm is marginal, it is relatively easy for the other firm to out-bid its rival for a subset of the inputs. Its bid reflects a sharing of the prospective profits among only this subset of input providers. In short, the central prediction of the simultaneous

[77] The assumption about inherent substitutability of downstream products is motivated by the Nielsen case, which will be discussed in section 4.5. In that case, conditional upon the same data inputs, the products from Nielsen and IRI were very similar, which indicates a low level of inherent product differentiation. This inherent substitutability implies that the downstream competition could be intense if IRI and Nielsen both exist in the market.

[78] The assumption about strong complementarity among upstream inputs is motivated by the Nielsen case. In that case, scanner data were collected from grocery chains in Canada. Although the data from the same regions were presumably functional substitutes, evidence indicated that a national data set, comprising data from all regions, was the product that Nielsen and IRI judged to be of the highest value.

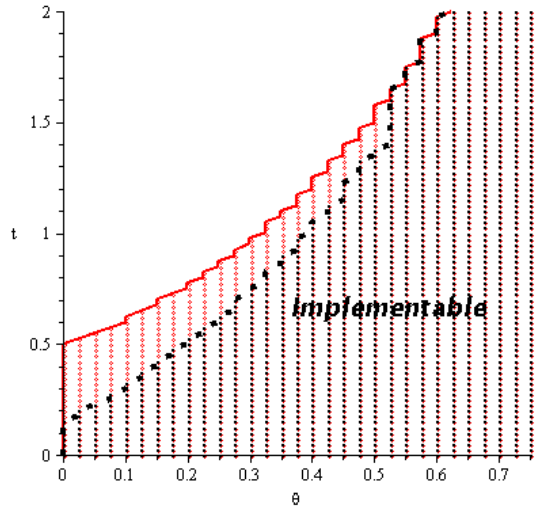


Figure 4.2: Implementable Condition Demonstration

bidding model is exclusivity, resulting from a single winner of all simultaneous bidding games, providing that three conditions hold: (1) sufficient complementarity upstream; (2) sufficient inherent substitutability downstream; and (3) informational conditions that restrict bids to dollar values rather than contracts.

4.5 Application: Nielsen

We have outlined four channels through which an incumbent firm and its buyers, suppliers or both have the incentive to enter exclusionary contracts – contracts that deter a rival from entering a market. In this section we illustrate the incentive with a Canadian competition policy case, *Nielsen.*[79] Nielsen, wholly owned by D&B, had a monopoly in Canada over the provision of market-tracking services for grocery store produce sales, when it was threatened in 1985 with the entry into the market by Information Resources Incorporated (IRI). IRI is a U.S. firm with which Nielsen shared the U.S. market in approximately equal market shares at the time. The products at issue in the case are a

[79] *Canada (Director of Investigation and Research) v. The D & B Companies of Canada Ltd. (1995), 64 C.P.R. (3d) 216 (Comp.Trib.)* ("Nielsen")

combination of software and information that allowed tracking of market shares, estimation of demand elasticities and responsiveness of demand to product promotions, and so on. The downstream buyers of these information products are mainly manufacturers of grocery products. The key inputs required are raw scanner data provided by the major grocery chains, 11 chains in Canada in 1985. Conditional upon the same raw data inputs, the Nielsen and IRI products were very similar but not identical.[80] Some important product differentiation arises, however, due to the fact that Canadian subsidiaries of U.S. firms prefer the product adopted by the U.S. parent because of complementarities in using the same software and informational products. In the upstream market, scanner data from grocery chains in the same regions were presumably functional substitutes, but evidence indicated a strong complementary in that a national data set, made up of data from all regions, was the product that Nielsen and IRI judged to be of highest value. In short, the market was characterized by strong complementarity in *upstream* inputs and strong substitutability between the *downstream* information products. Finally, we refer to Nielsen as the incumbent because it was established in the broad market for market-tracking services, but the scanner-based information products were in development in the mid-1980's.[81]

The case involved a challenge by the Canadian competition authority, the Director of Investigation and Research (now called the Commissioner of Competition), of two sets of Nielsen contracts. With the threat of IRI's entry starting in 1985, Nielsen had entered into 5 year exclusive contracts with all of the upstream grocery suppliers of scanner in 1986, contracts that contained liquidated damage clauses and prohibited the sale of scanner data to any other party. Nielsen had also entered into long-term (3-5 year) contracts with a set of downstream buyers (grocery product manufacturers); until then, Nielsen's downstream contracts had been evergreen contracts that were terminable on

[80] We label this feature as low *inherent* product differentiation in the discussion of our theoretical model.

[81] Nielsen introduced the full scanner-based information product in 1988, after the main events on which the case focussed.

8 month's notice (in those contracts entered as evidence). The Director's challenge of both sets of contracts before the Canadian Competition Tribunal was successful. The Tribunal prohibited the contracts.

4.5.1 Competition for Exclusive Contracts with Upstream Suppliers

As in our model in section 4.4, the incumbent Nielsen did not have a first-mover advantage. In fact, as Nielsen emphasized in its evidence, the potential entrant, IRI, was the first to offer an exclusive contracts. The bidding was not literally simultaneous as in our theory, but was concentrated in a few months in the summer of 1985; our adoption of the usual assumption of simultaneous competition is a better fit than usual to the facts of the case. And consistent with the model, the principal elements in each contract were the price for upstream data and parameters of exclusivity rather than more complex royalty schemes.

The market for rights to the data inputs, in short, was one in which competition was intense — but the competition was for rights to the upstream inputs, not competition within the output market. Does this type of competition in some sense substitute for competition within the market — or provide any welfare benefits at all? Under the facts of this case, the *substitutability* or low inherent product differentiation downstream and the *complementarity* of inputs upstream, the equilibrium outcome of competition for the rights to inputs was a monopoly no matter how intense the competition, i.e. no matter how symmetric the positions of Nielsen and IRI were in their potential for exploiting the monopoly position.

The socially optimal allocation of input is clearly an allocation to both firms — especially because the input as a non-rivalrous good can be supplied to the second firm at zero cost.[82] The benefits of the non-exclusive allocation are two-fold: providing greater

[82] An element of the case that makes this conclusion *sui generis*, i.e. not generalizable, is that the

product variety in the market (in allowing, for example, greater matching of software between Canadian subsidiaries and U.S. parents) and allowing price competition downstream instead of monopoly pricing. The model and economic principle — the conflict between privately and socially efficient contracts — generalize to the case where inherent product differentiation is strong enough that the equilibrium outcome is not a monopoly. Suppose that product differentiation is so strong that total industry profits would be maximized by the presence of both firms in the market. In general that industry profits (upstream and downstream) will be maximized by allocating some raw inputs exclusively to Nielsen and some to IRI: the difference in the allocation of inputs translates into greater differentiation and therefore less intense price competition in the output market. If the two firms were cooperatively choosing the allocation of inputs, and then competing, in general some exclusivity but not complete exclusivity may well result. Partial exclusivity can increase profits when it results in two firms competing because of the "competition-dampening effect" of exclusive dealing: increasing the sets of inputs to which firms have exclusive rights increases product differentiation in the final market, which dampens price competition, and raises equilibrium prices and profits. Firms' choices of how many input suppliers to sign up exclusively would trade off the private benefits of the competition-dampening effect with the costs of reduced product value. Again, however, the social optimum involves no exclusivity because this maximizes the value of each product to any purchaser (at zero social cost) and enhances downstream price competition, bringing prices closer to marginal cost.

Competition *for* the market in the form of competition for rights to upstream data inputs, in short, does not substitute for competition *within* the market. It does, however, yield one simple efficiency benefit. Suppose that the two firms that are bidding for

conclusion relies on the fact that the upstream data were produced at essentially zero cost as a by-product of production. In the case of patent rights, for example, the law would not properly strike down exclusivity since a contract would then merely transfer exclusive property rights granted by a patent to a downstream firm. And the collective acquisition by one firm of all patent rights would not be an example of anti-competitive patent-pooling if the upstream rights were complementary, analogous to this case. While the normative conclusions do not easily generalize, the positive analysis does.

exclusive rights have positive costs, rather than zero costs, with constant marginal cost. Under a mild restriction on demand, the result of the bidding game is that at least the "right" monopolist is chosen. Whichever monopolist, Nielsen or IRI, would produce the greater social surplus is the one that would win the game.[83] Two other aspects of the strategic interaction between the firms reviewed below, however, distort even this modest efficiency outcome and leave us with the Aghion-Bolton type of prediction that the higher cost (or lower surplus) firm may survive as a monopolist in this market.

The key effect of intense competition for exclusive rights, when the downstream firms are symmetric in demand and costs and product differentiation is relatively low so that monopoly is the outcome, is a *shift in monopoly rents upstream* to the suppliers of the raw data as the price for the data is bid up to the present value of resulting monopoly profits. The scarce input was the raw data, not the ability to manage a monopoly downstream. The suppliers of raw data, the grocery store chains, were principle beneficiaries of the contract exclusivity. Any asymmetry in the bidding game that Nielsen was able to create — to foreshadow the implications of the downstream contracts — acted to increase Nielsen's share of the increase in aggregate industry profits attributable to exclusivity.[84]

4.5.2 Nielsen's Downstream Contracts

The terms of Nielsen's contracts with downstream purchasers of their information products jumped from less than 1 year (terminable on 8 months' notice) to 3 to 5 years as soon as IRI attempted to enter the industry. The internal documents of Nielsen suggest that the management had just read the Aghion-Bolton working paper. These documents

[83]The restriction on demand is that the percentage difference in demand between Nielsen's product and IRI's product be independent of price. Under this assumption, the product generating the higher profit is also the product generating the higher total surplus.

[84]An outcome of the bidding game, arguably attributable to the intense competition, was that all contract lengths in the exclusive contracts were initially identical, at 5 years. Setting contract lengths the same across all input suppliers (if this can be committed to) maximizes the value of the bids facing the owners of the raw data because these owners know that when the identical contract terms are all up, the same competition for the next set of contracts will be intense and a higher price will be forthcoming at that time. Collectively at least, the suppliers of raw data would be better off accepting contracts with identical terms and this fact may explain the offer of identical contract lengths.

indicated that the strategic purpose of the shift in contract lengths was to deter the entry of IRI by "locking up" customers in long-term contracts. The contracts contained liquidated damages payable to Nielsen if the customer terminated the contract.

The horizontal externalities theory applies here because each client would view the probability of IRI entering — an event which with positive value for the client — almost unaffected by its own decision to accept the long-term contract. Only a small "bribe" in terms of a lower price would be necessary to induce the client to sign the long-term contract. The first Aghion-Bolton theory applies as well: the stipulated damage clauses was of low expected cost to the downstream customer at the time of contracting in part because even if IRI were to be successful in entering, IRI would in negotiations with the buyer effectively pay for part of the stipulated damage since this damage would reduce the joint surplus which the negotiations would allocate. By raising the stipulated damage, the incumbent and buyer in any downstream contract were implementing a transfer away from IRI, contingent on the states of successful entry, to the pair of them.

Just as in the horizontal externalities theory, the ability of the incumbent to discriminate in long-term contract offers was an important ingredient in implementing exclusivity. Nielsen did not induce all customers to sign long-term contracts but instead targeted the Canadian subsidiaries of U.S. customers of IRI. It was the loss of these buyers to which Nielsen was most vulnerable, and the gain from signing long-term contracts with them was the highest.

Finally, a vertical externality as analyzed in section 4.3 of this paper applies. The fact that Nielsen as the incumbent was able to enter the *downstream* contracts described, provided it with an asymmetric advantage over IRI in the *upstream* bidding game for the exclusive rights to the data. IRI's willingness-to-pay for the upstream data was surely reduced by the disadvantage it faced in overcoming the long-term contracts downstream. The long-term contracts downstream thus imposed a negative externality, and extracted a transfer, not just from IRI but also from upstream data suppliers in allowing Nielsen

to win the upstream game with lower bids.

One effect of this vertical externality is to negate even the modest efficiency property that we claimed for the upstream bidding game. It no longer follows that Nielsen would be forced out of the market in the event that it was not the "right" monopolist: the advantage transferred from the downstream contracting game to the upstream game leads to the possibility of an Aghion-Bolton type of inefficiency in allowing an inefficient incumbent to remain as a monopolist.

The *change* in contract length is key testable implication of the theory and the basis for the Tribunal's decision to strike down these contracts: the Tribunal was able to assume that the short contract length, established prior to the threat of IRI entry, optimally balanced the efficiency benefits and costs of longer term contracts in the absence of the competitive threat. The change in contract length was reasonably attributed entirely to the new threat of entry.

4.5.3 Renegotiation and Staggered Contracts

Let us return to the upstream contracts. After signing contracts with identical (5 year) terms with all of the data suppliers, Nielsen recognized that 5 years later (in the summer of 1991) it would potentially face the identical bidding war with IRI for the rights to the essential inputs. The prospect was again competition for the right to be the monopolist — competition that shifted rents upstream. Nielsen renegotiated contracts with two suppliers including Safeway, the largest supplier. The effect of contract staggering was not a monopoly — this market structure was already guaranteed by exclusivity whether contracts were staggered or not — the outcome was a barrier to entry*into the position of being the monopolist* in the market. Just as with the vertical externality discussed above, this staggering of contracts negates the modest efficiency property of the upstream bidding game. The social cost of this staggered contract strategy was, at a minimum, that

the most efficient monopolist would not necessarily occupy the market.[85] The profitability of the staggered contract strategy is not explained simply by its profitability to Nielsen. The two suppliers voluntarily renegotiated their contracts. It is the external effect or transfer of wealth away from the suppliers of data to the *pair* of parties undertaking any contract renegotiation that is the key to explaining the strategy.

4.5.4 Most-favored Nation Clauses

An additional issue that arose in Nielsen is that of preferred supplier contracts, or most-favoured nation (MFN) clauses in the upstream contracts. These were terms whereby Nielsen would be guaranteed that its price would not be higher than a price at which the data were subsequently sold to another buyer such as IRI. Two of Nielsen's contracts entered in 1994 contained MFN clauses, in addition to exclusivity clauses as Nielsen apparently recognized the risk that the latter would be struck down. In some circumstances, an MFN clause is reasonable. It ensures, for example, that the first purchaser of the input is not disadvantaged in downstream competition with a rival who is able to strike a more favourable price. (Because of the zero marginal cost of the input there is a risk that a lower price might be struck subsequently with a rival.)

Suppose in this case that exclusivity were struck down in these contracts. Could the MFN clauses, if they were allowed, have the effect of exclusivity? A example shows that

[85] The strategy of staggered contracts was not in and of itself challenged by the government in the case, for an obvious reason. The prohibition of staggered contracts would be an unworkable remedy. Requiring a firm to coordinate the beginning and ending dates of its contracts with suppliers would be simply too intrusive and inefficient. An additional aspect of upstream contracting is that the offer of contracts to the 11 upstream suppliers of the raw input (data) were not simultaneous, as we have assumed for simplicity in our discussion, but sequential. Because of the technological complementarity of the inputs, this introduced a *hold-out* problem: the last supplier of the complementary inputs in the sequence of offers-and-acceptances was in a position to hold out for a high price. (This problem is familiar, for example, in development of urban properties where a bundle of privately owned lots must be purchased in sequence by a developer.) IRI's strategy, in anticipation of the risk of not obtaining all necessary inputs, was to make each contract contingent upon successful completion of all of the contracts via an option on its part to abandon the contracts if not all grocers were signed up. This strategy, however, backfired. The mechanism put the last supplier of the input, Safeway, in a position to break the entire strategy (and to command a high price from Nielsen for doing so). This is what happened, and Nielsen won the competition for exclusive rights to data.

they could. To keep the analysis simple, imagine that there is a single upstream supplier of raw data, that the monopoly profits that could be earned with the data are 10 dollars and that the profits that could be earned by each duopolist in the market would be 3 dollars. (The monopoly profits thus exceed the sum of duopoly profits.) If the incumbent monopolist tried to bargain for a low price, say 5 dollars, for the input, then the MFN would not deter entry. The supplier of the raw data would willingly accept 3 dollars from the new entrant even with the MFN restraint leading to a reduction of 2 dollars in the incumbent's price. If the incumbent offered a price of 6.50, however, the entrant would be deterred. In short, the combination of MFN plus the offer of a *high* price deters entry. The Tribunal, convinced of this argument, struck down the MFN clauses, albeit with a time-limited order. The assessment of high prices in combination with the MFN clauses as exclusionary contrasts with the traditional legal view of *low* prices as exclusionary, as in predatory pricing cases.

4.5.5 The Impact of Nielsen: Exclusion via Implicit Contracts

Did the decision in this case transform the market for scanner-based information products from one with intense competition *for* the market to one with competition *within* the market, as in the U.S. No. IRI competes in 8 countries around the world, but the market remains a Nielsen monopoly in Canada.

The Tribunal recognized (*Nielsen*, p.95) that grocery retailers might decide to continue to offer their data to only one customer even once any exclusivity inducements by Nielsen were prohibited (although they regarded the possibility as unlikely). In fact, this is exactly what has happened. IRI is established in 8 countries around the world but not in Canada. The exclusivity agreements have continued in what economists would label *implicit contracts*: each grocery supplier of raw data has apparently recognized that if it were to break the implicit agreement by selling the data to IRI as well as Nielsen, then the downstream monopoly would soon be replaced by a duopoly of close substitutes in

which marginal costs were close to zero. The monopoly rents — which, as we have discussed, flowed almost entirely upstream to the grocers as suppliers of the scarce data — would disappear. While there is an obvious potential free-rider problem among grocery retailers in maintaining these implicit contracts, this free-riding problem is mitigated by the presence of a small number of large retailers whose participation would be necessary for IRI to enter.

The effect of these implicit contracts is that competition has not been restored in the market. The lesson, however, is not that the decision was wrong ex ante. It was uncertain at the time of the case as to whether the explicit contracts were necessary to sustain a monopoly. But these contracts were sufficient to sustain the monopoly and the Tribunal was right in prohibiting the contracts. The *Nielsen* decision was the right decision *ex ante* in spite of its complete lack of impact ex post.

4.6 Conclusions

This paper has synthesized the set of channels through which participants in a market have the incentive to enter into exclusionary contracts. Three of these theories operate in a market in which incumbency provides a first-mover advantage in offering contracts to buyers; the fourth yields exclusionary contracts, under some conditions, in a model in which two downstream firms bid simultaneously for the rights to upstream inputs. We examined an antitrust case that illustrated all four incentive channels as well as a set of strategic issues related to exclusionary contracts: the division of the rents from monopolization to the upstream firms and the winning monopolist, the dependence of this division in particular on the extent of symmetry between downstream bidders; the role of downstream contracts in rendering this division of rents more favorable to the winning bidder by rendering the bidding game more asymmetric, through the vertical externality which we introduced into the theory; the role of contract renegotiation in

staggering the contracts to further shift the division of rents and to create not a monopoly, but potentially a barrier to the success of the "right" monopolist; Most favored nation contracts as substitute for exclusive contracts; finally, the success of implicit as opposed to explicit exclusionary contracts.

The Aghion-Bolton perspective on exclusion is to identify the incentives for anti-competitive exclusionary contracts in terms of transfers from agents outside the contracts. The synthesis that we offer isolates each incentive in the simplest departure from a benchmark in which privately optimal contracts are efficient. The approach is valuable beyond the static framework that we have explored in this paper.[86]

[86]Suppose, for example, that in an evolving industry the probability of discovering the next generation technology is higher for firms operating in current market or that there is learning-by-doing of any other form, and that the set of buyers changes to some degree over time. In this dynamic market setting, exclusionary contracts can be explained in part as implementing a transfer from future buyers to the current market participants. Exclusionary contracts are potentially of even greater cost, and the Aghion-Bolton perspective more valuable, in dynamic market settings.

Bibliography

Aghion, P. and P. Bolton (1987). Contracts as a barrier to entry. *The American Economic Review 77*, 388–401.

Bernheim, D. and M. Whinston (1998). Exclusive dealing. *Journal of Political Economy 106*, 64–103.

Bork, R. (1978). *The Antitrust Paradox*. New York: Basic Books.

Chung, T.-Y. (1995). On strategic commitment: Contracting versus investment. *The American Economic Review 85*(2), 437–441.

Director, A. and E. H. Levi (1956). Law and the future: Trade regulation. *Northwestern University Law Review 51*, 281–296.

Fumagalli, C. and M. Motta (2006). Exclusive dealing and entry, when buyers compete. *The American Economic Review 96*(3).

Krattenmaker, T. G. and S. C. Salop (1986). Competition and cooperation in the market for exclusionary rights. *The American Economic Review 76*(2), 109–113.

Lopatka, J. E. and P. E. Godek (1992). Another look at alcoa: Raising rivals' costs does not improve the view. *Journal of Law and Economics 35*(2), 311–329.

Masten, S. E. and E. A. Snyder (1989). The design and duration of contracts: Strategic and efficiency considerations. *Law and Contemporary Problems 52*(1), 63–85.

Ordover, J. A., G. Saloner, and S. C. Salop (1990). Equilibrium vertical foreclosure. *The American Economic Review 80*(1), 127–142.

Rasmusen, E. B., J. M. Ramseyer, and J. Wiley, John S. (1991). Naked exclusion. *The American Economic Review 81*(5), 1137–1145.

Salop, S. C. and D. T. Scheffman (1983). Raising rivals' costs. *The American Economic Review 73*(2), 267–271.

Salop, S. C. and D. T. Scheffman (1987). Cost-raising strategies. *The Journal of Industrial Economics 36*(1), 19–34.

Segal, I. R. and M. D. Whinston (2000). Naked exclusion: Comment. *The American Economic Review 90*(1), 296–309.

Simpson, J. and A. L. Wickelgren (2007). Naked exclusion, efficient breach, and downstream competition. *The American Economic Review 97*(4), 1305–1320.

Spier, K. E. and M. D. Whinston (1995). On the efficiency of privately stipulated damages for breach of contract: Entry barriers, reliance, and renegotiation. *The RAND Journal of Economics 26*(2), 180–202.

Chapter 5

Conclusions

Chapter 2 studies the determinants of the size of equity stakes that downstream producers hold in their upstream input suppliers. This essay contributes to the literature on the following three aspects. First, in its analytic part, this essay introduces geographic distance into the study of vertical integration and predicts the effect of distance on the likelihood of backward integration under two different assumptions about distance costs. Different assumptions give different predictions, by which the empirical results then imply which one of the assumptions of distance costs is consistent with the data. Second, in the empirical part, this essay contributes to the literature by focusing on the *causal* effects of key variables. Testing the causal effects of key variables presents an econometric challenge—reverse causation. The control rights associated with equity stakes allow the downstream producers who hold the equity stakes of suppliers to manipulate the suppliers' operation decisions and then change the characteristics studied in the hypotheses. To address this concern, three instrumental variables—one for each key variable—are utilized. Third, this essay studies the *size* of equity stakes, instead of zero/one dummies. The panel data fractional probit estimation method recently proposed by Papke and Wooldridge (2008) is implemented to address this issue.

Two comments could be made on chapter 2. First, the firm-level data from the Japanese auto parts industry fit the model built by Acemoglu, Aghion, Griffith, and Zilibotti (forthcoming) very well, which suggests that this model framework could be useful in the study of vertical integration in other contexts. Second, it will be a natural extension to study international trade's impacts on the size of such equity stakes, i.e.

whether the emergence of Chinese auto industry could make the Japanese auto industry be less integrated.

Chapter 3 shows that the presence of multinational retailers has a robust and positive effect on China's exports. This effect is *not* driven by the direct market access provided by multinational retailers. Rather, retailer presence (proximity to purchasing centers and the placement of stores in a city) is correlated with the city's export fixed effect. Combined with the strict exogeneity test results, the evidence that changes in retailer presence occur prior to changes in city export capability gives some reason to believe the effect could be causal.

Three comments could be made on chapter 3. First, this essay contributes to the literature by studying multinational retailers' impact on *host* countries' exports. Previously, Basker and Van (2007, 2008) have studied the impacts of multinational retailers on their home countries' imports. Second, our two-stage estimation provides us good evidence for the statement that presence of multinational retailers improves the general export capabilities of origin cities in which they are located. It is in line with Javorcik and Li (2007)'s findings of the retailer-induced productivity improvement for Romanian food suppliers; however, our result does not provide *direct* empirical evidence for *specific* economic mechanisms through which the presence of multinational retailers improves the general export capabilities of origins. Nevertheless, our trade-based method may prove useful for other valuations of linkage and capability effects caused by foreign investments.

Chapter 4 re-examines a question that is central to competition policy: when do participants in a market have the incentive to enter into contracts that exclude potential entrants? The answers to this question, first developed in an article by Aghion and Bolton (1987) focus on contracts that extract rents from a potential entrant or on contracts that exploit a collective action problem among buyers. Under the conventional assumption of a first-mover advantage by an incumbent, this essay first considers a simple benchmark market structure. In this "Chicago" benchmark, a seller and buyers contract with

perfect competition in all markets outside those of the contracting parties (i.e. perfect competition among upstream suppliers and among potential entrants facing a common, uncertain cost). There is no incentive for exclusionary long-term contracts.

Departures from the benchmark in each of three directions yield predictions of long-term exclusionary contracts. These include the two existing theories as well as a third: that a long-term contract at one stage of the supply chain may extract rents at another stage of the supply chain. Incumbency is less likely to yield a first-mover advantage in contracts with upstream suppliers than in downstream contracts; we therefore consider as well a model in which two firms bid simultaneously for the rights to upstream inputs, with bids for exclusive rights being one available strategy, and then compete in a downstream market. The equilibrium allocation of input rights, when the equilibrium in this game does exist, maximizes combined profits of all firms in the industry. With sufficient complementarity among upstream markets and substitutability downstream, the equilibrium allocates all inputs to a single firm—thus excluding the second firm from the market.

The Aghion-Bolton perspective on exclusion is to identify the incentives for anti-competitive exclusionary contracts in terms of transfers from agents outside the contracts. The synthesis that we offer isolates each incentive in the simplest departure from a benchmark in which privately optimal contracts are efficient. The approach is valuable beyond the static framework that we have explored in this paper.

Bibliography

Acemoglu, D., P. Aghion, R. Griffith, and F. Zilibotti. Forthcoming. Vertical Integration and Technology Theory and Evidence. *Journal of European Economic Association*.

Aghion, P. and P. Bolton (1987). Contracts as a barrier to entry. *The American Economic Review 77*, 388–401.

Basker, E. and P. H. Van (2007). Wal-Mart as Catalyst to U.S.–China Trade. (0710). Available at http://ideas.repec.org/p/umc/wpaper/0710.html.

Basker, E. and P. H. Van (2008). Imports "R" US: Retail Chains as Platforms for Developing-Country Imports. available at: http://ssrn.com/abstract=1123282.

Javorcik, B. S. and Y. Li (2007). Do the Biggest Aisles Serve Brighter Future? Implications of Global Retail Chains' Presence for Romania.

Papke, L. and J. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*, 121–133.

Appendix 1.1 Distance Model One

The model built by Acemoglu et al. (forthcoming) follows the convention in the property-rights literature. The following assumptions are made. Both the supplier and the downstream producer are risk neutral and can undertake technological investments to increase the output of the relationship. Investment decisions cannot be transferred between the supplier and the downstream customer because tacit knowledge is required in the investment of each party. Investments and the outputs of the relationship are not verifiable, and there is no financial constraint for either party.[87] The assumption of no financial constraints implies that, by applying ex ante transfers between the two parties, the first best scenario can be achieved.

The timing of the game is as follows:

1. The supplier and downstream producer choose an organization form (ownership structure), $z \in VIB, NI$ and associated transfers, $T_p(z)$ and $T_s(z)$, such that $T_p(z)$+$T_s(z)$=0. $T_i(z)$ refers to the ex ante transfer to party i conditional on the organization form z. Since there is no financial credit, $T_i(z)$ could be negative.

2. The customer and the supplier then simultaneously choose their investments, e_p and e_s. When they make such investment decisions, they know their payoffs $\{O_p^z$, $O_s^z\}$ ex post if they disagree on the division of revenues. In other words, both e_p and e_s are functions of the organization form (z) chosen in the previous period.

3. The supplier and the downstream producer bargain over the division of revenue according to the symmetric Nash bargaining solution given the organization form and its associated alternative payoffs $\{O_p^z$, $O_s^z\}$ outside the specific relationship. Output is realized and shared.

[87] Following the literature, the term investment is equivalent to the term effort.

The production technology of the relationship is assumed to be

$$F(x_s, e_p, e_s) = (pe_p + se_s + \frac{1}{d})\phi x_s + (pe_p + 1)(1 - \phi). \tag{.0.1}$$

$1/d$ in the first parentheses makes this model different from the original model in Acemoglu et al. (forthcoming). It captures the idea that distance dampens the value of parts in which no special effort is exerted. The first part of the function is the output generated by the supplier and the assembler when the part is supplied. In other words, it is positive only when x_s is equal to 1. $\phi \in (0, 1)$ corresponds to the share of the producer's output accounted for by the inputs provided by the supplier. The parameters p and s measure the marginal effects of efforts by the downstream producer and the supplier, respectively. e_p and e_s denote the corresponding effort levels.

It is also assumed that the basic input is supplied at no cost and the cost functions are quadratic in terms of efforts:

$$\Gamma_p = \frac{1}{2}e_p^2, \qquad \Gamma_s = \frac{1}{2}\phi e_s^2$$

It is important to notice that the supplier's effort costs are multiplied by ϕ, which ensures that costs are proportional to the scale of operation. No economies of scale are involved in the production of the supplier.

I focus on the comparative study between backward integration and non-integration. The model is solved backward.

Backward Integration

The organization form chosen determines players' outside options. Outside options are the payoffs in the event of ex post breakup. Under backward integration, the assembler has the residual rights of control and keeps all assets and inputs that have already been produced. It is also assumed that without the support of the supplier, the assembler

cannot achieve all the benefits associated with the supplier's effort. Some benefits of the effort, such as fine-tuning the machine, cannot be realized until the last moment of transaction. A parameter of λ captures the proportion of such loss. Therefore, the outside options under backward integration are

$$\begin{aligned} O_s^{VIB} &= 0 \\ O_p^{VIB} &= F(x_s = 1, e_p, (1-\lambda)e_s) \\ &= \phi[pe_p + se_s(1-\lambda) + \frac{1}{d}] + (1-\phi)(pe_p + 1). \end{aligned}$$

The quasi-rents in backward integration are

$$\begin{aligned} QR^{VIB} &= F(x_s = 1, e_p, e_s) - O_s^{VIB} - O_p^{VIB} \\ &= \phi\lambda s e_s, \end{aligned}$$

which is strictly greater than zero, ensuring the rationality of this specific relationship. Based on symmetric Nash bargaining, the payoffs of the supplier and assembler follow:

$$y_i^{VIB} = O_i^{VIB} + \frac{1}{2}(QR^{VIB}).$$

In equilibrium, the effort levels are determined by the Nash equilibrium of a game where each party chooses its effort level to maximize its own revenue, given the investment of the other party and the ownership structure. The pair of efforts are

$$\begin{aligned} e_p^{*VIB} &= \max_{e_p}[y_p^{VIB} - \Gamma_p(e_p)] \\ &= p, \\ e_s^{*VIB} &= \max_{e_s}[y_s^{VIB} - \Gamma_s(e_s)] \\ &= \frac{1}{2}s\lambda. \end{aligned}$$

Hence the total social surplus under the organization form of backward integration is

$$\begin{aligned} S^{*VIB} &= F(x_s = 1, e_p^{*VIB}, e_s^{*VIB}) - \Gamma_p(e*_p) - \Gamma_s(e*_s) \\ &= \frac{1}{2}\phi s^2\lambda + \frac{\phi}{d} + \frac{1}{2}p^2 + 1 - \phi - \frac{1}{8}\phi s^2\lambda^2, \end{aligned}$$

which is to be compared with the corresponding surplus in non-integration.

Non-integration

Under non-integration, each player is independent. The supplier controls the parts that he has produced; the producer controls the output other than the parts. Their outside options are

$$\begin{aligned} O_s^{VIB} &= (se_s + 1)\theta\phi, \\ O_p^{VIB} &= (1 - \phi)(pe_p + 1), \end{aligned}$$

where $\theta \in [0, 1)$ is an inverse measure of how much the supplier loses if she sells the input outside the specific relationship. It is measured in the empirical analysis by the number of customers that the supplier has.

The quasi-rents in non-integration then equal

$$\begin{aligned} QR^{NI} &= F(x_s = 1, e_p, e_s) - O_s^{NI} - O_p^{NI} \\ &= \phi pe_p + (1 - \theta)\phi se_s + \phi(\frac{1}{d} - \theta), \end{aligned}$$

which is strictly greater than zero as long as $\frac{1}{d} > \theta$. Based on symmetric Nash bargaining, the supplier's and assembler's payoffs are:

$$y_i^{NI} = O_i^{NI} + \frac{1}{2}(QR^{NI}).$$

In equilibrium, the effort levels are

$$\begin{aligned} e_p^{*NI} &= \max_{e_p}[y_p^{NI} - \Gamma_p(e_p)] \\ &= p(1 - \frac{\phi}{2}) \\ e_s^{*NI} &= \max_{e_s}[y_s^{NI} - \Gamma_s(e_s)] \\ &= \frac{s}{2}(\theta + 1) \end{aligned}$$

Then the total surplus under the organizational form of non-integration is

$$\begin{aligned} S^{*NI} &= F(x_s = 1, e_p^{*NI}, e_s^{*NI}) - \Gamma_p(e*_p) - \Gamma_s(e*_s) \\ &= -\frac{1}{8}\phi^2 p^2 + \frac{1}{4}\phi s^2\theta + \frac{3}{8}\phi s^2 + \frac{\phi}{d} + \frac{1}{2}p^2 + 1 - \phi - \frac{1}{8}\phi s^2\theta^2. \end{aligned}$$

Comparison

The difference between these two structures equals $\triangle = S^{*VIB} - S^{*NI}$. In order to find the effect of distance on the likelihood of vertical integration, the first derivative of the difference with respect to d is made, i.e. $\frac{\partial \triangle}{\partial d} = 0$. It is equal to zero.

Proposition 1. *Based on the assumption that distance has no effect on the marginal effect of players' efforts, distance has no effect on the likelihood of backward integration.*

Appendix 1.2 Distance Model Two

The production technology of the relationship is assumed to be

$$F(x_s, e_p, e_s) = (pe_p + \frac{se_s}{d} + 1)\phi x_s + (pe_p + 1)(1 - \phi), \tag{.0.1}$$

where $\frac{1}{d}$ under se_s in the first parenthesis makes this model different from the original one in Acemoglu et al. (forthcoming). It captures the idea that distance only dampens the effectiveness of the supplier's effort. All the other part of this function is same as Acemoglu et al. (2005)'s model. The first part of the function is the output generated by the supplier and the assembler when the part is supplied. In other words, it is positive only when x_s is equal to 1. $\phi \in (0, 1)$ corresponds to the share of the producer's output accounted for by the inputs provided by the supplier. The parameters p and s measure the marginal effects of efforts by the downstream producer and the supplier, respectively. e_p and e_s denote the corresponding effort levels.

It is also assumed that the basic input is supplied at no cost, and the cost functions are quadratic in terms of efforts:

$$\Gamma_p = \frac{1}{2}e_p^2, \qquad \Gamma_s = \frac{1}{2}\phi e_s^2.$$

It is important to notice that the supplier's effort costs are multiplied by ϕ, which ensures that costs are proportional to the scale of operation. No economies of scale is involved in the production of the supplier.

I focus on the comparative study between backward integration and non-integration. The model is solved backward.

Backward integration

The organization form chosen determines the outside option of each player. Outside options are the payoffs in the event of ex post breakup. Under backward integration, the assembler has the residual rights of control and keeps all assets and inputs that have already been produced. It is also assumed that without the support of the supplier, the assembler cannot achieve all the benefits associated with the supplier's effort. Some benefits of the effort, such as fine-tuning the machine, cannot be realized until the last moment of transaction. A parameter of λ captures the proportion of such loss. Therefore under backward integration, the outside options for the two parties are

$$\begin{aligned} O_s^{VIB} &= 0, \\ O_p^{VIB} &= F(x_s = 1, e_p, (1-\lambda)e_s) \\ &= \phi[pe_p + \frac{se_s(1-\lambda)}{d} + 1] + (1-\phi)(pe_p + 1). \end{aligned}$$

The quasi-rents in backward integration are

$$\begin{aligned} QR^{VIB} &= F(x_s = 1, e_p, e_s) - O_s^{VIB} - O_p^{VIB} \\ &= \frac{\phi\lambda se_s}{d}, \end{aligned}$$

which is strictly greater than zero ensuring the rationality of the specific relationship. Based on the symmetric Nash bargaining, the payoffs of the supplier and the assembler are:

$$\begin{aligned} y_p^{VIB} &= \frac{\phi se_s}{d} - \frac{\phi\lambda se_s}{2d} + pe_p + 1, \\ y_s^{VIB} &= \frac{\lambda\phi se_s}{2d}. \end{aligned}$$

In equilibrium, the effort levels are determined by the Nash equilibrium of a game where

each party chooses its effort level to maximize its own revenue, given the other party's investment and the ownership structure. The pair of efforts are

$$\begin{aligned} e_p^{*VIB} &= \max_{e_p}[y_p^{VIB} - \Gamma_p(e_p)] \\ &= p, \\ e_s^{*VIB} &= \max_{e_s}[y_s^{VIB} - \Gamma_s(e_s)] \\ &= \frac{\lambda s}{2d}. \end{aligned}$$

Hence the total social surplus of backward integration is

$$\begin{aligned} S^{*VIB} &= F(x_s = 1, e_p^{*VIB}, e_s^{*VIB}) - \Gamma_p(e*_p) - \Gamma_s(e*_s) \\ &= \frac{4\phi s^2\lambda + 4p^2d^2 + 8d^2 + \phi s^2\lambda^2}{8d^2}, \end{aligned}$$

which is to be compared with the corresponding surplus in non-integration at the first stage.

Non-integration

Under non-integration, each player is independent. The supplier control the parts that she has produced; the producer has the output other than the parts. Their outside options are

$$\begin{aligned} O_s^{VIB} &= (se_s + 1)\theta\phi \\ O_p^{VIB} &= (pe_p + 1)(1 - \phi), \end{aligned}$$

where $\theta \in [0, 1)$ is an inverse measure of how much the supplier loses if she sells the input outside the specific relationship. In the empirical part, it is measured by the number of customers that the supplier has.

The quasi-rents in non-integration equal

$$\begin{aligned} QR^{NI} &= F(x_s = 1, e_p, e_s) - O_s^{NI} - O_p^{NI} \\ &= pe_p + \phi(1-\theta) + \phi s e_s(\frac{1}{d} - \theta), \end{aligned}$$

which is strictly greater than zero as long as $\frac{1}{d} > \theta$. Based on the symmetric Nash bargaining, it implies that the payoffs of the supplier and the assembler are:

$$\begin{aligned} y_p^{NI} =& pe_p + 1 - \frac{1}{2}\phi p e_p - \frac{\phi}{2} + \frac{\phi s e_s}{2d} - \frac{\phi\theta s e_s}{2} - \frac{\phi\theta}{2} \\ y_s^{NI} =& \frac{\phi s e_s}{2}(\theta + \frac{1}{d}) + \frac{\phi p e_p}{2} + \frac{\phi}{2}(\theta + 1). \end{aligned}$$

In equilibrium, the downstream producer's and supplier's efforts are

$$\begin{aligned} e_p^{*NI} &= \max_{e_p}[y_p^{NI} - \Gamma_p(e_p)] \\ &= p(1 - \phi/2) \\ e_s^{*NI} &= \max_{e_s}[y_s^{NI} - \Gamma_s(e_s)] \\ &= \frac{s}{2}(\theta + 1/d) \end{aligned}$$

Hence the total social surplus of non-integration is

$$\begin{aligned} S^{*NI} &= F(x_s = 1, e_p^{*NI}, e_s^{*NI}) - \Gamma_p(e*_p) - \Gamma_s(e*_s) \\ &= \frac{-\phi^2 p^2 d^2 + 2\phi s^2 \theta d + 3\phi s^2 + 4p^2 d^2 + 8d^2 - \phi s^2 \theta^2 d^2}{8d^2}, \end{aligned}$$

which is compared with the corresponding surplus in non-integration.

Comparison

The difference of the total social surplus between these two organization forms equals $\triangle = S^{*VIB} - S^{*NI}$. In order to pin down the effect of distance on the likelihood of

backward integration, the first derivative of the difference with respect to d is made, $\frac{\partial \triangle}{\partial d} = \phi s^2 \frac{(\lambda-2)^2+\theta d-1}{4d^3}$. According to the assumption that $1/d > \theta$ and $\lambda \in [0,1]$, it can be seen that the first derivative is positive, which implies that as distance increases, backward integration becomes better. In other words, the likelihood of vertical integration increases in distance.

Proposition 2. *If distance has a negative effect only on the marginal effect of the supplier's effort, as distance increases, the likelihood of backward integration becomes higher.*

Appendix 1.3 Equity Stakes Classifications

The level of equity stake associated with effective control rights in the Japanese companies is not very high, and there are various classifications available in the literature. I list three classification schemes of which I am aware of here.

IMF IMF classifies international investments. A target firm is defined as the subsidiary of the shareholder when it is owned the majority equity stake, and the shareholder has control rights over the subsidiary. When the equity stake is between 10% and 50%, the target firm is the associate of the shareholder corporation; when the equity stake is lower than 10%, the equity is regarded as portfolio investment, and the shareholder becomes passive, which implies they have no voice in the management of the firm.

Freshfields Bruckhaus Deringer This classification is based on Freshfields Bruckhaus Deringer's report in 2003.[88] It classifies the levels of equity stakes into the following categories: 1%, 3%, 10%, over one-third, 50%, over 50%, and two-thirds. Two levels are particularly important: over one-third and two-thirds. Over one-third equity of all voting rights is the minimal equity share that endows the shareholder with veto rights on the passage of resolutions of a shareholders' meeting concerning a set of specific important issues, such as purchasing the company's own shares, sale of the entire business or an important part of the business of the corporation, demerger by transferring the businesses to another company, etc. Ahmadjian and Lincoln (2001) states that "Any shareholder who owns two-thirds of all voting rights has the absolute right to pass the resolutions of a shareholders' meeting concerning" the set of matters mentioned above.(p. 699) Therefore, the over one-third equity share level actually offers de facto control because it prevents a two-thirds majority. Let

[88]Freshfields Bruckhaus Deringer is an international law company. Its website is `http://www.freshfields.com/en.asp`. The 2006 edition is now available online. It reflects recent reforms in the Japanese company law, Freshfields (2006). However according to the sample period that I consider, I think the changed edition should not be considered. The 2003 edition is upon request.

me restate the statement made by Fukao (1995), an expert on Japanese corporate governance: "shareholders' meetings can only veto on the matters stipulated by the corporate law or the articles of incorporation. However, by amending the articles of incorporation with two-third majority, a shareholders' meeting can vote on other matters, including ordinary business."(p. 99) Therefore, it is clear that the over one-third equity level prevents the absolute right associated with two-third equity shares. It provides the shareholder effective rights to prevent the target firm from working to the shareholders' disadvantage. The one-third equity share level is legally critical and frequently observed in the Japanese auto industry. The following are a few examples among assemblers.

- Toyota increased its equity shares in Daihatsu from 16.8% to 33.4% in 1995,[89] and Toyota also increased its investment in Hino Motors to 33.4% in March 2000; [90]
- Ford owns Mazada with 33.4% since 1996;
- Daimler Chrysler's stakes in Mitsubishi Motors is 34% since March 2000;
- Renault holds 36.8% equity shares in Nissan.

20% Ahmadjian and Oxley (2006) document that 20% is another critical equity stake in Japanese auto companies. "If an assembler owns 20% stakes of a supplier, the supplier attains Kanren gaisha status, as an important sign of affiliation. Many Japanese firms have special departments to assisting these affiliated companies, for example, in developing strategy and acquiring technology."(p. 218)

[89]In September 1998 Toyota further increased its investment in Daihatsu Motors to 51.2% and made Daihatsu its consolidated subsidiary.

[90]In March 2001, Toyota further increased its investment in Hino Motors to 36.6% and subsequently to 50.1% on August 31st, 2001.

Appendix 2.1 Chinese Data Sources

Table A-1: Gravity Variables Data Sources

Variable	Source
Countries' GDP and populations	World Bank Development Indicators
Chinese provinces' GDP and populations	China Data Online
The longitudes and latitudes of country capitals	CEPII
The longitudes and latitudes of province capitals	Map of World website
Chinese city land areas	China Data Online
Chinese ports	Lloyd's ports of the world (1995)

Appendix 2.2 Collinearity Problems Associated with Three Sets of Dummies

As we use specification 3.4.3, we soon realize that a large quantity of city–year and country–year dummies are dropped automatically by STATA, and the estimates of the key variable differ just when we run the same regression repeatedly. We implement specification 3.4.3 by using the code

`xtreg` X_{odt} `ot2-ot300 dt2-dt500, i(od).`

X_{odt} is the sum of exports from city o to destination country d in year t. ot2-ot300 include all city–year dummies except for the first one. dt2-dt500 include all country–year dummies other than dt1. The constant term is included in the specification. Dyadic fixed effects are controlled by "i(od)." Without adding any other variables which includes key variables, the specification will drop $(N_o + N_d + N_t - 3)$ dummies.

These origin–year and destination–year dummies are dropped because there are three sets of perfect collinearity problems imbedded in the specification. In addition, these three sets of perfect collinearity problems also explain why we cannot solve this problem by arbitrarily dropping $(N_o+N_d+N_t-3)$ dummies. Collinearity occurs between city–year and dyadic fixed effects, between country–year and dyadic fixed effects, and between the city–year and the country–year fixed effects. The collinearity between the city–year and dyadic fixed effects results from the fact that dyadic and city–year fixed dummies share the *same* number of years. After demeaning on each dyad, the sum of dummies over all years of the same city is a zero vector except for the city whose one city–year dummy has been dropped. In other words, for each city, one of its city–year dummies must be dropped in order to deal with the perfect collinearity problem resulting from city–year and dyadic fixed effects. Thus, when we use the code list above to implement specification 3.4.3, (N_o-1) city–year dummies are to be dropped. Similarly, for each country, one of its country–year dummies has to be dropped. Therefore, (N_d-1) country–year dummies are

dropped, when we use the previous code. The perfect collinearity between the city–year and the country–year fixed effects also results from cities and countries having the same number of years. For each year, the vector determined by adding up all cities' city–year dummies in that year is equal to the the vector arrived at by adding up all countries' country–year dummies in the same year. After inputting the previous code, $(N_t - 1)$ dummies, one for each year other than the first year, are dropped. In order to avoid the collinearity problem resulting from the city–year and the country–year fixed effects, one item (either a city or country) has to be dropped in *each* year.

The following simple example demonstrates these three sets of perfect collinearity problems. Suppose there are 2 countries, 2 years, and 2 cities.

$Prov$	$Country$	$Year$	myr	$m1$	$m2$	$m3$	$m4$	xyr	$x1$	$x2$	$x3$	$x4$
BJ	USA	96	1	1	0	0	0	1	1	0	0	0
BJ	USA	97	2	0	1	0	0	2	0	1	0	0
TJ	USA	96	1	1	0	0	0	3	0	0	1	0
TJ	USA	97	2	0	1	0	0	4	0	0	0	1
BJ	ARG	96	3	0	0	1	0	1	1	0	0	0
BJ	ARG	97	4	0	0	0	1	2	0	1	0	0
TJ	ARG	96	3	0	0	1	0	3	0	0	1	0
TJ	ARG	97	4	0	0	0	1	4	0	0	0	1

The perfect collinearity problem between city–year and country–year dummies: the vector achieved by adding up the 2 cities' city–year dummies in 96 is equal to the vector that results from adding up the 2 countries' country–year dummies in 96. Both vectors are equal to $(1, 0, 1, 0, 1, 0, 1, 0)^T$. For every city or country, there must be a dummy for 96 and a dummy for 97, respectively. Therefore, it is clear that by using city–year and country–year fixed effects simultaneously, we incur a perfect collinearity problem in each year. The previous code is equivalent to having $m2$-$m4$ and $x2$-$x4$ in this simple case.

Based on the analysis, we can see that there is still a perfect collinearity problem in 97, i.e. $m2{+}m4{=}x2{+}x4$.

Now we can turn to the perfect collinearity between the dyadic and city–year dummies. Following the previous matrix, we achieve the following matrix after demeaning the variables on each dyad.

$Prov$	$Country$	$Year$	myr	$m1$	$m2$	$m3$	$m4$	xyr	$x1$	$x2$	$x3$	$x4$
BJ	USA	96	1	$\frac{1}{2}$	$\frac{-1}{2}$	0	0	1	$\frac{1}{2}$	$\frac{-1}{2}$	0	0
BJ	USA	97	2	$\frac{-1}{2}$	$\frac{1}{2}$	0	0	2	$\frac{-1}{2}$	$\frac{1}{2}$	0	0
TJ	USA	96	1	$\frac{1}{2}$	$\frac{-1}{2}$	0	0	3	0	0	$\frac{1}{2}$	$\frac{-1}{2}$
TJ	USA	97	2	$\frac{-1}{2}$	$\frac{1}{2}$	0	0	4	0	0	$\frac{-1}{2}$	$\frac{1}{2}$
BJ	ARG	96	3	0	0	$\frac{1}{2}$	$\frac{-1}{2}$	1	$\frac{1}{2}$	$\frac{-1}{2}$	0	0
BJ	ARG	97	4	0	0	$\frac{-1}{2}$	$\frac{1}{2}$	2	$\frac{-1}{2}$	$\frac{1}{2}$	0	0
TJ	ARG	96	3	0	0	$\frac{1}{2}$	$\frac{-1}{2}$	3	0	0	$\frac{1}{2}$	$\frac{-1}{2}$
TJ	ARG	97	4	0	0	$\frac{-1}{2}$	$\frac{1}{2}$	4	0	0	$\frac{-1}{2}$	$\frac{1}{2}$

From this matrix, we can find that both $x1{+}x2$ and $x3{+}x4$ are equal to $(0,0,0,0,0,0,0,0)^T$. In other words, after demeaning, there is a perfect collinearity problem across the time-varying dummies for each city. The traditional method of dropping only the first city–year dummy cannot solve the collinearity problem for all the other cities. In order to deal with this problem, we must drop one city–*year* dummy for each city.

A similar argument also applies to the country-year dummies. In addition, we notice that the previous perfect collinearity problem between the city–year and country–year dummies still exists after demeaning the variables over dyads.

In summary, if we put all city–year and country–year dummies in the specification and use dyadic fixed effects, without adding any other key variables, STATA automatically drops $N_o + N_d + N_t - 1$ dummies *in total*. In order to solve the previous collinearity problems, the method for manually dropping the dummies is to:

- Drop one city–*year* dummy for each city;
- Drop one country–*year* dummy for each country;
- Drop one *city*–year *or country*–year dummy in each year.

The total numbers of dummies dropped or controlled have been tested by simulations.

Appendix 2.3 Functional Forms for Linkage Variables

Global procurement centers scour cities in their regions for product ideas. Search and communication costs impede the transfer of lists from city o to its nearest global procurement center. With s_o items available for each city, the proportion of the list that is successfully transmitted to its nearest global procurement centers decreases in the distance D_o from city o to the GPC with the standard functional form for distance costs found in gravity equations,[91] which is $\frac{s_o}{D_o}$. Suppose the probability for each item being chosen is $\pi \in [0, 1]$ and the number of stores of retailer r in country d is n_{dr}, then the total number of items produced in city o and expected to be picked up in country d equals $s_o \pi \sum_r \frac{n_{dr}}{D_{or}}$.

Stores also bring up product lists to their stores overseas. As before, the probability for each item being chosen by a store is assumed to be π, and n_{dr} stores of retailer r are assumed to be located in country d. Then $s_o \pi n_{dr}$ items proposed by a single store are expected to be picked up in country d. If there are n_{or} stores in city o and the product lists are non-overlapping, the total expected number of items produced in city o that are picked up by retailer r' stores in country d is $s_o \pi n_{dr} n_{or}$. Summing over the retailers, we could have $s_o \pi \sum_r n_{dr} n_{or}$.

There are three issues associated with the two measures discussed above. First, the assumption of non-overlapping lists is unrealistic. In reality there is probably some redundancy. Second, some products would be chosen in the absence of any stores. We incorporate these two ideas by adding 1 and raising the sum to a power that is presumably less than one to reflect the kind of diminishing returns that would arise from redundant lists. The numbers of products linked by global procurement centers and stores are therefore $(1 + s_o \pi \sum_r n_{dr} n_{or})^{\zeta}$ and $(1 + s_o \pi \sum_r n_{dr} n_{or})^{\eta}$ respectively. Third, when we apply these two measures to panel data, s_o enters time-varying city fixed effects and π is co-linear with the constant term. Then we use $ln(1 + \sum_r \frac{n_{dr}}{D_{or}})$ and $ln(1 + \sum_r n_{dr} n_{or})$ to

[91] Refer to Disdier and Head (2008)

measure the linkages generated by multinational retailers.

Appendix 3.1 Dual Exclusionary Contract Strategy

Suppose that we have ruled out Aghion Bolton I by assuming competitive entrants; ruled out Aghion-Bolton II by assuming a single downstream buyer; and ruled out the third theory by assuming a single upstream supplier. Then, in the context of a conventional private good, is there any incentive for exclusionary contracts? Conventional theory would suggest that in the case of a conventional private good, once a monopoly is in place at one level of the supply chain ($n = 1$) the monopolist has no incentive to enter into contracts that would establish additional barriers to entry at another stage of the supply chain.

We show that under these assumptions, there is in fact an incentive for exclusionary contracts. Vertical externalities along the supply chain, between the agents upstream and downstream from the incumbent, can lead to inefficient exclusion — via a vertical analog to the horizontal externalities in Aghion-Bolton II. We assume one primary supplier, one downstream buyer and perfectly competitive entry at a cost c_e. If no long term contracts have been struck, then after the cost of entry is realized the entrants compete with the incumbent in making offers to the suppliers and to the buyers. The long term contract with the seller calls for the exclusive supply of the input to the incumbent at a particular price and contains a liquidated damage or penalty that the supplier will pay if it breaches the contract by supplying to the entrants. (One option is complete exclusion, with an infinite liquidated damage.) The buyer needs at most one unit of the product and similarly may enter a long term contract with a liquidated damage if she decides not to purchase from the incumbent. In short, the market participants play the following game: (1) the incumbent chooses whether to make long-term contract offers to the seller, to the buyer, or both. Each contract specifies a price and a liquidated damage, or breach penalty. Throughout, we assume that the incumbent is committed to any long-term contract offer without the possibility of renegotiation. (2) the costs of the entrants are realized; (3) where the incumbent has not signed a contract with at least one of the

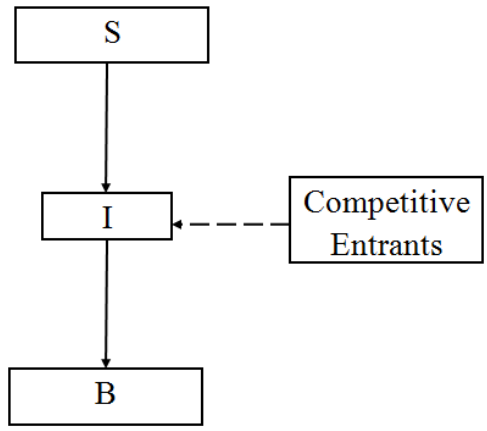


Figure A-1: Our Model Structure

agents, S or B, then ex post the incumbent makes a take-it-or-leave-it offer to this agent or agents. S and B can avail themselves of the competitive market as an alternative to the combination of the long term contracts and incumbent's ex post offer(s). Breaking a long term contract will, of course, involve payment of any breach penalties. (4) buyers choose between the offers, or may choose to retain the long term contract, if one was signed, and transactions take place. The structure is depicted in Figure A-1.

We define the exogenous variables in this model as follows. The cost of production upstream is assumed to be 0. The incumbent's cost of production (adding value at the intermediate stage) is given by c_I. The competitive entrants' cost of production is c_E. The cost c_E is random, with a continuous density $f(\cdot)$ and distribution $F(\cdot)$ on the interval $[0, \overline{c}]$. The buyer's value of the final product is v. The endogenous variables are the long-term supply contract (if it is struck), denoted by (w, d_S) where w is the price for the input and d_S is the liquidated damage; the long-term buyer contract, (p, d_B). The wholesale and final price in the event that a contract is not struck are denoted by w_N and p_N respectively.

The incumbent decides whether to offer a long-term contract to the supplier only, to the buyer only, or to both; and if contracts are offered the supplier and buyer decide whether to accept. Then nature draws the cost c_E from $F(\cdot)$. Ex post, the competitive entrants stand ready to offer the product at a price of c_E.

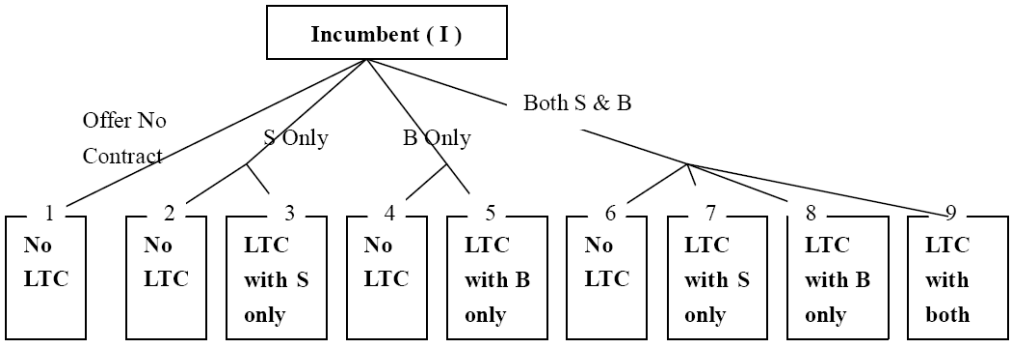


Figure A-2: Game Tree Summary

Figure A-2 provides a summary of the first stage of the game, the offer and acceptance decisions of the long term contract (LTC). We consider the sub-games that follow the nodes in this summary, with the aim of associating with each node a set of payoffs to the three agents, derived from the subgame, for various realizations of c_E. We will subsequently integrate over the possible realizations of c_E. This provides payoffs for the nodes at Figure A-2 and allows us to solve the game, characterizing when the incentive for long-term contract offers exists in this model.[92]

Realized Payoffs following No Long Term Contract

Consider first the nodes where no long-term contract has been entered (nodes 1, 2, 4, and 6). In this event, if $c_E > v$ then the incumbent is unaffected by the entrant and offers prices $w = 0$ and $p = v$, extracting the entire surplus, $v - c_I$. If $c_E < c_I$ then the competitive entrants supply the second input. The incumbent's payoff is 0, and we assume that the upstream supplier and the buyer share the surplus from the transaction, $v - c_E$, in shares λ and $(1 - \lambda)$. To stay within a non-cooperative game, we interpret these shares as the probabilities with which the upstream supplier or downstream buyer "discovers" the competitive entry first and makes a take-it-or-leave-it offer to the other party. In the intermediate range, for $c_E \in [c_I, v]$ the incumbent supplies the second input,

[92]For example, the payoffs to the buyer from the subgames at nodes 7 versus 9 will enter the buyer's participation constraint in the incumbent's design of an optimal pair of long term contracts.

Table A-2: Payoffs Following No Long-term Contract

Range of c_E:	$c_E < c_I$	$c_I \leq c_E \leq v$	$v < c_E$
π_I^{00}	0	$c_E - c_I$	$v - c_I$
π_S^{00}	$\lambda(v - c_E)$	$\lambda(v - c_E)$	0
π_B^{00}	$(1-\lambda)(v - c_E)$	$(1-\lambda)(v - c_E)$	0

but its price is constrained by the competitive entrants' price, c_E. Here the surplus from the transactions, which is $v - c_E$, accrues to the upstream supplier in the amount $\lambda(v - c_E)$, to the buyer in the amount $(1-\lambda)(v - c_E)$ and to the incumbent, $c_E - c_I$. Table A-2 summarizes the payoffs after the "no-contract nodes", conditional upon a realization of c_E, where the superscript 00 denotes that neither the supplier nor the buyer has entered the long-term contract.[93]

Realized Payoffs following Two Long Term Contracts

If a long-term contract with both parties is signed ex ante (node 9), then ex post the buyer and the supplier would, as a pair, benefit from breaching their long-term contracts providing that the entrants' price for the second input (i.e. the entrants' cost) is low enough to allow them to cover their combined net opportunity cost of leaving the long-term contract. Thus, breach of the long-term contracts occurs if $v - c_E - d_s - d_B > (v - p) + w$, i.e.

$$c_E < p - w - d_S - d_B \tag{3}$$

Note that the opportunity cost of leaving the contracts in (3) includes the breach penalties.[94]

The total payoff to the upstream supplier, after both long-term contracts are entered

[93] The superscript 11 will denote payoffs following two long term contracts; 01 payoffs following a contract with the buyer only and 10 payoffs following a contract with the seller only.

[94] We continue to assume that the gains from breach, which are $p - w - d_S - d_B - c_E$, are split between the supplier and the buyer, again in shares λ and $(1-\lambda)$. We interpret these shares as the probabilities that the supplier, or buyer respectively, discovers the competitive entrants first and sets the terms for a transaction that would leave the other party indifferent between breaching and not. For example, if the supplier discovers the entrants *first*, she will breach if (3) holds and then offer the buyer the final product at a price that extracts the entire buyer's gain from breaching.

Table A-3: Payoffs Following Dual Long-term Contracts

Range of c_E:	$c_E < p - w - d_S - d_B$	$c_E \geq p - w - d_S - d_B$
	breach	no breach
$\pi_I^{11}(p, d_B; w, d_S)$	$d_s + d_B$	$p - w - c_I$
π_S^{11}	$w + \lambda(p - w - d_S - d_B - c_E)$	w
π_B^{11}	$v - p + (1 - \lambda)(p - w - d_S - d_B - c_E)$	$v - p$

into and after the realization of a low value for c_E (satisfying (3)), is the payoff provided by the long-term contract plus the share of the gains from breach: $w + \lambda(p - w - d_S - d_B - c_E)$. Similarly, the buyer's payoff in this circumstance is $v - p + (1 - \lambda)(p - w - d_S - d_B - c_E)$.

Realized Payoffs following Long Term Contract with Buyer Only

Next we examine the payoffs when only the buyer has entered a long-term contract (node 5). Suppose that the buyer's long-term contract (p, d_B) with $p < v$, is in place. Ex post, the parties observe c_E, the incumbent offers an input price $\widehat{w}$ to the supplier, and the supplier and buyer decide whether to contract with the competitive entrants instead of having the buyer continue the long term contract. In deriving the payoffs for this subgame as a function of contract parameters, we must divide the contract parameters into two cases, depending on whether the incumbent's payoff is higher under continuance or breach of the long term contract: (A) $p - d_B \leq c_I$ and (B) $p - d_B > c_I$. In case A, the incumbent hopes that c_E is low enough that breach occurs, since the profits from breach, d_B, exceed $p - c_I$, the profits from continuation of the contract even if an input price $\widehat{w} = 0$ is offered by the incumbent and accepted. Of course in case A, a positive $\widehat{w}$ would never be offered.

Breach occurs in case A if the cost of obtaining the intermediate product from the entrants, c_E, plus the breach penalty, d_B, is less than p. The buyer and the supplier split the gains from breach, $(p - c_E - d_B)$ in proportions λ and $(1 - \lambda)$. In case A, the states (realizations of c_E) can be further divided into three events: states of breach,

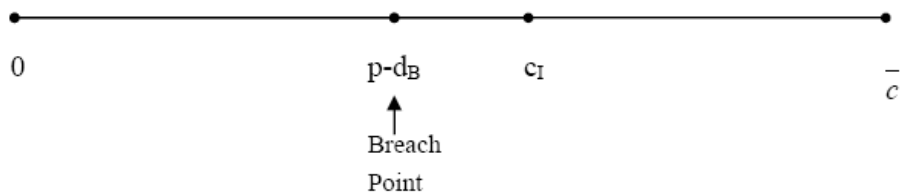


Figure A-3: Single Buyer Long-term Contract

$c_E < (p - d_B)$; inefficient continuation of the long term contract, $(p - d_B) \leq c_E < c_I$; and efficient continuation of the long term contract, $c_I \leq c_E$. This is depicted in Figure A-3. The following lemma shows that if the incumbent is restricted to offering a contract to the buyer only, it will satisfy $p - c_I < d_B$ so that inefficient continuation is possible (and a contract satisfying case B is in fact never optimal):

Lemma: If a long term contract (p, d_B) is offered to the buyer only, then it must satisfy $p - c_I < d_B$.[95]

The property that an inefficient contract would be struck with the buyer (under the supposition that a long term contract with only the buyer is possible), so as to extract rent from a party outside the contract, is familiar from Aghion and Bolton. It can be shown that the optimal contract extracts profit from the supplier with the buyer-incumbent pair essentially acting as a monopsonist against the supplier, following Aghion-Bolton.

As a result of the lemma, the payoffs following a contract with only the buyer are the payoffs in case A, $p - c_E \leq d_B$. These are provided in Table A-4.

Table A-4: Payoffs Following a Single Buyer Long-term Contract

Range of c_E:	$c_E < p - d_B$	$c_E \geq p - d_B$
Remarks:	breach	no breach
π_I^{01}	d_B	$p - c_I$
π_S^{01}	$\lambda(p - c_E - d_B)$	0
π_B^{01}	$(v - p) + (1 - \lambda)(p - c_E - d_B)$	$v - p$

[95] Proofs are contained in the end of this appendix. It is similar to the proof in section 4.3.

Realized Payoffs following Long Term Contract with Supplier Only

Using a parallel argument (or simply relying on the inherent symmetry of the buyer and input supplier in our model), the payoffs to the three players following a long term contract (w, d_S) with the input supplier only are those given in Table A-5.

Table A-5: Payoffs Following a Single Seller Long-term Contract

Range of c_E:	$c_E < v - w - d_s$	$c_E \geq v - w - d_s$
Remarks:	breach	no breach
π_I^{10}	d_s	$v - c_I - w$
π_S^{10}	$w + \lambda(v - w - d_s - c_E)$	w
π_B^{10}	$(1 - \lambda)(v - w - d_s - c_E)$	0

Optimal Long-term Contract Strategy

Finally, we analyze the incumbent's expected payoffs at the first stage of the game summarized in Figure A-2, taking into account the reaction of the buyer and seller to any contract offers. The incumbent has four broad strategies: offering no contract; offering long-term contract to the buyer only; offering long-term contract the supplier only; offering long-term contracts to both the supplier and buyer simultaneously; and of course within these broad strategy classes the incumbent must choose the optimal parameters to any contract offered, ensuring that the participation constraints are met. The payoffs to each strategy, as a function of contract parameters p, w, d_B, d_S are obtained by integrating the payoffs in Tables A-2 through A-5 with respect to the random variable c_E. These payoffs are provided in the proof section. The proposition below, proved in the proof section, provides our main result.

Proposition: *Offering a long term contract to each of the input supplier and buyer is always an optimal strategy for the incumbent.*

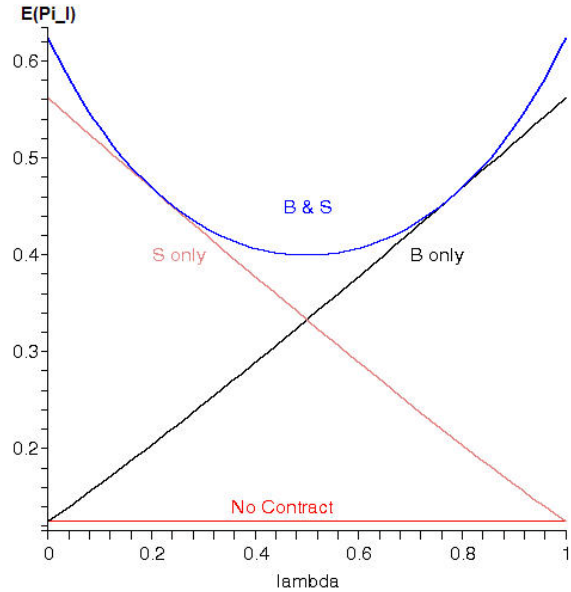


Figure A-4: Strategy Payoff Comparison

The profit from any strategy depends upon λ, the relative bargaining power of the buyer and the seller in negotiating an agreement to breach any contracts with the incumbent in order to contract with a low-cost entrant. Figure A-4 below provides the profit function for the incumbent, from offering an optimal contract with each of the parties, B and S, and dual contracts.

A long term contract with either party, dominates the no-contracting strategy, demonstrating the basic result: that the vertical externality alone generates an incentive for exclusionary contracts. The diagram reveals that the greatest profit comes from contracting with the party that has the lesser bargaining power, consistent with the basic idea that the role of long term contracts is to extract rent from the party excluded from the contract. The party with the greater bargaining power has the greater rent exposed to extraction through contractual externalities.

The optimal contract with the buyer satisfies $p - c_I < d_B$. This has two implications. First, the incumbent achieves greater profit when its contract is breached than when it is not breaches, as in the Aghion-Bolton model. Second, because the breach condition is $c_e + d_B < p$, this implies that there is a range of realizations of c_E, $c_E \in (p - d_B, c_I)$

where it would be efficient socially to breach, but breach does not occur.[96] This is the sense in which the long term contract acts as a barrier to efficient entry. Similarly for a long term contract with the supplier. Note that it pays either the buyer or supplier to establish the contract with the incumbent even though the party has a completely secure monopoly at its stage of the vertical supply chain. Normally, we think of a monopoly at one stage of a supply chain having no incentive to establish barriers to entry at another stage of the supply chain.[97]

Proofs for A Dual Contract Strategy

Payoffs from sub-games in Figure A-2

Following no long-term contract (nodes 1,2,4 and 6):

$\pi_I^{00} = \int_{c_I}^{v}(c - c_I)dF(c) + [1 - F(v)](v - c_I)$

$\pi_S^{00} = \lambda \int_0^v (v - c)dF(c)$

$\pi_B^{00} = (1 - \lambda) \int_0^v (v - c)dF(c)$

Following long-term contracts (w, d_S) and (p, d_B) to both agents (and omitting arguments of the functions):

$\pi_I^{11} = \int_0^{p-w-d_S-d_B}(d_S + d_B)dF(c) + [1 - F(p - w - d_S - d_B)](p - w - c_I)$

$\pi_S^{11} = \int_0^{p-w-d_S-d_B}[w + \lambda(p - w - d_S - d_B - c_E)]dF(c) + [1 - F(p - w - d_S - d_B)]w$

$\pi_B^{11} = \int_0^{p-w-d_S-d_B}[v - p + (1 - \lambda)(p - w - d_S - d_B - c_E)]dF(c) + [1 - F(p - w - d_S - d_B)](v - p)$

Following a long-term contract (p, d_B) with only the buyer:

[96] Recall that we are not allowing for contract renegotiation. The contracting parties could gain from renegotiating whenever $c_E \in (p - d_B, c_I)$. Renegotiation would not eliminate the barrier to entry aspect of long term contracts in a more plausible model with asymmetric information or post-contract specific investment.

[97] The one exception to the strict dominance of the dual contracting strategy are the tangency points in Figure A-4. We can understand this effect by considering the optimal level of liquidated damages in the dual contracting strategy. We find in simulations that when the bargaining power is extremely different between the two agents, a *negative* liquidated damage is optimal for one contract under the dual contracting strategy. At the values of λ where the optimal d_B is exactly zero, there is effectively no contract with the buyer and the dual contract strategy is equivalent to contracting with the supplier alone.

$\pi_I^{01}(p, d_B) = \int_0^{p-d_B} d_B dF(c) + (p - c_I)[1 - F(p - d_B)]$

$\pi_S^{01}(p, d_B) = \lambda \int_0^{p-d_B} (p - d_B - c) dF(c)$

$\pi_B^{01}(p, d_B) = (v - p) + (1 - \lambda) \int_0^{p-d_B} (p - d_B - c)] dF(c)$

Following a long-term contract (w, d_S) with only the upstream supplier:

$\pi_I^{10}(w, d_S) = \int_0^{v-w-d_S} d_S dF(c) + (v - w - c_I)[1 - F(v - w - d_S)]$

$\pi_S^{10}(w, d_S) = w + \lambda \int_0^{v-w-d_S} (v - w - d_S - c) dF(c)$

$\pi_B^{10}(w, d_S) = (1 - \lambda) \int_0^{v-w-d_S} (v - w - d_S - c) dF(c)$

Proof of Lemma

Lemma: *If a long term contract (p, d_B) is offered to the buyer only, then it must satisfy $p - c_I < d_B$.*

Proof: Note the following:

1. The addition of x dollars to each of p and d_B is equivalent to an ex ante transfer from the buyer to the incumbent. The implication is that the optimal contract will maximize the sum of expected profit of the incumbent and the expected surplus of the buyer.

2. The entrants make zero realized profits. Therefore the combined realized profits of the incumbent and the buyer can be expressed as the realized <u>total</u> surplus in the market minus the profits earned by the supplier.

3. The total surplus in the market is just v minus the cost of production, which is c_E in the states where the contract breached and c_I in the states where there is no breach. The expected total surplus is therefore

$$ES = \int_0^{(p-d_B)} (v - c_E) dF(c_E) + (v - c_I)[1 - F(p - d_B)]$$

Subtracting the payoff to the supplier yields the total expected profits to the incumbent and the buyer (the objective function maximized by an optimal buyer-only contract), in

case A:

$$\Pi \equiv E\pi_I^{01} + E\pi_B^{01} = \int_0^{(p-d_B)} (v - c_E)dF(c_E) + (v - c_I)[1 - F(p - d_B)]$$
$$-\lambda \int_0^{(p-d_B)} (p - c_E - d_B)dF(c_E)$$

Note that this objective, Π, can be written as a function of $\Delta \equiv (p - d_B)$:

$$\Pi(\Delta) = \int_0^{\Delta} (v - c_E)dF(c_E) + (v - c_I)[1 - F(\Delta)] - \lambda \int_0^{\Delta} (\Delta - c_E)dF(c_E)$$

We show $\partial\Pi/\partial\Delta < 0$ at $\Delta = c_I$ to show that a decrease in Δ below c_I is optimal, thus proving the lemma for case A:

$$\frac{\partial \Pi}{\partial \Delta}_{\rfloor \Delta = c_I} = -\lambda F(c_I) < 0$$

The marginal reduction of $\Delta \equiv (p - d_B)$ below c_E opens up an "inefficient continuation" set, but (through an envelope theorem effect) the first-order impact of this loss in efficiency is zero. The only first order effect is the extraction of λ additional dollars from the supplier in each of the breach states. The supplier's bears a share λ of the drop in gains from breach due to an increase in p or a decrease in d_B.

This shows that $\Delta = c_I$ is dominated by $\Delta < c_I$. It remains to consider the possibility of case B, $\Delta > c_I$ Since the payoffs are different in case B than in case A, the above proof does not extend directly. We cannot a priori assume $\widehat{w} = 0$ in case B since the incumbent may, ex post, want to pay the supplier a positive price $\widehat{w}$ so as to deter breach and allow the incumbent to realize profit $p - c_I - \widehat{w}$. Given an offer $\widehat{w}$, breach occurs in case B if the net cost to the supplier-buyer pair is less from the entrants than it is from the incumbent: $c_E + d_B < p - \widehat{w} \Longleftrightarrow c_E < p - \widehat{w} - d_B$. It is straightforward to show that under the supposition of case B it pays the incumbent to offer a supply price $\widehat{w}$, ($\widehat{w} = p - c_E - d_B$) that deters breach if and only if c_E exceeds c_I. Thus all contracts

with $\Delta > c_I$ achieve first-best efficiency, i.e. maximize S. The supplier earns $\widehat{w}$ plus a share λ of the gains from breach $p - d_B - \widehat{w} - c_E$. It is straightforward to show that the expected profit to the supplier is strictly reduced in a change in contract from $\Delta > c_I$ to $\Delta = c_I$. Since both of these achieve the same total surplus, this means that the sum of the payoff to the incumbent and buyer are higher under $\Delta = c_I$ than under $\Delta > c_I$. Thus any contract satisfying (B) is dominated. ■

Proof of Proposition

Proposition: *Offering a long term contract to each of the input supplier and buyer is always an optimal strategy for the incumbent.*

Dominance of dual contracting strategy over single contract:

We first prove the weak dominance of a dual contracting strategy over the strategy of contracting with only the supplier by (a) finding a dual-contracting strategy that is pay-off equivalent to the optimal supplier-only strategy, and then showing that (b) this strategy is dominated by other dual-contracting strategies, except at a particular value of λ where it is optimal. Parallel arguments show that the dual-contracting strategy weakly dominates a buyer-only strategy, and that a supplier-only strategy or a buyer-only strategy weakly dominates no contracting.

The optimal supplier-only contract solves

$$\max_{w,d_S} E\pi_I^{10}(w,d_s) \text{ s.t. } E\pi_S^{10}(w,d_s) \geq E\pi_S^{00}.$$

Using the payoffs provided in Appendix 1, we can write the Lagrangian for this problem as:

$$\begin{aligned} L_I^{10} &= \int_0^{v-w-d_S} d_s dF(c_E) + [1 - F(v-w-d_s)](v-w-c_I) \\ &\quad + \mu[w + \int_0^{v-w-d_S} \lambda(v-w-d_s-c_E)dF(c_E) - \int_0^{v} \lambda(v-c_E)dF(c_E)], \end{aligned}$$

where μ is the shadow price on the participation constraint. At the optimal contact

$$\frac{\partial L_I^{10}}{\partial d_S^*} = f(v - w^* - d_S^*)(v - w^* - d_S^* - c_I) + (1 - \lambda\mu)F(v - w^* - d_s^*) = 0.$$

Note that in this problem – and in any of our optimal contracting problems – a lump sum transfer can be implemented between the incumbent offering the contract and the party accepting the contract: an decrease in d_S by 1 dollar and a increase in w by 1 dollar transfers 1 dollar to the supplier in all states without affecting the incentive to breach. This transferable-utility aspect of the optimal contract means that the shadow price on the participation constraint equals 1. Substituting $\mu = 1$ yields

$$f(v - w^* - d_s^*)(v - w^* - d_s^* - c_I) + (1 - \lambda)F(v - w^* - d_s^*) = 0, \tag{4}$$

where the subscript * denotes the optimal long term contract with the supplier only. This equation describes the marginal trade-off with an increase in d_s. A marginal increase in d_S results in an efficiency cost as is the first part of this equation since $v - w^* - d_S^* < c_I$; but also it results in the extraction of $(1 - \lambda)F(v - w^* - d_s^*)$ in rents from the party outside the contract, i.e., the buyer.

Next, we note that any long term contract with the supplier only (w, d_s) is payoff-equivalent to a dual-contract strategy (p, w, d_s, d_B) in which $p = v$ and $d_B = 0$. This can be verified with a comparison of Tables 2 and 4. We can therefore show that the optimal supplier-only contract, (w^*, d_S^*) is weakly dominated by showing that within the space of dual-contracting strategy parameters, $(v, w^*, d_S^*, 0)$ is dominated.

The optimal dual contracting problem is

$$\max_{p,w,d_S,d_B} E\pi_I^{11}(p, w, d_s, d_B)$$

$$\begin{aligned} \text{s.t. } E\pi_S^{11}(p, w, d_s, d_B) &\geq E\pi_S^{01}(p, d_B) \\ \text{and } E\pi_B^{11}(p, w, d_s, d_B) &\geq E\pi_B^{10}(w, d_s). \end{aligned}$$

The transferability of utility means (shadow prices = 1) allows us to re-write this problem as

$$\max(E\pi_I^{11} + \ E\pi_S^{11} + E\pi_B^{11} - E\pi_S^{01} - E\pi_B^{10}).$$

The sum of first three expressions is the total social surplus and is equal to v minus the expected cost of production. Using the expressions above, this problem can be re-written as

$$\begin{aligned} \max \qquad L_I^{11} &= v - \int_0^{p-w-d_S-d_B} c_E dF(c_E) - [1 - F(p - w - d_S - d_B)]c_I \\ &\quad - \int_0^{p-d_B} \lambda(p - c_E - d_B)dF(c_E) - \int_0^{v-w-d_S} (1-\lambda)(v - w - c_E - d_S)dF(c_E). \end{aligned}$$

The first derivative of the objective with respect to d_B is

$$\frac{\partial L_I^{11}}{\partial d_B} = f(p - w - d_S - d_B)(p - w - d_S - d_B - c_I) + \lambda F(p - d_B). \tag{5}$$

To evaluate the partial effect of d_B on the incumbent's expected payoff at the value of $d_S = d_S^*$, $w = w^*$, $p = v$, and $d_B = 0$, we substitute these values into equation (5) and get

$$\frac{\partial L_I^{11}}{\partial d_B} = f(v - w^* - d_S^*)(v - w^* - d_S^* - c_I) + \lambda F(v). \tag{6}$$

The sign of expression (6) show*s* us how the incumbent's expected payoff under a dual-contract is affected by a small change of d_B in the neighborhood of the values replicating the optimal single supplier long-term contract. The first part of this equation is also present in equation (4), which is equivalent to

$$f(v - w^* - d_s^*)(v - w^* - d_s^* - c_I) = -(1 - \lambda)F(v - w^* - d_s^*).$$

After substituting the first part of expression (6), we find that except at the point of

$$\lambda^\dagger = \frac{F(v - w^* - d_s^*)}{F(v - w^* - d_s^*) + F(v)} \in [0, 1],$$

the incumbent's expected payoff $E\pi_I^{11}$ can be increased by either increasing or decreasing d_B from zero. Specifically, for $\lambda < \lambda^\dagger$, $\frac{\partial L_I^{11}}{\partial d_B} < 0$; $\lambda = \lambda^\dagger$, $\frac{\partial L_I^{11}}{\partial d_B} = 0$; $\lambda > \lambda^\dagger$, $\frac{\partial L_I^{11}}{\partial d_B} > 0$. Hence, in terms of the incumbent expected payoffs, the optimal single supplier's long-term contract (w, d_s) is weakly dominated by the optimal dual-contract (p, w, d_s, d_B).

Dominance of single contract strategy over the no-contracting strategy:

Next, we prove the dominance of the strategy of contracting with only the buyer over no contracting strategy by (a) finding a specific buyer-only long term contract which is not only payoff equivalent to no contracting and then showing that (b) this specific buyer-only contract could be further improved upon by decreasing p or increasing d_B. Therefore, we conclude that no contracting strategy is dominated by a buyer-only contracting strategy (provided that $\lambda > 0$).

First, we note that a buyer-only long term contract (p, d_B) is payoff equivalent to no contracting if $d_B = a$ and $p = a + c_I$, for any a, from a comparison of Tables 1 and 3.

The incumbent's expected payoff under no contracting (Appendix 1) is $E\pi_I^{00} = (v - c_I) - \int_{c_I}^{v} F(c_E)dc_E$. Let $a = E\pi_I^{00}$ and consider the long term contract $(p, d_B) = (a + c_I, a)$. We show first that this contract satisfies the buyer's participation constraint, i.e.

$E\pi_B^{01}(p, d_B) \geq E\pi_B^{00}$: Evaluated at the particular long term contract,

$$\begin{aligned} E\pi_B^{01} &= (v-p) + \int_0^{p-d_B} (1-\lambda)(p-d_B-c_E)dF(c_E) \\ &= (v-c_I-a) + \int_0^{c_I} (1-\lambda)(c_I-c_E)dF(c_E) \\ &= \int_{c_I}^{v} F(c_E)dc_E + \int_0^{c_I} (1-\lambda)F(c_E)dc_E. \end{aligned}$$

The buyer's expected payoff under no contracting is

$$E\pi_B^{00} = \int_0^v (1-\lambda)F(c_E)dc_E$$

Since $\int_{c_I}^v F(c_E)dc_E \geq \int_{c_I}^v (1-\lambda)F(c_E)dc_E$, we know that the contract satisfies the buyer's participation constraint.

Next, we demonstrate that the single buyer contract defined can be further improved upon by changing the values of p and d_B. The optimal buyer-only long term contract is

$$\max_{p,d_B} E\pi_I^{01}(p, d_B) \text{ s.t. } E\pi_B^{01}(p, d_B) \geq E\pi_B^{00}.$$

Using the payoffs provided in Appendix 1, and noting that the shadow price = 1, we can write the Lagrangian for this problem as:

$$\begin{aligned} L_I^{10} &= \int_0^{p-d_B} d_B dF(c_E) + [1-F(p-d_B)](p-c_I) + (v-p) \\ &\quad + \int_0^{p-d_B} (1-\lambda)(p-d_B-c_E)dF(c_E) - \int_0^v (1-\lambda)(v-c_E)dF(c_E), \end{aligned}$$

whence

$$\frac{\partial L_I^{01}}{\partial p} = -f(p-d_B)(p-d_B-c_I) - \lambda F(p-d_B);$$

$$\frac{\partial L_I^{01}}{\partial d_B} = f(p - d_B)(p - d_B - c_I) + \lambda F(p - d_B).$$

After substituting the value of p and d_B defined previously, we find

$$\frac{\partial L_I^{01}}{\partial p} = -\lambda F(c_I) < 0$$

$$\frac{\partial L_I^{01}}{\partial d_B} = \lambda F(c_I) > 0$$

and therefore changes in the parameters p and d_B can improve L_I^{01}. ■

Bibliography

Acemoglu, D., P. Aghion, R. Griffith, and F. Zilibotti. Forthcoming. Vertical Integration and Technology Theory and Evidence. *Journal of European Economic Association*.

Aghion, P. and P. Bolton (1987). Contracts as a barrier to entry. *The American Economic Review 77*, 388–401.

Ahmadjian, C. and J. Lincoln (2001). Keiretsu, Governance, and Learning: Case Studies in Change from the Japanese Automotive Industry. *Organization Science 12-6*, 683–701.

Ahmadjian, C. and J. Oxley (2006). Using Hostages to Support Exchange: Dependence Balancing and Partial Equity Stakes in Japanese Automotive Supply Relationships. *Journal of Law, Economics, and Organization 22*(1), 213–233.

Baldwin, R. and D. Taglioni (2006). Gravity for Dummies and Dummies for Gravity Equations. Working Paper 12516, National Bureau of Economic Research.

Basker, E. and P. H. Van (2007). Wal-Mart as Catalyst to U.S.–China Trade. (0710). Available at http://ideas.repec.org/p/umc/wpaper/0710.html.

Basker, E. and P. H. Van (2008). Imports "R" US: Retail Chains as Platforms for Developing-Country Imports. available at: http://ssrn.com/abstract=1123282.

Disdier, A.-C. and K. Head (2008). The Puzzling Persistence of the Distance Effect on Bilateral Trade. *The Review of Economics and Statistics 90*(1), 37–48.

Freshfields (2006). Shareholder rights in a japanes business corporation. accessed in October, 2006.

Fukao, M. (1995). Financial Integration, Corporate Governance, and the Performance of Multinational Companies.

Javorcik, B. S. and Y. Li (2007). Do the Biggest Aisles Serve Brighter Future? Implications of Global Retail Chains' Presence for Romania.

Papke, L. and J. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*, 121–133.