Title
Three Essays on Intra-firm Transactions

Author
Hong, YongKi

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Three Essays on Intra-firm Transactions

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

YongKi Hong

2022

ABSTRACT OF THE DISSERTATION

Three Essays on Intra-firm Transactions

by

YongKi Hong
Doctor of Philosophy in Economics
University of California, Los Angeles, 2022
Professor Jonathan E. Vogel, Chair

In this thesis, I empirically analyze the transactions between firms and their related parties. I construct a novel granular dataset containing South Korean firms' related-party transactions, including intra-group sales and purchases as well as loans and debts. On the topic that in many contexts has remained a black box due to the lack of data, this thesis aims to uncover novel stylized facts and analyze the determinants and implications of these transactions.

In Chapter 1, I focus on sales and purchases between related parties and study how measurement issues have critically influenced the literature's view on such trades. Despite the importance of intra-firm trades in theories of the firm, an empirical literature using proxy measures has documented surprisingly little such trade. I revisit this conclusion using the economy-wide firm-level data from Korea, where related-party trades are directly observable. I show that the true prevalence and volume of the trades are much greater than previous measures indicate. Past proxy measures that rely heavily on input-output tables appear to dramatically underestimate the trades, capturing only 17.6% of related parties that trade and 32.6% of their sales volume.

In Chapter 2, I propose novel methods to infer supply chains within business groups.

Research into topics such as vertical integration has strove to discern firms within business groups that constitute a supply chain aside from those held for unrelated reasons. Due to the lack of detailed data, extant literature has had to rely heavily on proxy measures of such information, despite mounting evidence against their accuracy. Using random forests algorithms with detailed intra-group trade data from South Korea, I propose alternative methods to infer trade within ownership networks. Compared to traditional proxies that rely on input-output table coefficients, this measure shows substantially improved performances while requiring only a small amount of additional information that is easily accessible to researchers.

Chapter 3 focuses on another key aspect of related-party transactions, the flow of capital through loans. Utilizing intra-group transactions data of South Korean firms, this paper provides novel evidence that business groups' internal capital markets are utilized for the group, instead of the private benefits of controlling shareholders. While higher cash-flow rights of controlling group owners do not predict lending relationships to the firms, empirical evidence consistently points to loans flowing to firms with higher marginal benefit of capital. Proximities within ownership and supply chains are highly correlated with lending relationships, suggesting that major factors behind intra-group capital flows include concerns of performance spillover between firms, as well as extending and supporting internal supply chains. Finally, this paper documents novel stylized facts on a specific type of loan, trade credit, that highlight the need to analyze them in a different context from non-trade credit loans.

The dissertation of YongKi Hong is approved.

Martin B. Hackmann

Pablo D. Fajgelbaum

Simon A. Board

Jonathan E. Vogel, Committee Chair

University of California, Los Angeles

2022

To my wife, family, and friends, who enable everything good in my life

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I have my advisor Jonathan Vogel to thank for everything it took to complete this dissertation. From the get-go, he has patiently and repeatedly showed me the way to go and the steps it would take to get there. Without his comments, patience, and willingness to shape me into a researcher, it would have been impossible to finish the long journey of Ph.D. I owe him a lifetime of gratitude and respect.

Professors Simon Board, Pablo Fajgelbaum, and Martin Hackmann have all graciously guided me throughout the Ph.D. process. Simon's intuitions and firm support, Martin's friendliness and willingness to discuss, and Pablo's research acumen were vital in finishing the program.

Comments and guidance from participants in the intra-department International Trade Proseminar, as well as the Industrial Organization Proseminar have provided continuous stimulus and momentum for me throughout the program. Among them, I want to express special gratitude to my friends Brian Pustilnik, Ivan Lavrov, Luke Yasheng Zhang, and Conor Foley, for their friendship and discussions. People say that to finish the Ph.D. program, talks with peers are as important as talks with advisors—through them I have repeatedly found the saying to be true.

UCLA Graduate Division, Department of Economics, and Samsung Foundation have graciously provided scholarships for me throughout the process, which granted me a chance to focus solely on my research. Chiara Paz and everyone else in the Department of Economics have always looked out after the needs of graduate students, and they deserve to know that everyone in our cohort is immensely thankful for their help.

On a more personal note, I thank everyone in my family. My parents and two brothers have been the pillars that I could lean on throughout life. Lastly, but by far the most importantly, I thank my wife, Seo-young Silvia Kim, who has been there from the start to the end, every moment of the way. It is truly her who made all this possible.

VITA

2007–2014	B.A. in Economics, Seoul National University, Seoul, South Korea
2014–2016	M.A. Candidate in Economics, Seoul National University, Seoul, South Korea
2016-2017	M.A. in Economics, University of Los Angeles, Los Angeles, California, U.S.A.

CHAPTER 1

Related-Party Trades in Vertical Integration

1.1 Introduction

An extensive literature on the theory of the firm focuses on why some trades are moderated within the boundary of a firm whereas others occur at arm's length. Transaction cost economics (Williamson, 1971, 1979; Klein et al., 1978) and the property rights approach (Grossman and Hart, 1986; Hart and Moore, 1990) are particularly influential perspectives on this issue. Both theories build on the premise that substantial vertical trades occur within entities.

This fundamental hypothesis has been infrequently tested; when tested, it has not fared well with data. Information on sales and purchases of inputs within-firm, or with related parties of a firm,[1] are not generally available. Hence, large-sample empirical works on the topic are scarce and are either reliant on proxy data that require strong assumptions or are based only on a specific, small portion of an economy.[2] Moreover, these empirical studies suggest that related-party trades are small and sparse, in a critical divergence from the theoretical literature.

Atalay et al. (2014)—henceforth AHS—were the first to empirically test this issue with a large sample, and this seminal paper has served as a benchmark for subsequent

[1]Related parties refer to entities connected through a sufficient amount of control or ownership, such as parent companies, subsidiaries, and so on. Exact definitions differ by study and dataset; the criteria used in this study is outlined in section 1.2.

[2]Ramondo et al. (2016) and Nunn and Trefler (2013) are among the strand of the literature that studies U.S. firms' trades with *foreign* subsidiaries by utilizing customs data.

researches. In the absence of data on intra-firm trades, AHS construct a novel proxy from U.S. establishments' shipments data in the Commodity Flow Survey, using geographical information as well as industry-level proxies. Their baseline results show that almost 50% of establishments with at least one related party do not sell anything to them; and even when they do, the sales comprise only a small portion of the sellers' economic activities. This was a surprising conclusion: if there is little vertical trade, how can facilitating trade be a central driver of integration? How can vertical integration enhance efficiency?[3]

In this paper, I present the first economy-wide direct measurement of related-party trades and show that there is, in fact, substantial trade within integrated firms. I construct a novel firm-level dataset that enables this direct observation. Exploiting a South Korean accounting requirement, I web-scraped firms' annual trades with each of their related parties from the side notes of *all* publicly available financial statements in Korea—similar to 10-K reports in the U.S.—between 2013–2019. As firms are explicitly requested to report related-party transactions, this data shows the trades for all firms in the economy above a set of size thresholds,[4] without having to rely on a proxy measure.

I find that firms utilize related-party trades–in terms of both prevalence and value–substantially more than the previous literature has documented. Almost all manufacturing firms with a related party appear to engage in related-party trades: 87.2% of the firms report either sales to or purchases from a related party, and 77.3% report related-party sales during a fiscal year. What is more, the trades in my data are a considerably larger part of firms' activities than past estimates. These results imply that vertically integrated

[3]This discussion has influenced the directions of pivotal economic policies. In September 2021, the Federal Trade Commission (FTC) issued a statement to withdraw support for the Vertical Merger Guidelines issued jointly with the Department of Justice in 2020. In the statement, the FTC cites AHS to argue that the guideline puts too much emphasis on the pro-competitive effects and efficiency gains from vertical mergers: for example, "we should be highly skeptical that EDM [Elimination of Double Marginalization] will even be realized" as "[in] many cases, vertical integration does not even prompt firms to provide the upstream input to its own downstream division." (FTC (2021))

[4]One representative threshold is the firm's total sales surpassing roughly $8.3 million U.S. dollars; see Section 1.2.

firms actively utilize related-party trades, consistent with a focus of the vast theoretical literature.

Subsequently, I explore why related-party trades are substantially larger in my data compared to previous studies. I show that rather than the simple difference in country, the main driver of the different results appears to be the data that is better suited to address the question. Specifically, I construct a dataset where related-party trades are directly observed, whereas the lack of such measurement has compelled previous works to infer them using proxy measures. In most cases, researchers observe only either relatedness or trades: that is, they observe trades but not whether the two sides are related, or see that two entities are related but not whether they trade.[5]

To infer internal sourcing without direct observation, researchers have long depended on a combination of (i) industry-level trade patterns represented by Input-Output Tables (IOT) and (ii) entity ownership; see, e.g., Alfaro et al. (2019), Atalay et al. (2019), Acemoglu et al. (2009), and Aghion et al. (2006). For example, consider two entities, A_i and A_j, with a common owner and belonging to industries i and j, respectively. The common approach regards A_i to be purchasing from A_j only if industry i uses output from industry j more than some arbitrary threshold according to IOT. Unfortunately, this approach may not provide a close approximation for the presence of trades between A_i and A_j. First, while there may exist some IOT threshold above which A_i is likely to buy from A_j, we require data on intra-party trade to know the value of this threshold. More generally, related-party trade patterns may differ fundamentally from the general economy upon which the IOTs are built.

Indeed, I show that when compared with the directly observed data, the proxies' accuracy

[5]In a concurrent work, Garg et al. (2021) also directly measure related-party trades by utilizing a regional dataset from a state of Karnataka in India and find larger trades compared to AHS. Relative to Garg et al. (2021), the dataset utilized in this paper can represent an entire economy that is more advanced and has strong contractual enforcement, visualize international related-party trades in addition to domestic, and include flow of services as well as physical goods. Furthermore, this paper utilizes the data to construct more accurate predictive algorithms that allow future work without access to intra-firm trade data to better predict these trades.

is heavily sensitive to the choice of cutoffs, and is generally low. A common coefficient cutoff of 1% (e.g. AHS, Aghion et al. (2006)) captures only 17.6% of related-party pairs that trade in the Korean data, and less than one-third of the sales volume. In fact, using Korean data together with the 1% cutoff yields a result that remarkably resembles AHS's results. Similarly, AHS's robustness check that drops the IOT requirement and only utilizes geographical information yields results that are remarkably similar to those obtained from the true Korean data. Yet another common cutoff used in the literature of having a positive total requirements coefficient (e.g. Alfaro et al. (2019)) is in turn too lenient and results in assuming almost all of related parties are trading.

These results suggest that the IOT-based proxies for related-party trades, when used on its own, appear to have generated incorrect conclusions and an inconsistency between theory and data.[6] Yet these same proxies have been central in addressing a range of questions. For example, the proxies have been used to discern production chains within integrated firms and subsequently answer questions such as why firms only integrate specific parts of the production chain (Alfaro et al., 2019), what factors induce more vertical integration (Acemoglu et al., 2009; Blyde and Molina, 2015; Alfaro et al., 2016), and also to separate out vertical from horizontal FDI (Fajgelbaum et al., 2015; Alfaro and Charlton, 2009). Given the importance of these questions and the sensitivity of inferring related-party trades on the assumptions used, there is a strong need for a better way of inferring trades within ownership structures.

The rest of the paper proceeds as follows. Section 1.2 discusses the original data collection and contents. Section 1.3 presents the related-party trade of Korean firms and discusses the underestimation problem of the proxy measure. Section 1.4 presents robustness checks. Section 1.5 concludes.

[6] Ramondo et al. (2016) also notes that the IOT coefficients are not correlated with the observed U.S. parent companies' trades with foreign subsidiaries in their data.

1.2 Data

This paper draws on several sources to construct data. I first collect and construct the key dataset of firm-to-firm related-party trades, then combine it with existing firm-level databases to match firm characteristics to both sides of the trades. Here, I describe the data collection and matching.

Related-party trade data. — The primary strengths of the data used in this paper are that related-party trades are observed directly, for a large sample of firms, and with high credibility. In most cases, researchers either observe trades without knowing whether the trade partners are related or observe related firms but not whether they trade. This is the first economy-wide dataset that incorporates both components at the same time. Moreover, the reports undergo external audits and government scrutiny, ensuring the results' accuracy.

A South Korean accounting requirement enables this direct observation of related-party trades.[7] All Korean firms satisfying a set of size thresholds (henceforth *reporting firms*)[8] are legally required to get an annual external audit and make the reports—analogous to 10-K reports in the U.S.—publicly accessible. Crucially to this paper, in the side notes of the reports, the firms need to disclose all annualized trades with each of their related parties.[9]

[7]The International Financial Reporting Standards (IFRS), a widely used accounting standard, requires related-party transactions to be reported (IAS 24). However, partly due to the complexity of creating a dataset from the document-based information, this section of the financial reports have not received much academic attention from economists until recently. Santioni et al. (2020), which analyzes Italian firms' related-party loans and debts, is a good example of the recent efforts to utilize this information. Compared to other countries that have adopted IFRS, South Korea stands out as it had a similar reporting requirement even before the country's adoption of IFRS in 2011.

[8]In the last year of the data, fiscal year 2019, firms satisfying at least two of the following four thresholds were required to receive an external audit and disclose 10-K reports: (1) sales of $\geq$ approximately 8.3 million USD (originally 10 billion Korean Won), (2) assets of $\geq$ 10 million USD (12 billion KRW), (3) debts of $\geq$ 5.8 million USD (7 billion KRW), or (4) more than 100 employees. The size threshold varies slightly over the data period, reflecting factors such as inflation; details can be found in the data appendix.

[9]As this information is reported only in the side notes and not on the main financial statements, it has never been built into a database with the same scope and detail as this paper. The two largest firm-level

I scraped and cleaned the filings from an official website[10] maintained by a Korean government agency (*Financial Supervisory Service*), which is comparable to the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database maintained by the U.S. Securities and Exchange Commission.[11] I examine all 190,725 available financial reports from FY 2013–2019: after ruling out reports that were found to be inadequate or inaccurate by external auditors,[12] or had issues in scraping, 176,657 (92.6%) reports from 44,701 firms are used in the analysis. The main body of the analysis utilizes only the firm-years when the reporting firms are confirmed to have had at least one related party. After the exclusion of singleton firms, 121,519 annual reports from 30,390 firms are used in the analysis that follows.

A related party is defined primarily by the control—that is, the voting rights—that one firm possesses over the other. In this way the definition of related party utilized in this paper is based on *control*, and not mechanically on the share of outstanding stocks owned.[13] For example, if firm A owns a right to name the entire board of directors of firm B, A and B are deemed related parties even if A owns only a small share of B's outstanding stocks.

Related parties in this data are commonly connected through majority control rights, but

---

databases in Korea each contain a part of these data: TS2000 and KISVALUE. TS2000 reports firm-to-firm related-party trades but only for approximately 3,000 very large firms that are publicly traded on the stock market. KISVALUE, on the other hand, reports the trades for more firms, close to about 20,000, but all trades are aggregated to the firm-level only (i.e., how much a firm trades with each of its related parties is invisible). In Appendix 1.B I show that when my data is limited to the same scope as TS2000 or aggregation as KISVALUE, it largely reproduces the numbers in the existing databases.

[10]http://dart.fss.or.kr/

[11]A number of studies have used the EDGAR database to utilize information in the sections of 10-K filings that were not traditionally included in large databases (Hoberg and Maksimovic, 2015; Hoberg and Phillips, 2016). While 10-K filings of U.S. firms also include related-party trade information, the Korean filings offer a steep advantage for digitizing as they are primarily filed in table formats, not in sentences.

[12]Specifically, if external auditors issued a *disclaimer of opinion* as they could not obtain sufficient audit evidence from the company, I exclude the report.

[13]Details can be found on the financial reporting standards used in Korea (1024 K-IFRS § 11).

minority control rights are sufficient in some cases.[14] In the first category, related parties are within a network connected from the reporting firm with controlling voting power. It does not have to be a direct subsidiary or a direct parent company: regardless of the number of ownership links between the reporting firm and the related party, trades between the two are reported as long as each link satisfies one side having a controlling voting right of the other. Second, a trade partner is also deemed a related party if it is *directly* linked to any entity in the first category with 20% or more voting power. In a rough summary, for two firms to be deemed a related party, up to one link between firms of 20% voting power is allowed, as long as all other links are above 50%.[15]

Firms report the transactions separately for each of their related parties. I process the transactions into four categories—sales, purchases, loans, and debts.[16] In the analysis, I utilize 558,114 firm-year-related party triples.

The financial reports disclose the related parties' names and their relationship with the reporting firm. The relationships are reported in discrete categories, with information on (1) whether a link with less than 50% voting rights is included and (2) in which direction the control runs: in other words, whether the trade partner is a subsidiary, a parent, or whether there is a lateral component in the direction of ownership—as in 'sibling' or 'uncle' firms.

The financial statements, including the side notes, go through strong government scrutiny on top of external audits. An intentional misreporting of the information can lead to strong punishments: in a widely reported case in 2018, public stock trading of *Samsung Biologics* was suspended as the firm had allegedly intentionally misreported the firm's

[14] This section outlines a rough guide: detailed criteria can be found in the Data Appendix 1.B and are based on the financial reporting standards (1024, 1028, 1110 K-IFRS).

[15] Related parties are defined with different levels of control, or shareholding, depending on available data. Majority shareholding is commonly required (AHS, Ramondo et al. (2016)) but lower thresholds are also utilized, especially with the customs data (Ruhl, 2015).

[16] In practice, firms often choose to report the numbers in more detailed categories than what the guideline requires. The numbers are re-summarized into the four categories (sales, purchases, loans, and debts) during data cleaning. Complete reporting requirements are provided in the appendix 1.B.

relationship with a related party.[17]

Firm Information and Matching. — Aside from the transactions and relationships with related parties, all firm-level information is matched to the firms in the scraped dataset from three existing databases: the ORBIS database by Bureau van Dijk and the two largest firm-level databases in Korea, KISVALUE and TS2000. The linked firm-level variables include total assets and sales, industry, and location. Each reporting firm is matched to the databases using 10-digit firm identifiers assigned by tax authorities.[18] The related parties of the reporting firms are matched by name, as no other identifying information is provided on the financial statements. Extensive checks on the matches are undertaken to ensure accuracy.[19] Some are inevitably unmatched to firm-level information: in robustness checks in Section 1.4, I show that these cases are not vital to the main results.

Importantly, the ORBIS ownership database is used to discern the set of firms that have at least one related party. This paper excludes singleton firms from the sample as they inherently cannot engage in related-party trades. However, the 10-K reports do not provide a firm's full list of related parties, instead only displaying those that have *transacted* with the reporting firm—be it sales, purchases, loans, or debts—during the fiscal year. Hence, when relying solely on the reports, a firm that has related parties but is not actively trading with any will be marked as a singleton and excluded from the sample. I use the ORBIS database to supplement the 10-K reports, as it provides for each firm the list of all other firms in the database that share the same *ultimate owner*.[20]

---

[17]WSJ, Jeong and Martin (2018), "South Korea Regulator Says Samsung BioLogics Violated Accounting Rules" (https://www.wsj.com/articles/south-korea-regulator-says-samsung-biologics-violated-accounting-rules-1531407128). The case is currently on trial (May 23, 2022).

[18]The identifier utilized in this paper is the *Business Registration Number*, an identifier for each business used by the National Tax Service (the Korean equivalent of the Internal Revenue Service in the U.S.).

[19]The matching and cleaning process is akin to the procedures outlined in Alfaro-Ureña et al. (2020) for firm-to-firm trades; for example, when the related party also files its own 10-K report, I examine the reports from both sides to verify the match. See Data Appendix 1.B for details.

[20]The ORBIS database's definition of related party differs in detail from my data based on Korean financial statements. For example, in the ORBIS database, the minimum ownership percentage criterion

Throughout this paper, I take a conservative approach and use the set of firms reported to have a related party either in ORBIS or in the 10-K reports.[21]

Input-Output Table. — This paper utilizes IOTs to measure the accuracy of the previous literature's proxy measures and show the extent of possible biases. I use the 2015 Bank of Korea Input-Output Table (*use table*) throughout the entire data period, as it is the only publicly available version offering the most finely disaggregated sectors (381 commodities and 278 industries). Using the official concordance tables, I then link the IOT with 5-digit Korean Standard Industry Classification (KSIC-10) codes and also with 5-digit 2017 NAICS codes.[22]

1.3 Sales and Purchases with Related Parties

In this section, I analyze the Korean firms' trades with related parties. Two findings emerge. First, most firms appear to engage in related-party trades, and the relative importance of these trades is much greater than what has been established in the literature. Second, the discrepancy from existing literature appears to stem from a considerable underestimation inherent in the proxy measure that has been widely used to infer related-party trades; specifically, measures based on IOT coefficients show poor performance in predicting the trades.

---

is 25% while it is 20% in the Korean data. While these minor differences may affect the sample of firms that I consider, the significance of them is expected to be very small: observations with smaller than 50% ownership share are already scarce in the ORBIS database (3.34%). While it is possible that there could be a bunching between 20% and 25% ownership shares, there is no reason to anticipate the bunching to exist or the magnitude of it to be consequential.

[21] As the ORBIS database only provides the most recent ownership information for each firm, the set of firms that have related parties in ORBIS in 2019 are utilized throughout the entire data period. I contend that any possible biases from this limitation are not significant to the main results in Table 1.1. First, any bias will include singleton firms in the sample that I consider, more so for the earlier years of the data, and therefore result in underestimating firms' related-party trades. As the main finding of the paper is that related-party trades are larger and more prevalent than previously thought, this paper's results will only be stronger without the possible biases. Second, the possible biases are likely small, as the attrition rate is small: using a sub-sample of all public firms in Korea, for which the complete list of firms' related parties are observable, I confirm that 91.3% of firms with a related party still do after 3 years.

[22] As is common, concordance tables are not one-to-one for a number of industries. Where one IOT code is matched to multiple industry codes, I split the IOT coefficients equally between industries.

1.3.1 Usage of Related-Party Trades by Korean Firms

While the data spans all industries, the main analysis will primarily utilize reports from firms in manufacturing industries. The reasons for this are threefold. First, it enables a more clear comparison with existing studies, a vast majority of which focus on the manufacturing sector. Second, it allows for a greater emphasis on trades of goods rather than services. Last, a focus on the manufacturing sector reduces concerns about possible effects from transfer pricing. While transfer pricing is monitored closely in Korea,[23] manufacturing firms have even less room to maneuver as the fair market value is easier to calculate for manufactured goods than services.

Panel 1 of Table 1.1 reports the share of related-party sales in total firm sales. Following the convention used in AHS, I present quantiles of distributions of the shares. I first note that most firms report positive quantities of related-party trades. 77.3% of manufacturing firms report selling to a related party during the fiscal year, and in fact, 87.2% of the firms have either sales to or purchases from related parties.

Moreover, sizes of the related-party sales appear substantially greater than what the literature has previously found. In Panel 2 of Table 1.1, the findings of AHS are presented as a benchmark. The results from Korean manufacturing firms illustrate a greater importance of related-party sales throughout the entire distribution. The 75th percentile firm in AHS sells only 7.0% of its sales to related parties, while the corresponding number is 17.8% for Korean firms.[24] In the 90th percentile entity, the gap is even larger, where the share of related-party sales is close to 60% in Korea but is less than 40% in AHS.

While the overall level of trades is higher, other general characteristics of the distribution

[23] All related-party trades that show more than (i) a 5% difference from the market price *or* (ii) a $250,000 difference in total value from the fair market value are subject to an investigation by Korean tax authorities.

[24] Note that the comparison is based on the Manufacturing industries in Korea, while AHS is based on establishments in the Manufacturing, Mining, Wholesale, and select Retail industries that are included in the Commodity Flow Survey. In robustness checks, on Table 1.2, I show that using all industries in CFS generates almost identical results.

Table 1.1: Share of Sales to Related Parties

Related Party Share of Firm Total Sales	Percentiles (%)				Fraction (%)		Weighted Mean (%)	N
	50th	75th	90th	95th	=0	≥1		
Panel 1: Korean Firms								
Manufacturing Firms	2.7	17.8	57.4	90.6	22.7	2.1	33.6	54,042
All Firms	1.3	15.1	62.4	95.8	30.3	3.0	24.0	121,519
Panel 2: Literature								
AHS (2014) - Main Result	0.1	7.0	37.6	69.5	49.7	1.2	16	67,500
Panel 3: IOT Requirement								
Manufacturing Firms	0.8	8.7	37.4	69.7	26.6	1.1	16.9	26,077
All Firms	0.1	3.5	24.4	52.5	36.7	0.7	9.7	52,298
Panel 4: Requirements à la AHS								
Manufacturing Firms	0.2	4.7	30.9	69.0	34.3	1.2	9.9	18,194
All Firms	0.0	1.9	19.1	50.5	42.7	0.8	6.4	35,955
Panel 5: Intensive Margin Only								
Manufacturing Firms	7.0	31.7	82.0	99.4	13.5	3.8	38.1	26,232
All Firms	4.8	30.2	86.3	100.0	19.3	4.7	26.8	52,696

NOTE.—The table reports the share of firms' sales that is sold to related parties. N consists of firm-years that have at least one related party: due to missing firm-level information, total N used on the table is slightly smaller than the 125,044 available firm-level reports. The weighted mean is weighted by the size of each reporting firm's total sales.

are similar to the literature (AHS, Ramondo et al. (2016)). The internal trade share distribution is highly skewed, although to a lesser degree. Also, the share of related-party sales is positively correlated with the selling firm's size. The weighted mean of the shares of related-party sales, weighted by the total sales of firms, is significantly larger than the unweighted mean or the median. However, a difference is that this paper shows much higher shares of intra-party sales overall and, importantly, a much longer left tail: a vast majority of manufacturing firms are selling to at least one of their related parties, even if the relative share is small.

1.3.2 Underestimation Problem in Using Input-Output Tables to Infer Vertical Trading between Related Parties

If the present Korean data show such a consistent and significant difference from the existing literature, what causes this divergence? Since each paper on the topic uses a different dataset, the apples-to-oranges problem renders a rigorous decomposition of the source of the differences difficult. In this section, I highlight a single main driver of the disparity: the low accuracy of the widely-used proxy for related-party trades. Specifically, using IOTs to construct proxies appears to be responsible for much of the inaccuracy.

Confronted with a lack of data on related-party trades, IOTs are widely used to proxy for the related parties that trade with each other. This process is based on Fan and Lang (2000) and utilizes firms' ownership and industry information, which are often easily available to researchers. Assume firm A is in industry J and has related parties $a_1, a_2, \cdots, a_n$, each in industries $j_1, j_2, \cdots, j_n$, respectively. The commonly used method assumes that A sells to a_i only if the industries J and j_i trade *significantly* according to the IOT. The criteria for significant trade varies by studies: in AHS, the criteria is more than 1% of industry J's output being used as an intermediate input in industry j_i, but studies have utilized different cutoffs.

These criteria based on IOT cutoffs are then used to study firms' vertical integration decisions. As an example, Korea Yakult, a firm in the *Manufacture of dairy products and edible ice cakes* industry, has a related party in the same industry (IOT coef. 0.042), one in the *Manufacture of noodles, macaroni, and similar farinaceous products* industry (IOT coef. 0.003), and another in the *Manufacture of truck and motor vehicles for transportation of goods and special purpose* industry (IOT coef. 0.001).[25] Based on industry-level trading patterns, with the 1% cutoff, one would assume that Korea Yakult is selling only to the first related party, and not to the second and third.

[25] The coefficients refer to the proportions of Korea Yakult's industry's intermediate sales that are directed to each related party's industry.

While a paucity of true data have necessitated the use of proxy measures, using IOT coefficients to infer related-party trade inherently entails limitations. Firstly, while IOTs show technological input requirements of industries, they do not account for the idiosyncrasies of related-party trades. That is, how firms source from related parties may fundamentally differ from the general economy upon which the IOTs are built. For example, data shows that Korea Yakult purchases a large amount from the related party that produces special purpose vehicles, in spite of such trade being a very small share of an average dairy-producing firm's behavior. This is because Korea Yakult operates a large fleet of small refrigerating motor vehicles that work as roaming retail stores for the firm's products. As these are highly specialized vehicles that no other company uses, their production is integrated into the firm to reduce holdup costs; due to the exact same reason, the IOT coefficients are small. Second, industry codes may not be able to reflect the firm's entire line of businesses, which is problematic especially for the multiproduct firms.

Moreover, even when taking IOT coefficients as accurate reflections of related-party trades, without actual data to optimize the accuracy of proxy measures, the traditional method is inherently simple and reliant on arbitrary components. It utilizes a single cutoff of input-output table coefficients to categorize a complex decision, and the choice of cutoffs vary widely among studies. Aghion et al. (2006) and Monarch et al. (2017) are among those that echo AHS's criteria. Alfaro and Charlton (2009), Blyde and Molina (2015), and Alfaro et al. (2019), among others, use criteria based on total requirement coefficients, thus taking into account indirect input use by industries as well. However, the specific cutoffs utilized in previous papers range from 0.05 in the first study, 0.0001 in the second and zero in the last, in their baseline specifications.[26] Lastly, papers including Acemoglu et al. (2009), Fort (2017), and Altomonte et al. (2018) do not utilize cutoffs, but instead utilize the coefficients to construct vertical integration indices for each firm.

The benchmark example, AHS, ingeniously utilizes shipments and geographical data

[26]Based on each specific cutoff, predictions of which related parties are trading produce widely different results. Papers often provide robustness checks where they use different cutoffs and show that the qualitative results do not change. However, this is not always the case.

along with the proxy method. The paper is unique in two dimensions. First, by utilizing shipments data, it can speak to the *sizes* of related-party trades, while the traditional proxies mostly concentrate on approximating the *existence* of trades between two related parties. Second, it exploits the geographical locations of entities to achieve a more accurate inference.

AHS's proxy for U.S. establishments' related-party sales are constructed with shipments data in the Commodity Flow Survey (CFS). For each establishment in the survey, CFS samples shipments and provides information on their counts, values, and destination zip codes. However, whether the shipment is internal to the same firm is not observable. The paper, therefore, classifies a shipment as related-party sales only if it is sent to zip codes where (i) a related party of the sending establishment is located *and* (ii) that related party is in an industry that uses more than 1% of the sender's industry's output, according to IOT.[27] In a way, this method treats a group of entities as possible participants of related-party trades by first using the IOT cutoff. Then, it utilizes geographical information to distinguish the shipments that are likely sent to the group.

To demonstrate the extent of limitations that stem from utilizing the IOT cutoffs, I re-create the main results after imposing the same data-generation process used in the previous literature. Panel 3 of Table 1.1 reports the results from the Korean data when counting only the sales to the related parties that satisfy the IOT cutoffs (1%).[28] It is

[27]AHS recognize that retail and wholesale industries are not represented well in the IOT they use: the tables treated the two industries as a single sector. To deal with this problem they utilize other information, such as the Annual Retail Trade Survey and Annual Wholesale Trade Survey. While the current study's Korean tables share the same limitations, the solutions cannot be shared and therefore this paper's comparisons only rely on IOTs. In the robustness checks I show that this limitation does not affect the results.

[28]In Panels 3 and 4 of Table 1.1, if a firm is selling to related parties that are missing industry information, I assume that a fraction of the firm's sales—the 4-digit reporting firm industry's average share of sales to related parties—are to related parties that satisfy IOT coefficient cutoffs. This is a conservative assumption in the sense that it would make the effect of additionally required restrictions (IOT requirement, and the extra restrictions in AHS) smaller. That is, this when additional restrictions are imposed on the data, this assumption would result in an overestimation of the fraction of firms that sell to at least one related party. In section 1.4, I show that alternative treatments of related parties without industry information makes only a minor difference in the results.

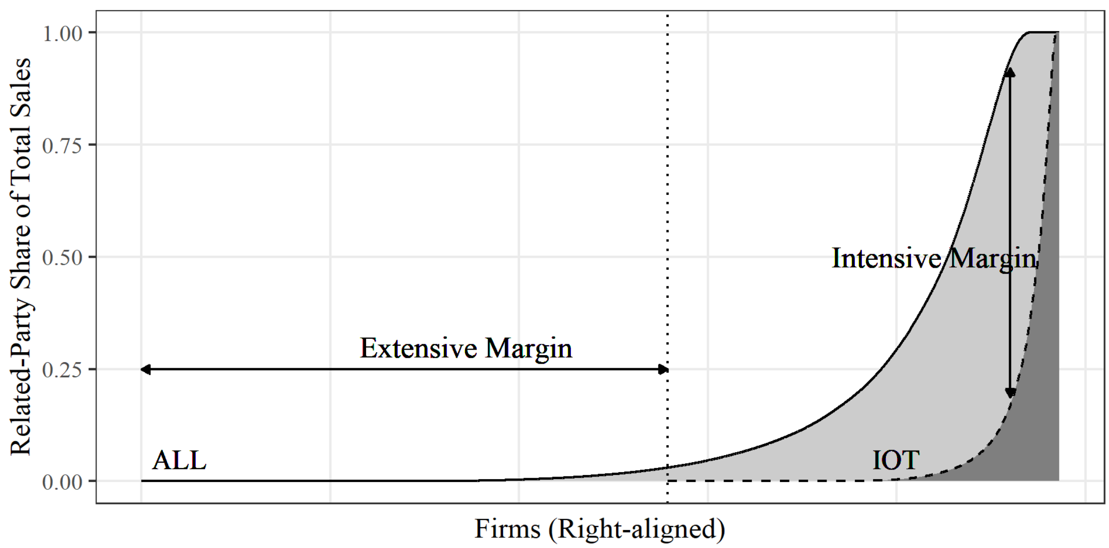
immediately apparent that the limitations of the proxy measures take a substantial toll: the results are now much more similar to AHS, where the use of related-party sales is almost identical to the existing literature for the 90th and 95th percentile firms. In fact, only 17.6% of reporting firm–related party pairs identified in the full Korean dataset satisfy the IOT cutoff, representing 32.6% of the total value of trade in the data.

Panel 4 constructs a more complete comparison with AHS by imposing two additional limitations that pertain to the paper's data. First, as the CFS only provides detailed destination information for domestic sales, AHS only utilizes sales to domestic related parties. Also, only the related parties connected through majority shareholdings are included, while the present study's Korean data also includes a small group of related parties linked with minority voting power, as described in section 1.2. The two additional limitations produce smaller, but still significant changes to the result. The results of Panel 4 now look surprisingly similar to AHS. In this way, the restrictions in the customary data generation process can influence how the true state of vertical trades is perceived.

Furthermore, the effect from the customary data-generation process can be broken down into both extensive and intensive margins. When *tradeable* related parties are defined too narrowly by utilizing IOT cutoffs, related-party trades appear too modest in size as only the trades with a much smaller set of related parties are counted (*intensive margin*). However, the narrow definition also limits the total sample of firms that we consider by reducing the scope of 'firms that have at least one related party' (*extensive margin*). In other words, in a common data-generation process, the firms with related parties only in industries falling short of the 1% cutoff would have been excluded from the sample altogether, deemed as having 'no (tradeable) related party.'

The extensive margin becomes apparent when we compare the size of sample firm-year pairs in Panel 1 with Panels 3 and 4 in Table 1.1. While there is a total of 54,042 manufacturing firm-year pairs that have a related party during the year, only 26,077 (48.2%) are confirmed to have at least one related party in an industry that satisfies the IOT coefficient requirement. Moreover, related-party trades of only 18,194 (33.7%)

Figure 1.1: Comparison of sample limitation on intensive and extensive margins



NOTE.—The figure shows the extensive margin and the intensive margin based on imposing a more stringent definitions of related parties. "ALL" refers to the distribution of firms' share of total sales that is directed to related parties using all firms in the Korean data. "AHS" refers to the same distribution with only the firms that have at least one related party satisfying all three additional restrictions required in the literature (equivalent to Panel 3 of Table 1.1). Note that a simple overlay of the distributions is presented, so that the same horizontal location does not imply the same firm.

firm-years are confirmed to remain after imposing all requirements in AHS. Hence, the extensive margin suggests that the IOT proxies limit researchers' focus to a much smaller set of firms and their activities.

One caveat is on the treatment of related parties that could not be matched with firm-level information such as industries. The size of the extensive margin varies depending on the treatment of the related parties that were not successfully matched with the industry information. The main analysis reports the most conservative specification that would make the sum of the intensive and extensive margin (i.e., the difference between Panels 1 and 4 of Table 1.1 appear the smallest). This, however, likely overestimates the extensive margin. In the robustness check, I report alternative specifications: while the relative size of extensive margin may change, the margin remains consistently large and the qualitative results remain unchanged.

Lastly, the breakdown implies that the intensive margin is much larger than what is visible from a simple comparison. Panel 5 of Table 1.1 displays the effect when only taking the intensive margin into account: in other words, using the sample of firms that have at least one related party satisfying the IOT cutoff, but measuring their true related-party trades. Now the effects appear strikingly large—more than a quarter of the sampled firms sell more than 30% of their sales to related parties. The now excluded set of firms on average have a smaller number of related parties, and hence a smaller share of related-party sales. This result demonstrates that when analyzing related-party trades of a fixed number of firms (i.e., by removing the extensive margin), the limitations of the standard proxy measure will be much more pronounced than for the baseline result.

1.4 Robustness Checks

In this section, I present two sets of robustness checks of the results presented in Section 1.3. The first set of checks investigate avenues where imperfections in the present dataset could have affected the results. Several possibilities are explored: using a different data source to define the set of 'firms with at least one related party'; alternative treatment of the trades with related parties that were unsuccessfully matched to industry information; inclusion of more reporting firms in industries that are covered in CFS; lastly, an alternative treatment of retail and wholesale industries, which are represented with insufficient detail in input-output tables. The results consistently support the main conclusion, with only minor discrepancies from Table 1.1.

In the second set of robustness checks, I also show that analyzing specific segments of the data generates results that are in line with intuition: the results are robust to using related-party purchases as opposed to sales, and the results are also much higher for firms in industries with prior belief of high usage of related-party trades.

Note that the results of robustness checks are presented primarily for the reporting firms in manufacturing industries for the sake of conciseness, unless declared otherwise. The results from the full sample of firms in all industries are consistent with the manufacturing sample, and are listed in the appendix Table 1.A.1.

1.4.1 Alternative Treatments of the Data

First, I test an alternative sample selection, by using a different method to find the set of firms that have at *least one related party*. As Section 1.2 describes, the sample in Table 1.1 is a union of two groups of firms: (i) those that have related-party transactions in the data year, or (ii) the firms that have related parties in the ORBIS database. However, to the extent that only the firms with some related-party transactions are represented in group (i), inclusion of the firms only in (i) may still exert an upward pressure. In Panel 2 of Table 1.2, I only use group (ii) as the sample. The results confirm that the concern is unfounded. In fact, related-party sales appear even larger in this alternative specification compared to the benchmark results presented in Panel 1 of Table 1.2.[29]

The next two robustness checks examine treatments of the related parties that were

[29] Panel 1 of Table 1.2 simply provides the manufacturing firms' results from Table 1.1 for a concise comparison with the results from robustness checks.

not able to be matched to industry information. As detailed in Section 1.2, related-party trade data from Korean financial statements provide only limited information about the related parties. Information on the related parties, such as industry, total sales, and cost of goods sold are matched from multiple existing firm-level databases by the firms' names. However, some related parties are inevitably unmatched to firm-level information, which may affect the results. Specifically, the results of the exercises that impose additional constraints on the data may be impacted. For example, If a related party's industry is not known, then whether it satisfies an IOT cutoff is also unavailable.[30]

The main analysis in Table 1.1 utilizes the reporting firm's industry-level averages in place of missing values. That is, using all related parties *with* industry information, I first calculate the share s_i of related-party sales of firms within each 4-digit KSIC industry i that are directed to the related parties that satisfy the IOT cutoffs. Then if a reporting firm in i sells to a related party that is missing industry information, the share s_i of such sales is assumed to satisfy the IOT cutoffs. Assuming that the share of related party sales that satisfy the cutoffs are not systematically different between the matched and the unmatched groups, this assumption will not bias the results of Table 1.1.

However, this is a strong assumption. In the Appendix, Table 1.A.2 shows that the matched and unmatched groups exhibit differences in both the sizes and types of trades with reporting firms. While there is no evidence that the differences extend to the firms industry, a validation of this assumption is in order.

Panels 3 and 4 of Table 1.2 test two assumptions on opposing ends of the spectrum and confirm that the main results remain consistent in either case. Panel 2 is calculated assuming that all unmatched related parties do not satisfy the IOT cutoff. In this specification, imposing proxy-generating processes eliminate the largest share of observed

[30]Note that the limitation discussed here does not affect the calculation of Korean firms' related-party sales' share in Panel 1 of Table 1.1. Any possible effects are limited to the results in Panels 3 and 4 of the same table.

Table 1.2: Share of Sales to Related Parties, Manufacturing Firms' Reports Only

Related Party Share of Firm Total Sales	Percentiles (%) 50th	75th	90th	95th	Fraction (%) =0	≥1	Weighted Mean (%)	N
Panel 1: Main Result - Share at Reporting Firm Industry-Level								
All trades	2.7	17.8	57.4	90.6	22.7	2.1	33.6	54,042
IOT Requirement	0.8	8.7	37.4	69.7	26.6	1.1	16.9	26,077
AHS Requirement	0.2	4.7	30.9	69.0	34.3	1.2	9.9	18,194
Intensive Margin Only	7.0	31.7	82.0	99.4	13.5	3.8	38.1	26,232
Panel 2: All Firms with a Related Party in the ORBIS Database								
All trades	4.8	25.6	74.0	98.3	20.4	3.3	37.4	25,696
IOT Requirement	0.7	8.5	37.2	66.6	27.6	0.9	17.8	15,717
AHS Requirement	0.0	2.0	22.8	55.9	44.5	0.9	11.1	8,018
Intensive Margin Only	8.5	35.2	86.4	99.9	13.4	4.3	40.1	15,826
Panel 3: Assume No Related Party without Industry Information Vertically Related								
All trades	2.7	17.8	57.4	90.6	22.7	2.1	33.6	54,042
IOT Requirement	0.0	2.2	22.2	54.2	54.2	0.8	8.3	26,236
AHS Requirement	0.0	2.0	23.2	61.4	52.2	1.0	5.0	18,276
Intensive Margin Only	7.0	31.7	82.0	99.4	13.5	3.8	38.1	26,232
Panel 4: Assume All Related-Parties without Industry Information Vertically Related								
All trades	2.7	17.8	57.4	90.6	22.7	2.1	33.6	54,042
IOT Requirement	0.8	9.2	36.4	67.0	31.5	1.0	24.9	49,234
AHS Requirement	0.2	4.2	22.2	50.3	36.1	0.7	12.5	35,888
Intensive Margin Only	3.2	19.2	60.0	92.2	21.1	2.3	34.2	49,237
Panel 5: All Sales to RW industries as Vertically Related								
All trades	2.7	17.8	57.4	90.6	22.7	2.1	33.6	54,042
IOT Requirement	1.2	10.5	41.2	73.2	24.2	1.0	21.8	30,772
AHS Requirement	0.4	6.3	34.5	73.4	29.4	1.2	11.4	21,804
Intensive Margin Only	6.3	29.1	78.2	98.8	14.1	3.5	36.9	30,950
Panel 6: Reporting Firms in All CFS industries								
All trades	2.5	16.8	56.7	90.6	22.3	2.1	31.3	66,126
IOT Requirement	0.5	6.9	32.8	63.4	27.2	0.9	15.0	29,455
AHS Requirement	0.1	4.1	28.3	64.9	35.1	1.1	9.1	19,558
Intensive Margin Only	8.5	36.1	87.7	99.9	10.7	4.4	38.9	19,669
Panel 7: Related-party Purchases								
All trades	3.8	17.5	41.8	62.0	23.8	0.8	28.8	53,175
IOT Requirement	0.9	8.4	24.9	40.9	26.3	0.2	14.9	23,368
AHS Requirement	0.2	4.6	18.9	35.3	37.2	0.1	7.3	18,701
Intensive Margin Only	7.7	25.5	52.8	73.4	14.1	1.4	32.2	23,483
Panel 8: Reporting Firms in Industries with Prior of High RPT Use								
All trades	5.0	27.4	79.4	99.4	20.8	3.8	36.9	7,745
IOT Requirement	4.6	25.6	72.4	94.4	21.3	2.5	23.2	5,347
AHS Requirement	1.6	15.9	63.2	94.4	27.4	2.4	13.5	4,198
Intensive Margin Only	10.5	42.1	93.6	100.0	13.2	5.3	38.0	5,347

NOTE.—This table reports the share of firms' sales sold to related parties. N shows all firm-years that have at least one related party. Due to missing firm-level information, N is slightly smaller than the 125,044 available firm-level reports. The weighted mean is weighted by the size of each reporting firm's total sales.

related-party transactions, and therefore the common proxies' limitations appear the strongest.

In Panel 4, the opposite case is tested where all unmatched related parties are assumed to satisfy the IOT cutoff. The results do not show a large difference from the main results. This suggests that even when the IOT cutoffs are applied just to the set of related parties with known industry information, the limitations are already strong. Here, note that both the intensive and extensive margins of the restrictions are affected. The intensive margin effect is smaller than in Table 1.1, as the IOT cutoff now generates false negatives. On the other hand, the extensive margin is smaller as well: now, with IOT and AHS requirements, a larger number of reporting firms have 'at least one tradeable related party'. Hence, the number of observations show less dramatic difference between the true observation and the simulated results.

However, regardless of the specification the key takeaways remain unchanged. First, descriptive statistics for firms' related-party trades remain unaffected and much larger than existing literature's estimates. Second, common restrictions on data consistently assert a substantial downward pressure on estimating the related-party sales, in both intensive and extensive margins.

The fourth robustness check tests how the treatment of the retail and wholesale industries affects the results. IOTs in most cases, including the one utilized in this study, do not define these industries finely enough.[31] For example, Korean IOT treats the entire retail and wholesale industries as one segment, while dividing the manufacturing industries into 234 sectors. Moreover, IOTs show the value of intermediates used to *produce* the retail and wholesale services, which does not accurately represent the flow of good and services that go through these industries.

Panel 5 of Table 1.2 treats all sales to related parties in the retail and wholesale industries

[31] Many previous studies that utilize IOTs recognize this issue as well (Fan and Lang, 2000; Acemoglu et al., 2009). AHS utilizes information from the Annual Wholesale Trade Survey and the Annual Retail Trade Survey to address this issue: similar data is unavailable for South Korea.

as satisfying the IOT criteria. Again, this is the most conservative measure that would make the impact of imposing IOT requirements the smallest. However, the results remain qualitatively unchanged. While the estimates with additional restrictions appear slightly larger, the differences are small.

The last robustness check in this section fully mirrors AHS in terms of which industries' reporting firms are used. AHS's sample is from the Commodity Flow Survey (CFS), which contains establishments in mining, manufacturing, wholesale, and select retail industries. As the main results in Table 1.1 limit to manufacturing industries only, or use all industries, in Panel 6 of Table 1.2 I report the results using the same set of industries as in CFS. Again, this produces only very minor differences in results.

1.4.2 Analysis of Distinct Sub-Segments of the Data

Panel 7 examines firms' purchases from their related parties, instead of the sales. In particular, this robustness check explores whether firms' reliance on related parties is different for input purchases compared to sales of output. For instance, while the median firm sells 2.7% of its total sales to related parties, it may be purchasing a much higher share of its needed inputs from them. I report Panel 7 using purchases from related parties as shares of the firms' cost of goods sold (COGS). The differences from the main result are minor: more than 75% of firms purchase from their related parties, and the limitations of proxy measures substantially underestimate the outcomes.

The final robustness check computes the share of sales for firms in industries where existing literature has underscored the importance of related-party transactions. Specifically, Panel 8 utilizes 15 4-digit industries that are reviewed in Lafontaine and Slade (2007).[32]

[32] The procedure is almost identical to a robustness check performed in AHS, with small differences in the specific industries considered. The included industries are coal mining, petroleum refining, footwear manufacturing, soft drink bottling, organic chemicals manufacturing, cement manufacturing, auto parts manufacturing, aircraft parts manufacturing, iron ore mining, pulp manufacturing, and ship building. The last three industries are not included in AHS due to CFS's scope or confidentiality. Removing them from consideration creates no qualitative change in the results, and in fact it makes the related-party trades appear slightly larger than is reported in Panel 8 of Table 1.2.

Reassuringly, the results confirm prior beliefs. The share of trades are larger at every reported percentile based on the alternative construction: the 75th and 90th percentile values are 27.4% and 79.4%, approximately 10 and 20 percentage points greater than the corresponding values in Table 1.1, respectively.

1.5 Conclusion

A major theoretical focus in vertical integration has been on its ability to facilitate the trade of goods and services along production chains. While the lack of related-party trade data has made direct observation or measurement of the trades difficult, the empirical literature has utilized proxy measures to infer them and has found trades within related parties to be surprisingly small.

I construct a novel firm-level data on Korean firms' related-party trades from their 10-K reports, and demonstrate for the first time the true size, direction, and prevalence of these trades. In contrast with the existing empirical literature, most firms appear to be engaged in related-party trades and the trades are shown to assume a substantially larger share of the firms' total sales and purchases.

A commonly used proxy for related-party trades, based on IOTs, appears to have caused much of this disparity between the theoretical and the empirical literatures. The traditional approach to creating a proxy only captures the tip of the iceberg, missing a much larger share of trades between related parties in seemingly unrelated industries. Out of all related entities that purchase from the reporting firms, only 14.4% are in industries that satisfy the common IOT requirements, and the sales to them represent only 31.7% of total related-party sales in the economy.

This strongly signals that (i) vertical integration involves active trading, and that (ii) related-party trades are utilized in contexts that are tailored to specific circumstances or needs of the firms, and do not simply follow the economy-wide trade patterns represented in input-output tables. Using IOTs to infer related-party trades from the network of related firms, it seems, is a perilous approach that should be utilized with caution.

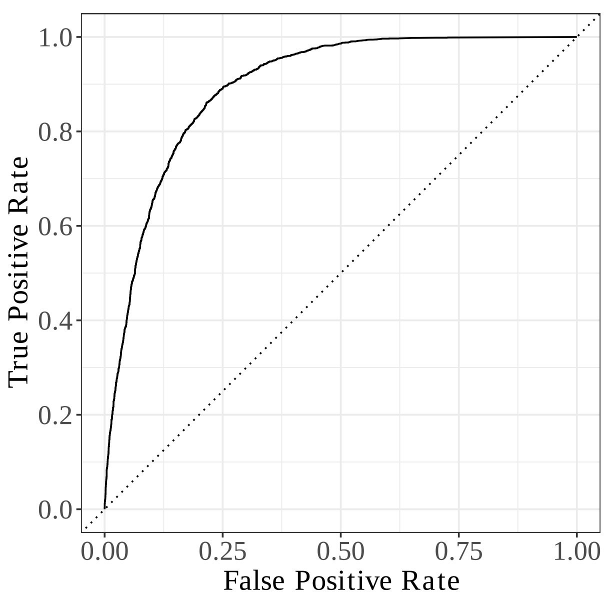
The usage of the proxy, however, was inevitable in many papers due to the lack of available data. Given this, what should researchers do when faced with this lack of data about the true state of the world? How can we better infer the existence and the magnitude of related-party trades from more commonly available data sources?

1.A Appendix: Tables and Figures

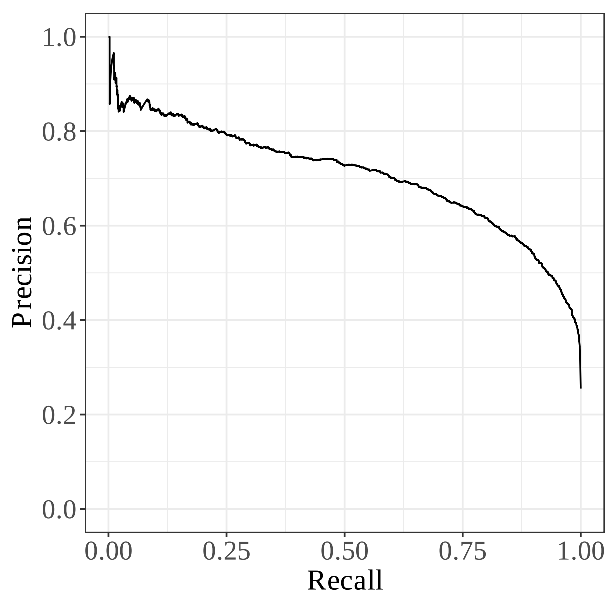
1.A.1 Figures

Figure 1.A.1: Performance of Algorithm based on 2019 Data

(a) ROC Curve



(b) PR Curve



1.A.2 Tables

Table 1.A.1: Share of Sales to Related Parties, Firms in All Industries

Related Party Share of Firm Total Sales	Percentiles (%) 50th	75th	90th	95th	Fraction (%) =0	≥1	Weighted Mean (%)	N
Panel 1: Main Result - Share at Reporting Firm Industry-Level								
All trades	1.3	15.1	62.4	95.8	30.3	3.0	24.0	121,519
IOT Requirement	0.1	3.5	24.4	52.5	36.7	0.7	9.7	52,298
AHS Requirement	0.0	1.9	19.1	50.5	42.7	0.8	6.4	35,955
Intensive Margin Only	4.8	30.2	86.3	100.0	19.3	4.7	6.8	52,696
Panel 2: All Firms with a Related Party in the ORBIS Database								
All trades	2.2	20.9	74.3	99.1	28.2	3.9	25.6	62,147
IOT Requirement	0.0	3.2	22.7	51.2	37.0	0.6	10.1	33,354
AHS Requirement	0.0	0.7	12.6	42.4	50.1	0.6	6.5	18,087
Intensive Margin Only	5.3	31.6	86.8	100.0	19.3	4.9	27.7	33,649
Panel 3: No Related Party without Industry Information Vertically Related								
All trades	1.3	15.1	62.4	95.8	30.3	3.0	24.0	121,519
IOT Requirement	0.0	0.3	10.9	36.1	66.3	0.5	5.3	52,761
AHS Requirement	0.0	0.4	11.9	40.8	63.8	0.6	3.7	36,197
Intensive Margin Only	4.8	30.2	86.3	100.0	19.3	4.7	6.8	52,696
Panel 4: All Related-Parties without Industry Information Vertically Related								
All trades	1.3	15.1	62.4	95.8	30.3	3.0	24.0	121,519
IOT Requirement	0.1	5.3	31.5	65.9	40.5	1.3	15.0	107,028
AHS Requirement	0.0	2.5	19.3	49.4	43.4	0.8	8.1	78,090
Intensive Margin Only	1.7	17.1	66.4	97.1	28.1	3.2	24.5	107,017
Panel 5: All Sales to RW industries as Vertically Related								
All trades	1.3	15.1	62.4	95.8	30.3	3.0	24.0	121,519
IOT Requirement	0.2	5.2	28.3	57.9	32.2	0.8	12.9	65,461
AHS Requirement	0.1	3.4	23.4	56.3	35.3	0.9	7.5	46,205
Intensive Margin Only	4.1	26.7	81.6	99.8	19.4	4.2	25.9	65,909
Panel 6: Reporting Firms in All CFS industries								
All trades	2.5	16.8	56.7	90.6	22.3	2.1	31.3	66,126
IOT Requirement	0.5	6.9	32.8	63.4	27.2	0.9	15.0	29,455
AHS Requirement	0.1	4.1	28.3	64.9	35.1	1.1	9.1	19,558
Intensive Margin Only	6.6	30.2	80.6	99.2	13.2	3.6	36.2	29,661
Panel 7: Related-party Purchases								
All trades	3.1	20.0	62.3	95.5	26.5	4.2	27.0	105,563
IOT Requirement	0.2	5.5	21.5	39.7	33.2	0.5	11.2	37,944
AHS Requirement	0.0	3.0	16.0	33.1	41.6	0.5	5.9	29,928
Intensive Margin Only	7.7	29.9	71.1	99.7	14.6	4.9	30.6	38,110
Panel 8: Reporting Firms in Industries with Prior of High RPT Use								
All trades	5.0	27.4	79.2	99.4	20.8	3.7	36.9	7,767
IOT Requirement	4.6	25.6	72.4	94.4	21.3	2.5	23.2	5,347
AHS Requirement	1.6	15.9	63.2	94.4	27.3	2.4	13.5	4,203
Intensive Margin Only	10.5	42.1	93.6	100.0	13.2	5.3	38.0	5,347

NOTE.—The table reports the share of firms' sales that is sold to related parties. Sample N consists of firm-years that have at least one related party, regardless of whether the reporting firm has any related-party transactions: due to missing firm-level information, total N used in the table is slightly smaller than the 125,044 available firm-level reports. The weighted mean is weighted by the size of each reporting firm's total sales.

Table 1.A.2: Comparison of Related-Parties with vs. without Industry Information

RPT	Info	$\text{Share}_{>0}$	q_{25}	q_{50}	q_{75}	q_{100}	Mean
Sale	Yes	0.642	0	20.0	548.0	247,063.0	3,211.6
	No	0.538	0	1.8	281.3	150,477.3	1,850.5
Purchase	Yes	0.614	0	15.1	540.5	190,664.2	2,788.5
	No	0.495	0	0	255.0	126,707.5	1,608.5

NOTE.—This table compares trades with related parties that are (i) successfully matched to industries and (ii) not, using pooled observation throughout 2013-2019. For each of the two groups, $\text{Share}_{>0}$ reports the share of firm-year-related party triples that have positive amount of each type of trade. q_n reports the size of the trades for the n-th percentile. As q_{25} is already at the smallest possible value of zero, q_0 is omitted from the table. Columns 4-8 uses Millions of Korean Wons as units. The table reports the numbers after truncating 0.5% of observations from each end.

1.B Data Appendix

1.B.1 Related-Party Trade Data

The related-party trades are web-scraped from side notes of firms' annual financial reports. This section describes the sample of firms included in this data and the scraping process.

Firms[33] in South Korea that satisfy a set of size thresholds are legally required to receive an annual external audit and publicly disclose the reports. The thresholds have undergone a small update during the data period. While the details of which are outlined in the table below, data from the only period affected by this change, fiscal year 2019, does not demonstrate a material difference from other years.

FY start date	Types of Firms	Criteria*
Before 2018.11.1	Corporations	Total Assets $\geq$ $10m
		Publicly traded or will be in the next FY
		Total Assets $\geq$ $5.8m *and* Total Debts $\geq$ $5.8m
		Total Assets $\geq$ $5.8m *and* Employees $\geq$ 300
On 2018.11.1 –	Corporations**	Total Assets $\geq$ $41.6m
		Total Sales $\geq$ $41.6m
		Satisfies 2 of $\begin{cases} \text{Total Assets} & \geq \$10\text{m} \\ \text{Total Debts} & \geq \$5.8\text{m} \\ \text{Total Sales} & \geq \$8.3\text{m} \end{cases}$

** All monetary units are converted to USD from the original Korean Won using a rough exchange rate of 1200 Won = 1 USD.*

*** Also includes limited companies from FY starting on 2019.11.1 and after.*

[33] More specifically, a type of firms, corporations, are subject to this reporting requirement during this paper's data period. As corporations are consistently more than 94% of all firms in Korea during the same period, I use the terms interchangeably.

1.B.2 Definition of Terms

1.B.2.1 Related Parties

On the 'Related-Party Trade' section of the side notes to the financial reports, firms disclose the names and relationships of the related parties that are involved in trades with the reporting firm. Specifically, the reporting firms are required to disclose relationships with the related parties in the following seven categories.

(1) Parent firm (holding a majority voting power)
(2) Parent firm (holding a minority voting power, or a joint parent firm)
(3) Subsidiary (holding a majority voting power)
(4) Affiliate (holding a majority voting power)
(5) Jointly owned subsidiary
(6) Board members of reporting firm or its parent firm
(7) All other related parties ('sibling'/'cousin' firms, etc.)

In practice, firms often report the relationships using much more finely defined categories. In that case, I re-categorize them into the seven categories.

1.B.2.2 Variables Reported ("Trades")

The firms are required to disclose the following categories of transactions with their related parties. From data, all transactions corresponding to sales, purchases, debts, and loans are taken and aggregated into the correct category. Note that the provision of debt guarantees or collateral are excluded from data used in the main analysis.

(1) Sales and purchase of goods (final and intermediate)
(2) Sales and purchase of real estate or other assets
(3) Sales and purchase of services
(4) Lease
(5) R & D
(6) License

(7) Loans, debts and other investment

(8) Provision of loan guarantee / collateral

(9) Uncompleted contract

(10) Provision of debt payment on behalf of the other

1.B.3 Matching Firm-level Information to Related-Party Trade Data

The related-party trade data from Korea provides a unique firm-level 10-digit identifier (Business Registration Number) for the reporting firm. Firm-level information is easily matched to the reporting firms, using the identifier, from three existing firm-level databases: KISVALUE, TS2000, and ORBIS. KISVALUE and TS2000 are the two largest firm-level databases in Korea. The two databases, when combined, contain all reporting firms in my related-party trade sample. The ORBIS database is compiled by Bureau Van Dijk, and includes a large subset from the related-party trade dataset as well as a larger number of firms outside of South Korea.

Values from the three databases are consistent but often display minor differences. In case of a difference in numerical accounting data (e.g., total sales, cost of goods sold, etc), I assign priority to the information from TS2000, KISVALUE, then ORBIS, following the number of reporting firms that can be matched with each database. When the databases report different industry codes for the firms, (i) the most detailed industry information is used, and (ii) if the industry codes display the same level of details or digits, the databases are given the same priority as the accounting data. Note that information from separate financial statements (at the individual firm level) is used instead of the consolidated financial statements, which report combined information of parent companies with their subsidiaries.

On the other hand, for the related parties that are reported, the only available information is their firm names. The related parties are then matched to data in existing databases by the names. The firm names are cleaned, then matched with 10-digit identifiers by (i) historical firm name data from the DART website as well as from existing domestic databases, then in case the firm names cannot be matched to the domestic 10-digit identifiers, it is then matched by (ii) firm names in the ORBIS database.

The name-matching process goes through extensive checks, and conservative criteria are used to check validity of matching in order to ensure accuracy. Here I summarize the key steps taken. The key is to use a dual-reporting feature: if firm A reports trades with firm B that is large enough, firm B would also be reporting trades with firm A in the same year. Then, the trades reported by firm A should be equal to the trade reported by firm B. In the following, firm A denotes the original reporting firm and firm B denotes one of A's related parties.

1. If a unique firm identifier is matched with firm B,
   (a) If there is an annual financial report from the matched firm identifier in the given year, the match is regarded correct only if B's financial report also lists firm A as its related party.
   (b) If there is no annual financial report from the matched firm identifier in the given year, then the match is assumed as a correct match.
2. If multiple firm identifiers are matched with firm B's name,
   (a) If there is a unique firm identifier that has an annual financial report in the same year that lists firm A as its related party, then that firm identifier is regarded as a correct match.
   (b) If there are multiple identifiers that have annual financial reports in the same year that lists firm A as its related party, or if there is no such 10-K report, then no firm identifier is deemed a correct match.
3. After (1) and (2), if firm B is not matched with an existing firm-identifier from Korea, firm B's name is checked against the ORBIS database. Information from the ORBIS database is deemed correct only if it is a unique match with a firm identifier.

After the matching process, 52.1% of the observations are matched with information on the related parties.

The related-party trade numbers go through a similar verification process. The process is largely in line with the data verification process outlined in the Data Appendix of Alfaro-Ureña et al. (2018).

1.B.4 Comparison of data with existing databases

This section provides a test of validity of the scraped related-party trade data's accuracy by comparing the results with two existing databases. In general related-party information has rarely been utilized as it is listed only in the side notes of financial statements. The rare exceptions are KISVALUE and TS2000, the two largest firm-level databases in Korea, which each contain a part of the data. Here, I show that comparison of the scraped data with existing databases yield almost identical results.

KISVALUE provides the related-party data aggregated at the reporting firm's level only. Even though trades with individual related parties are not visible, the database includes a large subset of firms in my related-party trade data: it includes 19,464 firms, compared to the 30,390 firms in my own. In Table 1.B.1 I report the comparison of related-party trades for the sample of firms that are included in KISVALUE and the scraped data at the same time. The results confirm accuracy of the scraping process. The difference between my data and KISVALUE is minuscule. Moreover, the related-party trades appear slightly larger in the existing database, suggesting that the results in Table 1.1 are not overestimating, if underestimating slightly.

It is worth noting that KISVALUE reports NA values for the firms that do not engage in any related-party trades, as well as for some firms that post zero trades. Therefore, by utilizing only the firms with non-NA related-party trade values in KISVALUE, as in Panel 2, the numbers may overrepresent the firms with some level of related-party transactions, be it sales, purchases, loans, or debts. In Panel 1, I mitigate the effect by including more firms. First, using my scraped data I identify the firms that have no related-party trades (but have all NA values in KISVALUE), then include them in KISVALUE samples as having zero related-party trades.

Table 1.B.1: Share of Sales to Related Parties, Comparison with KISVALUE

Related Party Share of Firm Total Sales	Percentiles (%)				Fraction (%)		Weighted Mean (%)	N
	50th	75th	90th	95th	=0	$\geq$1		
Panel 1: All firms with RPT in KISVALUE: Sales								
Scraped - All Industry	1.4	13.8	54.5	89.7	29.2	2.4	27.0	85,776
KISVALUE - All Industry	1.6	14.3	55.5	90.0	26.9	2.0	26.2	85,850
Scraped - Manufacturing	2.6	16.6	52.5	85.4	22.0	1.8	33.3	41,053
KISVALUE - Manufacturing	2.9	17.0	52.7	84.8	19.7	1.4	32.9	41,077
Panel 2: All firms with RPT in KISVALUE: Sales								
Scraped - All Industry	2.8	18.4	63.3	94.4	18.4	2.8	27.8	74,456
KISVALUE - All Industry	3.1	19.0	64.2	94.2	15.7	2.3	27.0	74,530
Scraped - Manufacturing	4.0	19.9	58.0	89.8	13.0	2.1	34.2	36,793
KISVALUE - Manufacturing	4.3	20.2	57.7	89.4	10.4	1.6	33.5	36,817
Panel 3: All firms with RPT in TS: Sales + Purchases								
Scraped - All Industry	13.5	40.8	81.6	101.0	4.8	5.6	48.9	11,768
TS2000 - All Industry	12.2	39.0	81.2	101.6	11.9	5.8	42.0	12,440
Scraped - Manufacturing	14.4	39.5	73.8	96.4	4.7	4.0	58.0	7,717
TS2000 - Manufacturing	13.1	37.8	70.9	94.6	12.5	3.8	56.1	7,957

NOTE.—The table compares the related-party trades reported in KISVALUE with the scraped and cleaned data used in this paper.

In contrast, TS2000 provides the data defined at the most similar levels of details to my own, but for much narrower scope of firms and with different variable definitions. The database reports the trades for 2,284 public firms in Korea, compared to the 30,390 firms used in the main paper. Moreover, the reporting firms in this sample are significantly larger: the median total sales of firms in this sample is roughly $61 million USD, while for the full sample it is $16 million USD.[34] Lastly, TS2000 divides the trades differently: the sales and purchases are divided into those related to the firm's core business operations, and those that are not (*operating* vs. *non-operating* income and cost), and *others*, a category assigned when the database could not determine whether

[34]The median total sales of reporting firms in KISVALUE's related-party trade sample is $27 million USD.

a reported transaction is about the core operation or not. The problem is, this *others* category was defined too liberally: the numbers are too large, and it does not distinguish sales from purchases, simply listing the sum of reported numbers.

Panel 3 of Table 1.B.1 compares the related-party trades reported by TS2000 with the scraped data. To account for the *others* category, here the numbers reported are the sum of sales and purchases, as a share of each firm's total sales.

1.B.5 Industries with Prior of High Related-Party Trades Reviewed in Lafontaine and Slade (2007)

Online appendix of Atalay et al. (2014) already provides a succinct overview of the industries and the individual papers that study them. Specific 4-digit KSIC industry codes that are used in this paper are listed here.

- Mining of coal and lignite (KSIC 0510)
- Manufacture of ice and non-alcoholic beverages; production of mineral waters (KSIC 1120)
- Manufacture of footwear (KSIC 1521)
- Petroleum refineries (KSIC 1921)
- Manufacture of basic organic chemicals (KSIC 2011)
- Manufacture of cement, lime and plaster (KSIC 2331)
- Manufacture of parts and accessories for motor engines (new products) (KSIC 3031)
- Manufacture of parts and accessories for motor vehicle body (new products) (KSIC 3032)
- Manufacture of power transmission devices and electrical and electronic equipment for motor vehicles (new products) (KSIC 3033)

- Manufacture of other parts and accessories for motor vehicles (new products) (KSIC 3039)
- Manufacture of parts and accessories for motor vehicles (remanufacturing products) (KSIC 3040)
- Manufacture of engines and parts for aircraft (KSIC 3132)
- Mining of iron ores (KSIC 0610)
- Manufacture of pulp (KSIC 1711)
- Building of ships and floating structures (KSIC 3111)

CHAPTER 2

Inference of Vertical Relatedness Through Machine Learning

2.1 Introduction

How firms transact within the boundaries of the firm has been a focus of vast literatures. The intra-firm transactions help answer a range of fundamental questions in fields such as industrial organizations, international trades, as well as corporate finance. The transactions show which parts of production chains have been integrated into a firm, allowing studies into the reasons and the subsequent effects of such integration. They are also critical in studying how multinational corporations utilize Foreign Direct Investments (FDI) in their supply chains. Moreover, the intra-firm trades enable a distinction of whether corporate structures and controls are utilized to enhance efficiency, or instead promote private gains of those in control.

However, the paucity of data has prevented researchers from obtaining a clear picture of the intra-group transactions. Firm-level databases predominantly reflect firms' total economic activities, and do not show trades with individual trade partners. In the rare cases where entity-to-entity transactions are observed, whether the trade partners exist within the boundaries of the firm is not observed.

Literature has therefore utilized related information to infer, or proxy for, whether two related parties trade. The methods utilize a combination of (i) industry-level trade patterns represented by Input-Output Tables (IOT) and (ii) entity ownership; see, e.g., Fan and Lang (2000), Alfaro et al. (2019), Atalay et al. (2019), Acemoglu et al. (2009), and Aghion et al. (2006).

At the same time, a rare literature utilizing granular intra-firm trade data has warned that such proxies may be inaccurate. Ramondo et al. (2016), using U.S. international corporations' trades with their foreign subsidiaries, that while the IOT-based proxies are positively correlated with having a subsidiary in an industry, they are not related to intra-group flow of goods. Hong (2022) shows, through utilizing extensive South Korean firm-level data, demonstrates that the IOT-based proxies are inaccurate. The paper shows that the proxies with a 1% cutoff leads to a large underestimation of intra-group trades in Korea, capturing only 17.6% of related parties that trade and 32.6% of their sales volume.

In this vein, I propose alternative proxy measures that provide a more accurate inference of related-party trades. Here, I exploit the unique opportunity to utilize the true data on related-party trades from South Korean firms (2013–2019) to test and compare each method's predictive performance. Using supervised machine learning, I build a number of prediction mechanisms that enable researchers with only widely available information on firms—sizes, countries and industries—and the ownership links between them to infer the existence and magnitude of related-party trades. Despite requiring none to only a small amount of additional information other than the IOT coefficients, the proposed measures greatly improve performance metrics such as accuracy, precision, and recall.

Moreover, the new measures imply that a set of novel factors may be strongly correlated with whether firms engage in intra-group trades. Variables such as the size of the group . The size of the firms were also shown to be prominent factors, as was shown in Ramondo et al. (2016) and Atalay et al. (2014). The result further implies that proxy-measurements based solely on industry information are inherently prone to mismeasurements, and points to new directions of research.

It should be made clear, however, that the proposed measure is a first step toward a correct direction, and is not perfected at this moment. This paper, utilizing a various performance tests, makes a simple and reasonable argument that the method is substantially better than the IOT-based ones: the major reasons being that it utilizes various firm and group-level information, as well as being optimized on out-of-sample performance. However, The algorithms are currently only based on observations from South Korea and therefore are not free from questions on its external validity towards data from other economies. Moreover, the random forests method, even though it has been shown to perform strongly in various domains, does not fully utilize the network-based nature of the data. As researchers construct more intra-group transaction data, this paper's algorithms can be extended to integrate them and methods refined in the process.

2.2 Alternative Proxies from Supervised Machine Learning

In this section, I propose an alternative proxy by utilizing supervised machine learning. This is the first such measure that is optimized based on how it performs to predict actual intra-party trades. Hence it is able to (i) move away from the arbitrary cutoffs used in the previous measures, and (ii) utilize a wider range of information. Here, the goal is to come up with an algorithm that only requires *commonly available data* to produce predictions on which pairs of related parties are trading and which are not, so that it can be used in more standard data environments.

2.2.1 Data and Methodology

Supervised machine learning offers a number of advantages for this task. First, it is best suited to out-of-sample predictions, compared to the more traditional econometric toolbox that focuses on in-sample fit of a model. As the desired end-product is an algorithm that researchers can apply to their own dataset, I opt for machine learning to optimize performance and avoid overfitting to my own data. Second, the models are flexible as they allow for a large number of predictors as well as non-monotonic relationships between the outcome and the predictor (Athey and Imbens, 2019). Lastly, when the set of potential predictors is large and the key predictors are not clearly identified, researchers can establish the relative importance of each predictor (Breiman, 2001).[1]

Specifically, I apply the random forests approach to generate predictive algorithms classifying pairs of related parties as either trading or not trading. As the approach is widely in use, I only provide a brief description. The random forests approach estimates an individual decision tree by sequentially splitting the data based on optimized cutoffs of the most informative predictors. To illustrate a simple example, if the IOT coefficient is the only predictor, a tree would find a cutoff (or cutoffs) of the coefficient that best divides the data into groups of trading pairs and non-trading pairs. In practice, each tree is drawn from many predictors and cutoffs. A random forest aggregates a multitude of decision trees that are created by bootstrapping different subsets of both the data and the predictors. This aggregation aims to address the limitations of relying on a single tree, such as the volatility of the results and an overdependence on the variables used in early splits.

In order to form a prediction on which pairs of related parties are trading and which are not, I use a variety of possible predictors—referred to as 'features' in the machine learning literature—derived from related literature. The predictors include a range of basic industry, firm, and group-level characteristics; for a complete list, see Table 2.2.1.

[1]Machine learning is increasingly being used by economists in the academic domain. See, for example, Kleinberg et al. (2018), Fuster et al. (2021), and Li et al. (2021). Athey and Imbens (2019) and Mullainathan and Spiess (2017) provide excellent reviews of machine learning applications for economists. Efron (2020) presents a succinct comparison between predictive algorithms and standard regression techniques.

Table 2.2.1: List of Predictors and Related Literature

Category	Predictors
1. IOT coefficients	Direct and total requirements, share of sales to RP industry (Atalay et al., 2019; Alfaro et al., 2019)
2. Firms' basic information	Size (assets, sales) of firms (Ramondo et al., 2016)
3. Group's basic information	Size of group (number of firms) (Ramondo et al., 2016)
4. Industry contractibility	Index derived from Rauch (1999) (Rauch, 1999; Nunn, 2007)
5. Control over RP	Dummy for whether RP is a subsidiary (Antràs and Chor, 2013)
6. Location of RP	Dummy for whether RP is a domestic firm (Antras and Foley, 2015)
7. Industries of firms	2-digit KSIC (3-digit NAICS) codes (Lafontaine and Slade (2007), AHS)

NOTE.—This table provides the list of predictors used in Section 2.2, and examples of the papers that discuss relevance between the predictor and related-party trades.

The first set of features involves both parties' industry-specific characteristics, including (i) IOT coefficients, (ii) a measure of industry contractibility, and (iii) industry dummies. IOT coefficients reflect technologically determined intermediate input needs for each industry, and have been most widely used to infer related-party trades. I include three different types of coefficients that have each been utilized in the literature: the direct requirements coefficients, total requirements coefficients, and the share of intermediate sales directed to a specific industry.[2] Also, I include a measure of how contractible an industry's output is—or more specifically, how relationship-specific it is—to account for potential holdup problems.[3] Past studies have also pointed to specific industries as more likely participants in intra-party trades. To account for this, in some specifications I also include dummy variables indicating the industry codes of both firms.[4]

The second set of predictors measures the size of firms and groups. Both AHS and Ramondo et al. (2016) report that a firm is more likely to engage in related-party trades when that firm is larger and when it belongs to a larger group. I use firms' total assets and total sales, and the number of firms within a group as possible predictors. The relatively parsimonious nature of the variables in this set have an added benefit for future use, as total assets and sales are often the most widely available information in firm-level databases.[5]

---

[2]Direct requirements show the share of industry i's intermediate inputs that come directly from another industry j, while total requirements represent both direct and indirect inputs from j. The last coefficient, the share of industry i' intermediate sales directed to j, is utilized in AHS.

[3]I use a measure of relationship-specificity developed and used in Rauch (1999). This method classifies commodities by whether they are sold on organized exchanges, have a reference price in a trade publication, or neither. By reflecting the depth of the potential market for the commodity, the degree of potential holdup problems is inferred. Following Nunn (2007), I create a dummy variable indicating whether a good falls into the first two categories. Rauch's classification that was revised in 2007 groups goods into 1,189 SITC Rev.2 industries. I use official concordance tables to match them with appropriate KSIC or NAICS codes of each firm.

[4]I use 3-digit NAICS and 2-digit KSIC codes in order to maintain a similar level of disaggregation.

[5]Total assets and sales are matched to firms from existing databases, following the process described in Hong (2022). However, not all related parties in the data are successfully matched with assets and sales information. These cases are primarily driven by firms that are too small and thus not included in the databases at hand. Therefore, missing values are assigned 0, and a dummy variable indicating imputation is added.

The last set of predictors is loosely defined as a firm's 'control' over the related-party in trade. I include an indicator of whether the reporting firm controls a *majority* of the related party's voting power; in other words, whether the related party is a subsidiary of the firm. In contrast, control over the firm's affiliates with *minority* vote-holding, or other related parties such as parent companies, parent companies' other subsidiaries, etc. are deemed weaker and are therefore distinguished with this dummy variable. In a similar vein, I also include an indicator showing whether the related party is domestic or foreign. This partly reflects the firm's possible control and supervision over the affiliate's activities (Antras and Foley, 2015); at the same time, it is expected to pick up differences in domestic vertical ownership and FDI.

The related-party trade dataset for South Korea is scraped from the firms' financial statements, and shows the intra-firm sales and purchases. The construction of this data, as well as the detailed definitions, are as defined in Hong (2022).

For all publicly traded firms in Korea, TS2000 reports the list of their related parties. I merge the primary related-party trade dataset with the list from TS2000, keeping only those firms that appear in both datasets. The consolidated data contains the complete list of all public firms' related parties, as well as how much each firm trades with them. As a result, I utilize a total of 2,607 firms over 2013–2019 that have 420,428 firm-year-related party triples. Among the triples, 105,367 (25.1%) report transactions in terms of either sales, purchases, loans or debts, while the rest do not trade with the reporting firms. The main body of this section will primarily report the algorithms trained with the firms in manufacturing sector only. This leaves 1,600 firms over the same data period, with 197,021 firm-year-related party triples. Among the triples, 56,158 (28.5%) report intra-party transactions.

The data for each year is then randomly divided into training and testing sets according to an 80:20 split at the firm-level. The principle is to have no overlapping information between the two sets: thus, no reporting firm will appear in the same year's training set *and* testing set. I train each year's prediction algorithm separately using the `R` package

Table 2.2.2: Confusion Matrix: Predict Related-Party Trades Based on IOT

(a) IOT Cutoff ≥ 0.01

		Reference	
		Trade	No Trade
Prediction	Trade	6,058	7,115
	No Trade	16,796	52,277

(b) Total Requirements > 0

		Reference	
		Trade	No Trade
Prediction	Trade	22,854	59,392
	No Trade	0	0

NOTE.—Both tables represent confusion matrices of utilizing only a single cutoff of the chosen IOT coefficients to discern the pairs of related parties that trade, from those that do not. *Prediction* denotes whether the method in question predicts the pairs to trade, and *Reference* denotes whether the pairs actually trade. Table (a) shows the results of predictions based on the cutoff of 1% of the seller industry's intermediate sale going to the buyer industry. Table (b) shows the performances of using the total requirements coefficient having a strictly positive value. The data spans all publicly traded manufacturing firms in Korea, over 2013-2019, and their related parties.

`caret`.

2.2.2 Prediction Results

Several key measures are utilized to assess how well a prediction is executed. As an illustrative example, I show classification performances of two widely used methods from the literature that each rely on a single IOT coefficient cutoff.

Table 2.2.2a presents the confusion matrix for predictions of related-party trades using 1% of the IOT coefficient as the cutoff.[6] At first glance the method appears to perform relatively well; it classifies 70.9% of the related party pairs correctly.

However, this simple *accuracy* measure masks underlying problems. The correctly classified observations are composed primarily of the pairs that do not trade, while not many of the trading pairs are picked up. Out of the related-party pairs that trade, this method detects only 6,058 (26.5%); in other words, the *recall* rate is low.[7] Moreover, the group of pairs predicted to trade is instead composed of more pairs that actually do not. Only 6,058 (46.0%) of the pairs predicted to be trading are true positives; therefore, the *precision*

[6] A confusion matrix divides pairs of related parties in the data according to how well a method predicts whether the pair trades.

[7] Recall = True Positives / (True Positives + False Negatives)

Table 2.2.3: Out-of-Sample Confusion Matrix, ML Algorithm from 2019 Data

		Reference	
		Trade	No Trade
Prediction	Trade	440	178
	No Trade	249	1,270

rate of this method is low as well.[8] In summary, attempts to study vertically trading entities through this proxy measure would not only focus on a small portion of those that are actually trading, but the constituted sample would in fact consist of more non-trading entity pairs than trading pairs.

Table 2.2.2b, using another proxy from the literature, displays a different problem. When we deem the pairs with a positive total requirements coefficient to be vertically integrated, the cutoff proves to be too lenient, not categorizing any firm-related party pairs as non-trading.[9] As such, while it detects all (27.8%) observations that are actually trading and records 100% recall, both its accuracy and precision are very low at 27.8%.

In comparison, supervised machine learning provides much more accurate and consistent predictions. Table 2.2.3 reports the performance of an algorithm trained with data from the most recent year, 2019, when applied to the same year's out-of-sample testing set. The algorithm detects 62.7% of the actual trading pairs, a substantial jump from 26.5% in Table 2.2.2a. Moreover, among the pairs predicted to be trading, 71.1% are in fact doing so, again showing a sizable increase from 46.0% and 27.8% based on the more traditional measures. This result is all the more notable considering that the model requires only a small amount of additional information over the traditional measures.

The improved performance is not limited to 2019, but is consistent throughout all years in data. Table 2.2.4 shows the out-of-sample performance of algorithms built with each

[8]Precision = True Positives / (True Positives + False Positives)

[9]This threshold has proven to exclude only minimal fraction of observations in other datasets as well. Alfaro et al. (2019) reports that in the WorldBase dataset that they utilize, 98.0% of the related parties of parent firms satisfy the total requirements criteria.

Table 2.2.4: Prediction Performance Metrics: 2013-2019 Algorithms

Year (1)	Accuracy (2)	Precision (3)	Recall (4)	Specificity (5)	AUC (6)	PR-AUC (7)
2019	0.800	0.712	0.639	0.877	0.861	0.708
2018	0.816	0.686	0.658	0.879	0.888	0.719
2017	0.770	0.678	0.547	0.876	0.848	0.701
2016	0.757	0.658	0.586	0.844	0.830	0.700
2015	0.703	0.657	0.593	0.781	0.778	0.712
2014	0.820	0.672	0.660	0.880	0.881	0.737
2013	0.757	0.648	0.599	0.837	0.826	0.730

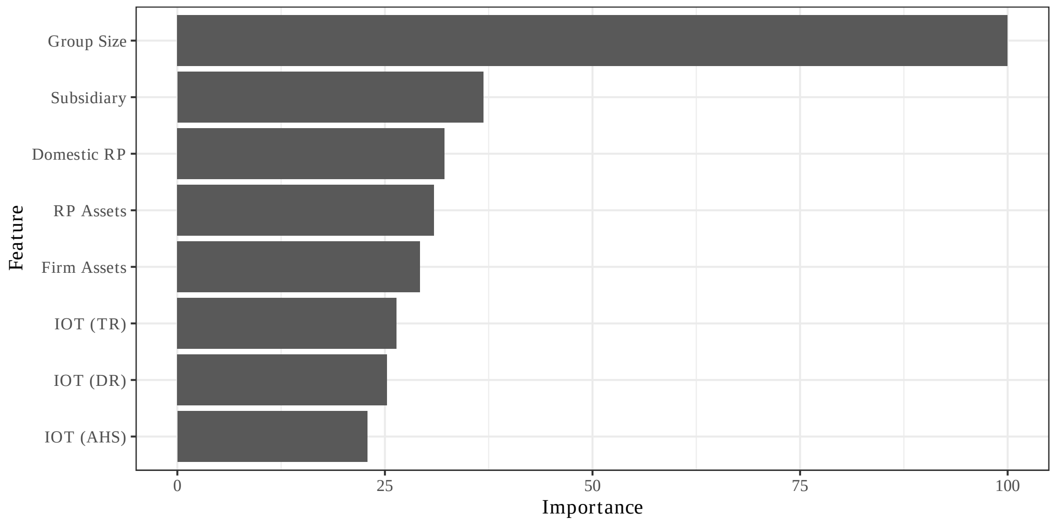
NOTE.—This table reports the prediction performance metrics of algorithms created from each year's training data, tested on the same year's out-of-sample testing dataset. Information from all manufacturing public firms in Korea and their related parties are utilized.

year's data. Here, a separate algorithm is trained each year with how 80% of firms in the year's data trade with their related parties, then is tested on the remaining 20%. In accuracy, precision, and recall, the performance gains over the traditional cutoffs are strong and consistent. Specificity, which measures how well the mechanism detects the non-trading pairs, is consistently strong as well.[10]

What is more, the algorithms fare well in other key metrics that are widely used in the prediction literature. Appendix Figure 2.A.1a plots the Receiver Operating Characteristic (ROC) curve from 2019. This curve plots the tradeoffs between true positive rates ($recall$) and false positive rates ($1 - specificity$) as the threshold becomes more relaxed for declaring a pair to be trading. As the method becomes more lenient in declaring a pair to be trading, any classification approach detects more true positives and produces higher recall. At the same time, it is more likely to misconstrue non-trading pairs as trading, and have higher false positive rates. Encapsulating this curve, the AUC score calculates the *Area Under the (ROC) Curve* which provides a measure of the estimated probability that a positive case, in this case a trading pair, will be ranked higher by the algorithm than a negative case (Hosmer Jr et al., 2013). As column (6) of Table 2.2.4 reports, the AUC scores consistently report a strong result.

[10]Specificity = True Negatives / (True Negatives + False Positives)

Figure 2.2.1: Importance of Top Variables, 2019 Algorithm



When the positive cases are scarce such that there is a large class imbalance, AUC scores could be overly optimistic (Davis and Goadrich, 2006). In this case, the PR-AUC scores are often used to evaluate the model performances. This score calculates the area under Precision-Recall curves, plotted in Appendix Figure 2.A.1b, which shows the tradeoff between *precision* and *recall* as the predictive threshold changes. While the class skew is not strong in this paper, with 27.8% trading, that PR-AUC scores are consistently strong sufficiently addresses any possible concerns.

In addition to reporting algorithm performance, I report the top 8 predictors in terms of variable importance in Figure 2.2.1. The importance is calculated based on how information from each variable decreases mean node impurity: this is closely analogous to the residual sum of squares in regression. The most important variable is given the index value of 100, while other variables are given values in relative terms to it. The most notable trait from the variable importance plot is the large importance of group size. While Ramondo et al. (2016) has shown that group size could be a significant predictor of related-party trades, that its relative importance far surpasses the other features is surprising.

Moreover, the IOT coefficients, although important, nevertheless appear far from the most decisive predictors. Other variables that represent reporting firms' control over the related party are shown to be more important, such as whether a majority of the related party's voting rights are owned by the firm (*Subsdidiary*) or whether the related party is located domestically. Then follow the sizes of both firms in terms of total assets, and only lastly the input-output coefficients.

2.3 Conclusion

Measuring intra-firm transactions has often been an integral part of answering fundamental economic questions. To study issues such as vertical integration, global supply chains within multinational corporations, and the relationship between corporate ownership structures and efficiencies, often the researchers are required to know which related parties buy and sell from each other.

Such importance of the measure, coupled with the lack of actual data, has long compelled researchers to rely on a set of proxy measures despite the amounting evidence against their accuracy. Previous works show that the IOT-based proxy measures are not significantly correlated with existence of related-party trades (Ramondo et al., 2016), as well as that a common form of the measure is prone to a large underestimation of the prevalence and size of the trades (Hong, 2022). However, there has been no viable alternative to the proxy measures.

To this end, I utilize random forest method to construct predictive algorithms that researchers can apply to their own datasets. While the method requires only a small amount of additional publicly available data, it shows a marked improvement in performance. Moreover, I show that variables such as group and firm sizes, as well as the firm's control over the related party, are important predictors of whether two related entities are trading.

I emphasize that the proposed method is only a first step in a better direction, rather than a perfected measure. It is a demonstration of what can be achieved by leveraging actual data even with relatively simple methods, and may benefit from integrating new data and methods. For example, while its strong out-of-sample prediction performance is documented in various ways, the lack of data currently prohibits verification of its performance across a diverse array of countries and contexts. By expanding the related-party trade data that this algorithm relies on, it can be updated to account for possible country-specific nuances. This a highly achievable goal as numerous countries share the accounting requirements that this paper exploits. Moreover, the methods can benefit from accounting for the network-based nature of the data, which the random forests method does not allow for.

This paper indicates that widely used proxy measures appear to have caused notable biases in our perception of the size and prevalence of vertical trades, and presents possible alternatives to the proxies. Then, a natural question follows: what other aspects of vertical integration could be better understood by this improvement in the long-standing measurement problem? I intend to pursue this in separate papers.

2.A Appendix: Tables and Figures

Figure 2.A.1: Performance of Algorithm based on 2019 Data

(a) ROC Curve (b) PR Curve

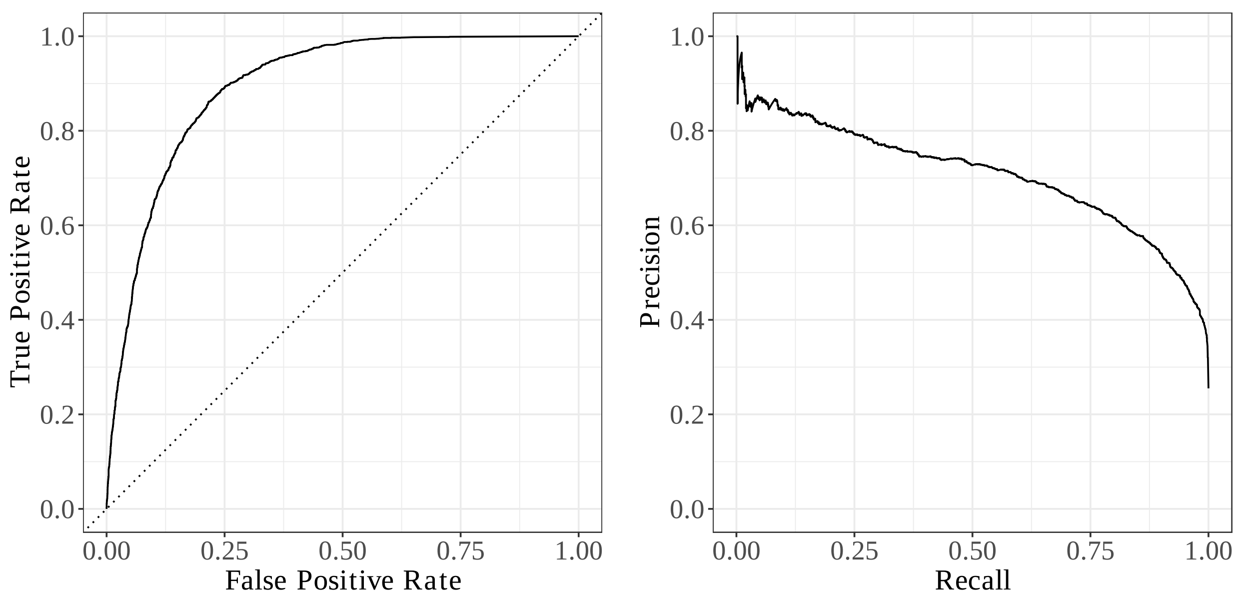


Table 2.A.1: Comparison of Related-Parties with vs. without Industry Information

RPT	Info	$\text{Share}_{>0}$	q_{25}	q_{50}	q_{75}	q_{100}	Mean
Sale	Yes	0.642	0	20.0	548.0	247,063.0	3,211.6
	No	0.538	0	1.8	281.3	150,477.3	1,850.5
Purchase	Yes	0.614	0	15.1	540.5	190,664.2	2,788.5
	No	0.495	0	0	255.0	126,707.5	1,608.5

NOTE.—This table compares trades with related parties that are (i) successfully matched to industries and (ii) not, using pooled observation throughout 2013-2019. For each of the two groups, $\text{Share}_{>0}$ reports the share of firm-year-related party triples that have positive amount of each type of trade. q_n reports the size of the trades for the n-th percentile. As q_{25} is already at the smallest possible value of zero, q_0 is omitted from the table. Columns 4-8 uses Millions of Korean Wons as units. The table reports the numbers after truncating 0.5% of observations from each end.

Table 2.A.2: Performance: Predict 2019 Intra-party Trades

Year (1)	Accuracy (2)	Precision (3)	Recall (4)	Specificity (5)
2019	0.797	0.711	0.627	0.878
2018	0.906	0.828	0.807	0.940
2017	0.883	0.806	0.722	0.939
2016	0.859	0.758	0.687	0.921
2015	0.846	0.740	0.633	0.922
2014	0.825	0.709	0.602	0.908
2013	0.825	0.709	0.552	0.921

NOTE.—Metrics for 2013–2018 represent the performances of each year's algorithms in predicting intra-party trades of 2019. In doing so, I utilize all firms in 2019 that have not been used in the training sets of each year: in this way, there are no overlaps between training and testing datasets. The metrics for 2019 represent the out-of-sample performances of 2019 algorithm.

Table 2.A.3: Performance: Predict 2013 Intra-party Trades

Year (1)	Accuracy (2)	Precision (3)	Recall (4)	Specificity (5)
2019	0.825	0.692	0.708	0.873
2018	0.825	0.696	0.733	0.864
2017	0.845	0.712	0.756	0.880
2016	0.842	0.723	0.763	0.876
2015	0.871	0.776	0.774	0.910
2014	0.887	0.801	0.830	0.911
2013	0.768	0.668	0.604	0.850

NOTE.—Metrics for 2013–2018 represent the performances of each year's algorithms in predicting intra-party trades of 2019. In doing so, I utilize all firms in 2019 that have not been used in the training sets of each year: in this way, there are no overlaps between training and testing datasets. The metrics for 2019 represent the out-of-sample performances of 2019 algorithm.

Table 2.A.4: Out-of-Sample Confusion Matrix, ML Algorithm from 2019 Data

		Reference	
		Trade	No Trade
Prediction	Trade	432	176
	No Trade	257	1,272

Table 2.A.5: Prediction Performance Metrics: 2013-2019 Algorithms (target: AUC)

Year (1)	Accuracy (2)	Precision (3)	Recall (4)	Specificity (5)	AUC (6)	PR-AUC (7)
2019	0.797	0.711	0.627	0.878	0.861	0.708
2018	0.816	0.694	0.639	0.887	0.889	0.719
2017	0.770	0.678	0.547	0.876	0.848	0.699
2016	0.755	0.663	0.564	0.853	0.829	0.695
2015	0.705	0.660	0.595	0.783	0.779	0.712
2014	0.820	0.672	0.660	0.880	0.881	0.734
2013	0.768	0.668	0.604	0.850	0.829	0.730

NOTE.—This table reports prediction performance metrics of algorithms created from each year's training data, tested on the same year's out-of-sample testing dataset. To create this table, algorithms that optimize the ROC AUC scores were utilized, while 2.2.4 uses algorithms that optimize PR-AUC scores. Information from all manufacturing public firms in Korea and their related parties are utilized.

CHAPTER 3

Not Me, Us: How Firms Allocate Intra-Group Loans

3.1 Introduction

Empirical evidence is inconclusive on whether transactions within business groups benefit the entire group. One strand of the literature emphasizes the negative side, such as tunneling, where controlling shareholders of a group abuse their control to reallocate profit to firms where they hold higher rights to cash flow, earning higher private profits at the expense of hurting minority shareholders (Johnson et al. 2000; Bernard et al. 2010). On the other hand, some have argued that the benefits of intra-group transactions accrue to the entire group. Such transactions, the studies argue, enable groups to source inputs more efficiently (Williamson 1979; Klein et al. 1978) as well as reallocate capital more efficiently across member firms (Gopalan et al. 2007; Almeida et al. 2015).

The lack of consensus is, in part, driven by a long-standing lack of detailed data. Granular data on the inner-workings of business groups are hard to obtain. Therefore, researchers often have to rely on a small number of firms that are typically on the largest end of the size distribution or utilize data that is aggregated to the group or firm-level, instead of utilizing the full variations in firm-to-firm transactions. Moreover, most existing datasets are able to only visualize specific subsets of intra-group transactions, such as loans or equity transfers, instead of being able to analyze how each type of transaction interacts with the others.

This paper contributes to the literature by constructing an extensive and granular dataset of intra-group transactions for a large sample of firms representative of an entire economy, South Korea. Utilizing a local accounting requirement, the constructed dataset is able to provide information on intra-group firm-to-firm loans and debts, as well as sales and purchases, for more than 34,000 firms across the span of 9 years (2013–2021). To the best of my knowledge, this is the only dataset with such comprehensive scope of transactions and firms at the same time.

Utilizing the data, this paper suggests that, at least in the recent South Korean context, intra-group loans were driven primarily by firms' efforts to efficiently reallocate resources. Firms do not appear to provide loans more heavily to the firms with higher group-owner cash-flow rights (hereinafter CFR), contrary to what the tunneling hypothesis predicts. Instead, the intra-group flow of capital seems to be much more concretely influenced by concerns over efficiency, favoring individual firms with higher profitability, higher capital intensity, and lower short-term liquidity. This finding closely echos the literature that finds, in different contexts, that intra-group reallocation of resources seems to have boosted efficiency.

However, through which mechanisms can the group find the reallocation of capital through intra-group loans beneficial, especially more than delegating the reallocation to the external finance sector? This paper provides empirical evidence for two possible explanations: spillover concerns and supply chain management. Specifically, I show that such transactions are carried out more intensively between firms close to each other in either the ownership chains or supply chains. First, the relationship between proximity within ownership chains and intra-group loans supports and strengthens the idea from previous literature (Gopalan et al., 2007) that a major factor behind intra-group loans is the concern over a member firm's performances spilling over to other members. Second, the patterns I document about supply chains provide early-stage evidence showing that groups utilize loans to assist in extending the intra-group supply chains, often further downstream.

Lastly, this paper concludes by constructing new stylized facts on a novel component of intra-group loans: trade credit. Literature has delved into determinants of trade credits

(Antras and Foley, 2015; Klapper et al., 2012) as well as how firms' use of credits changes when faced with a short-term economic shock (Love et al., 2007; Garcia-Appendini and Montoriol-Garriga, 2013). However, how *business groups* utilize this particular type of short-term loans differently and how they react to short-term shocks in this context have not been studied thoroughly. As a starting point in filling the gap, I document that in intra-group trades, firms utilize a significantly smaller amount of trade credits compared to arm's length trades. Moreover, I detail preliminary evidence showing that intra-group use of trade credits contracted sharply during the short-term economic contraction from the COVID-19 pandemic, in a direction opposite from arm's length trades.

This paper extends the small number of papers that use granular data to study intra-group loans, which largely utilize information from notes to financial statements in different economies. Gopalan et al. (2007) use loans within Indian business groups to argue that the loans benefit the entire group by deterring defaults of financially weaker members, thereby preventing negative spillovers. Santioni et al. (2020) focuses on the interplay between intra-group loans and external debts for Italian firms. In a unique example, Hong (2022) uses the intra-group *sales and purchases* information for a larger sample of firms in South Korea. However, none have been able to visualize both the intermediate goods trade and the loans simultaneously.

The closest research to this paper is Buchuk et al. (2014), which uses intra-group loans data in side notes to Chilean firms' financial statements. They find that while firms with high CFR do receive more loans, there is no evidence that minority shareholders lose out from such transactions. The first part of this paper reproduces Buchuk et al. (2014)'s main analysis with much more comprehensive and granular data, reducing the analysis's reliance on a small subset of firms that are also the largest firms in an economy. Also, by including the intra-group trade data, this paper is able to analyze how firms' incentives to use loans to protect and foster intra-group supply chains.

The paper proceeds as follows. Section 3.2 describes the data and its construction. Section 3.3 discusses the empirical specification and results on the factors that influence the flow of intra-group loans. 3.4 documents new stylized facts on intra-group trade credits. 3.5 concludes.

3.2 Data

3.2.1 Annual Intra-Group Transactions

South Korean corporations are required to report transactions with related parties on the side notes to their financial statements if the firms satisfy a set of size thresholds.[1] Moreover, the financial reports are made available on a website maintained by *Financial Supervisory Service*, a South Korean government agency. I web-scraped the information from the notes to financial statements and cleaned the text to create the intra-group transactions dataset.

The intra-group transactions are reported separately for each of the firm's related parties. In practice, many firms report the transactions in detailed subcategories. I re-categorize the more granular transaction information into the six categories that are clearly identifiable for all firms: sales, purchases, accounts receivables, accounts payables, loans, and debts.

Table 3.2.1 shows a sample of how information is presented in the corresponding section of firms' financial reports. In the table, the reporting firm reports, for each related party that it had any transaction with, how much sales, purchases, loans, and debts it had with each of the related parties during the fiscal period. The table comes in many shapes in reality, which I clean to construct the final dataset.

The requirement to report intra-group transactions is not unique to South Korea. For example, this requirement is included in the International Financing Reporting Standards, an accounting standard that has been adopted by more than 100 countries.

[1] From the fiscal year 2020 onward, the firms in my sample satisfy at least two of the next four criteria in the previous fiscal year (converted to USD at 1200 KRW = 1 USD): (1) total assets $\geq$ 10 million USD, (2) total debts $\geq$ 5.8 million USD, (3) total sales $\geq$ 8 million USD, (4) number of employees $\geq$ 100. The thresholds in previous fiscal years only show small differences.

Table 3.2.1: Sample Report on Related-Party Trades

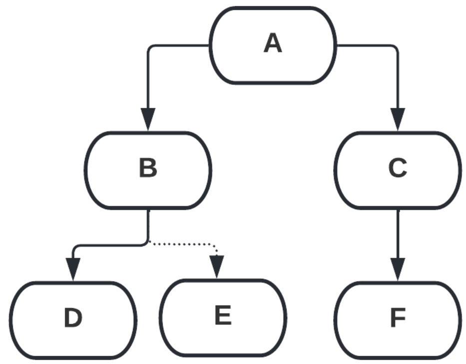
					Loans		Debts
Related Party	Relationship	Sales	Purchases	A/R	Other Loans	A/P	Other Debts
A Noodle	Parent	-	-	-	-	-	-
B Dairy	Subsidiary	-	-	-	-	-	-
C Mini Truck	Affiliate	-	-	-	-	-	-

NOTE.—The table shows an example of how the data is reported in the side notes to Korean firms' financial statements. A/R refers to Accounts Receivables, and A/P refers to Accounts Payables.

However, the information has not received strong attention from academia. This largely stems from two facts: (1) the information is only in the side notes, and not the main section, of the financial reports, and (2) it is written out in texts in various formats and is therefore difficult to transform into a uniform dataset. In fact, the information has mostly been utilized in very small samples that were hand-collected, and has chiefly been utilized in the corporate finance literature, where researchers are more accustomed to the nooks and crannies of financial statements. Buchuk et al. (2014) utilizes the intra-group loans and debts information for Chilean listed firms, spanning 1990–2009 and around 1,000 firm-year observations. Santioni et al. (2020) utilizes a rare example of an Italian firm-level dataset that includes this information, and hence is able to visualize intra-group loans for a large sample of more than 400,000 firm-years. Hong (2022) uses this information for a large sample of South Korean firms (2013–2019) spanning 610,000 firm-years, but has focused only on the sales and purchases of intermediate goods and services.

This is the first data, to the best of my knowledge, that simultaneously shows the trades of goods and services, trade credit, and non-trade credit loans and debts for a large sample of firms. Through the scraping process, all publicly available existing financial reports between FY 2013–2021 are utilized, constructing a database of 34,129 firms and more than 790,000 firm-years.

Figure 3.2.1: Sample Ownership Structure Between Firms



NOTE.—The figure shows a sample structure of ownership between firms. A to F refer to firms in the group, and the direction of arrows correspond to the direction of ownership. The solid lines refer to ownership of more than 50%, and the dotted lines refer to ownership of more than 20%, but less than 50% in the firm below.

3.2.2 Measurement of Ownership Structures and Cash-Flow Rights

Aside from the intra-group transactions, firms also report the ownership structures of the group in this section. Each firm reports, in discrete categories, the relationship that it has with a related party. From the discrete categories, I construct two categorical variables that summarize relevant information about ownership structures: *Direction*, *Direct Ownership*, and *CFR*.

The first, *Direction*, shows the direction of control with the related party. *Above* (*Below*) is when the related party is above (below) the reporting firm in the ownership tree, and *Lateral* is when the control is exercised through a lateral component without direct control between the two. In Figure 3.2.1, in firm B's financial reports, *Above* would refer to firm A, *Below* refers to firms D and E, and *Lateral* to C and F.

The second, *Direct Ownership*, shows whether there are direct shareholdings between the two firms, or if there are only indirect shareholdings. Again using firm B's financial reports from Figure 3.2.1, firms A, D, and E have direct shareholdings while firms C and F do not.

Lastly, *CFR*, a binary variable, reflects how strongly the group holds control over the related party. South Korean accounting requirements state that a member of a group must report the transactions with a related party, as long as the *combined* ownership of the controlling shareholder over the related party is larger than 20%. When the firms do report related-party transactions, related parties that are connected with the firm only with links of larger than 50% of ownership are listed in separate categories from those with links below 50%. Using this, I create the *CFR* dummy variable, which shows the likelihood of the controlling shareholders' ownership of the related party being high and hence works as a proxy for CFR. In the financial statements of firms A, B, C, D, and F, firm E, which is linked from the group only with a weak ownership link, will be categorized as *Low CFR*, as opposed to *High CFR*.

Utilizing *CFR* as a proxy for CFR is, admittedly, a weak proxy in the absence of better data. For each trading related party, firms only report whether (i) more than 50% of shares of the related party are owned by the group, or (ii) less than 50% but more than 20%. While the more detailed information on CFR in each firm is more desired, the controlling shareholder's CFR in group (i), on average, is higher than CFR in group (ii), and I utilize this variation.

One merit of utilizing the *CFR* variable is that it accounts for both the direct and the indirect shareholdings. Firms need to account for detailed indirect holdings, including the shares owned by families of controlling shareholders, as well as the more standard indirect ownership through other firms. Compared to only utilizing the latter, this variable can better capture the private incentives of the group owners.[2]

At the same time, the measure is limited in that it is binary, as well as that it cannot account for the cases where the controlling shareholder has high CFR in the related

[2]The related-party transactions section of the financial statements go through external auditing as well as government scrutiny, and the scrutiny extends to whether the indirect shareholdings have been accurately calculated. In a well-known case, in 2018, Samsung Biologics was suspended stock trading because the firm had allegedly intentionally miscalculated the indirect shareholdings to a subsidiary of its own and had categorized it as 'weakly' owned by the group (Jeong and Martin, 2018), The Wall Street Journal.

parties through non-voting shares, but has little control through voting shares. The latter case, however, is unlikely to drive the results, and the binary variable can still show the overall effects.

3.2.3 Firm Financial Data

Aside from the intra-group transactions, all other firm-level characteristics are merged from three established databases: KISVALUE, TS2000, and ORBIS by Bureau van Dijk. KISVALUE and TS2000 are the largest firm-level databases in South Korea and cover all firms required to publicly disclose annual financial reports. The ORBIS database maintained by Bureau van Dijk, on the other hand, covers firm-level information around the world.

The information is matched to the reporting firms by 10-digit firm-level tax identifiers. On the other hand, firms do not provide detailed identifiers for the related parties that appear on the reports and instead only provide the names of the firms. The related parties' financial data are matched to the firms by the names and go through extensive checks to ensure accuracy. Details of the matching and verification process are outlined in the appendices of Hong (2022).

3.2.4 Quarterly Financinal Statements

In order to analyze short-term fluctuations in intra-group transactions in Section 3.4, I also utilize quarterly financial statements. As only publicly listed firms are required to prepare and disclose quarterly financial information, the sample size is reduced. For 2,740 listed firms, the quarterly intra-group transactions are visible from the first quarter of 2018 to the last quarter of 2021. In the end, intra-group transactions between the total of 388,099 firm-quarter-related party triples are observed.

The original quarterly statements reflect each firm's accumulated activities over the year until the quarter in which the reports are prepared. For example, a firm's third-quarter statements reflect its intra-group sales from the first fiscal quarter to the third, and the amount of intra-group loan outstanding on the last day of the third fiscal quarter. Therefore, for flow variables such as sales or purchases, I calculate the difference with the previous quarter to obtain the trades that transpired during the quarter.

3.3 Providers and Receivers of Intra-Group Loans

Using the data, I study whether factors that explain the intra-group loan provision are more in line with the tunneling hypothesis or the efficiency hypothesis. The two hypotheses provide different predictions on the characteristics of the firms that are net providers of loans within a group, and those that are net receivers.

The efficiency hypothesis predicts that intra-group loans benefit the entire group. That is, the loans should provide benefits that are either unobservable or unobtainable to the providers of external finance. In the first case, the firms' successes or failures may have externalities or spillover effects on at least some members within the group. Then, it would be rational for the other members of the group to provide liquidity to prevent negative spillovers (Gopalan et al., 2007). In the second case, the group may have better information about the firm in need of financing, such as in terms of long-term profitability or solvency, and therefore may be willing to provide loans on better terms.

Hence, the efficiency hypothesis predicts that firms with the following characteristics are more likely to become the net receivers of intra-group loans (Buchuk et al., 2014): (1) higher capital intensity, (2) higher cost of external financing, (3) higher profitability, and (4) smaller short-term liquidity.

H1. The efficiency hypothesis predicts that a firm is more (less) likely to be a net receiver (provider) of loans from (to) a related party if the firm has (1) higher capital intensity, (2) higher cost of obtaining external finance, (3) higher profitability, and (4) smaller short-term liquidity than the related party.

On the other hand, the tunneling hypothesis predicts that all other things being equal,

(1) the firms with higher CFR are more likely to receive loans, and (2) the firms with lower CFR are more likely to become the net providers of the loans. I also account for this possibility by utilizing the proxy for firms with low CFR.

H2. The tunneling hypothesis predicts that a firm is more (less) likely to be a net receiver (provider) of loans from (to) a related party with low CFR.

Using all five variables, I test whether they explain which of two firms in a pair of related parties provides loans to the other. To this end, I estimate the model in equation (1), in Table 3.3.1, where Φ denotes the logit distribution function. Here I utilize annual transactions between pairs of firms within the same groups.

$$\begin{aligned} Prob(y_{ijt}=1) = & \Phi(\beta_0 + \beta_1 CFR_{jt} + \beta_2 \Delta_{ijt} CapitalIntensity + \beta_3 \Delta_{ijt} \ln{(Assets)} + \\ & \beta_4 \Delta_{ijt} Profitability + \beta_5 \Delta_{ijt} Liquidity + \beta_6 DirectOwnership_{ijt} + \\ & \beta_7 Direction_{ijt} + \beta_8 I(Sales_{ijt} > 0) + \beta_9 I(Purchases_{ijt} > 0) + \\ & \delta_i + \gamma_t) \end{aligned} \tag{1}$$

In columns (1) and (2), y_{ijt} is whether firm i being a net lender to firm j. In columns (3) and (4), it is whether i is a net receiver of loans from firm j. Each type of dependent variable is calculated once only with non-trade credit loans (*Loans*) and then with the total of non-trade credit loans and trade credits (*Gross Loans*). All regressions were carried out after accounting for the reporting firm and year fixed effects separately, and with standard errors clustered at the reporting firm level.

Note that the data only contains transactions between related parties. Therefore, the data consists of pairs where (i) j is a net lender to j, (2) i is a net receiver of loans from j, (3) or i and j do not have a lending relationship. Due to data limitations, I cannot observe lending relationships with unrelated parties.

CFR_{ijt} is as defined in Section 3.2 and corresponds to whether the group owner's CFR in firm j is likely to be high in year t, and is the variable that accounts for the tunneling hypothesis.

The next four variables account for the efficiency hypothesis: (1) capital intensity, calculated with the share of tangible assets over total assets, (2) cost of obtaining external finance, proxied with the log of total assets, which has been used in the literature as strongly and inversely correlated with the cost, (3) profitability, calculated with earnings before interest and tax (EBIT) divided by total assets, and (4) liquidity, calculated by the difference between current assets and current debts, divided by the firm's total assets. All four terms are expressed in differences of the variables between i and j (Δ_{ijt}), which will ultimately dictate the loans relationship—e.g., the *more profitable* firm among the pair is more likely to receive loans, and so on.

The subsequent four variables, $DirectOwnership_{ijt}$, $Direction_{ijt}$ and whether the firm has any intermediate sales to (purchases from) the related party, account for the possibility that firm i has more information about j, as well as more exposure to externalities from j's performance. $DirectOwnership_{ijt}$ and $Direction_{ijt}$ are variables corresponding to more details about the ownership structure, and are as defined in 3.2.

The first implication from Table 3.3.1 is that the evidence is not supportive of the predictions of the *tunneling* hypothesis, or *H2*. Note that the coefficient for low CFR is negative for all four specifications. The related party of a firm having a smaller CFR appears to simply lower the likelihood of the two firms having a lending relationship instead of strictly favoring one side. This is inconsistent with the tunneling hypothesis, which predicts that loans flow from firms with low CFR to high CFR firms. While this is still an early result—lack of data on detailed CFR, as well as the interest rates and other terms of the loans, is required for a more definitive statement—the data is inconsistent with the theoretical predictions, which state that the low CFR of the related party would be negatively correlated with a net lending relationship to the related party, and positively with a net receiving relationship from it.

On the other hand, the results in Table 3.3.1 robustly support the implications of the efficiency hypothesis (1) to (4), or *H1*. Firms are less likely to be a net provider of loans when they are more capital intensive than the related party, more profitable, have smaller external financing costs (larger log assets), and have more short-term liquidity. In this way, the flow of capital within groups appears to benefit the group instead of favoring the controlling shareholders' interests by preying on minority shareholders.

3.3.1 Factors Behind the Efficiency Hypothesis

However, if intra-group loans are not abused by the controlling shareholders, why are they extended, and what factors can explain the pattern of their extension? More specifically, how can the group make the reallocation of capital through intra-group loans beneficial to the group, possibly more so compared to the external finance sector? This paper provides evidences that emphasize two possible explanations.

First, the loans may be driven by concerns over a member firm's performances *spilling over* to others. Gopalan et al. (2007) argue that in Indian business groups, concerns over a member firm's bankruptcy having negative externalities on other firms, e.g., through equity stakes or heightened scrutiny from external creditors, influence the flow of intra-group loans. Moreover, they find that for group members that have closer managerial connections, the spillover concerns can be stronger and therefore have larger effects.

This paper is in line with Gopalan et al. (2007)'s findings and adds to the literature by showing that the direction and proximity of ownership have direct bearings on the capital flows within groups. Having a direct ownership relationship with another member of the firm, regardless of the direction, is strongly and positively correlated with a lending relationship being formed. Considering that the proximity in ownership chains is directly correlated with spillover effects from one to the other, this result shows that concerns over such influences are important factors that govern intra-group loans.

Table 3.3.1: Net Providers and Receivers of Intra-group Loans, Logit Models

	Net Provider		**Net Receiver**	
	(1) Loans	(2) Gross Loans	(3) Loans	(4) Gross Loans
Low CFR	-0.223^{***}	-0.113^{***}	-0.125^{***}	-0.228^{***}
	(0.036)	(0.030)	(0.036)	(0.031)
Δ_{ijt}Capital Intensity	-0.001^{**}	-0.001^{*}	0.000	0.001
	(0.000)	(0.001)	(0.001)	(0.001)
Δ_{ijt}log (Assets)	0.138^{***}	0.157^{***}	-0.079^{***}	0.040
	(0.007)	(0.006)	(0.007)	(0.006)
Δ_{ijt}Profitability	-0.068^{*}	-0.061^{*}	-0.039	-0.055
	(0.034)	(0.028)	(0.035)	(0.031)
Δ_{ijt}Liquidity	0.803^{***}	0.691^{***}	-0.679^{***}	-0.723^{***}
	(0.026)	(0.023)	(0.028)	(0.024)
Direct Ownership	0.602^{***}	0.293^{***}	-0.042	0.146^{*}
	(0.098)	(0.077)	(0.085)	(0.072)
Direction_{ijt} : *Below*	-0.034	0.120	0.188^{*}	0.246^{**}
	(0.102)	(0.081)	(0.090)	(0.076)
Direction_{ijt} : *Above*	-0.238^{*}	-0.227^{**}	0.669^{***}	0.657^{***}
	(0.095)	(0.075)	(0.082)	(0.069)
$\text{Sales}_{ijt} > 0$	0.857^{***}	1.434^{***}	-0.143^{***}	-0.774^{***}
	(0.026)	(0.023)	(0.024)	(0.020)
$\text{Purchases}_{ijt} > 0$	-0.120^{***}	-0.820^{***}	1.315^{***}	1.554^{***}
	(0.023)	(0.019)	(0.025)	(0.022)
Num. obs.	87,466	107,105	84,878	106,576
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note.—This table presents coefficients from logistic regressions. The dependent variables are net provider, a dummy variable with a value of one in firm-year-related party triples with a positive value of net intra-group loans from the reporting firm to the related party, and net receiver, also a dummy variable with a value of one in firm-year-related party observations when there is a negative value of net loans from the reporting firm to the related party. In columns (1) and (2), the dependent variables are calculated only using the non-trade credit intra-group loans. Columns (3) and (4) combine the trade credit and the non-trade credit loans. All regressions include the reporting firm fixed effects and year fixed effects, and the standard errors are clustered at the reporting firm level. The sample consists of South Korean firm pairs in business groups and extends from 2013 to 2021. Variations in sample size utilized in each regression come from the number of firms that are successfully matched to all independent variables.

Moreover, the effects of the direction of ownership also support the same theory. The performances of subsidiaries directly appear on the parent companies' books, influencing the parent firms' profits as well as how they are viewed in the market. Therefore, the spillover theory predicts that loans flow from parent firms to subsidiaries. The results in Table 3.3.1 confirms that this is indeed true.

The second possible factor behind intra-group loans is that they are utilized by the group to support and accelerate the growth of its nascent new ventures. In particular, this paper finds empirical evidence suggesting that intra-group loans are utilized to extend the firms' internal supply chains further downstream.

In Table 3.3.1, I find a positive correlation of a lending relationship forming from the seller of intermediate goods to the buyers. In other words, the flow of intermediate goods and the flow of intra-firm loans tend to flow in the same direction, and lending relationships in opposite directions are less likely to form. This trend does not conform with the *tunneling* hypothesis but is suggestive of a pattern where groups extend the intra-group supply chain further downstream. In other words, the intra-group loans help the downstream subsidiary, even at its fledgling stage, to absorb the sales of intermediate goods from its upstream affiliates. In this way, the new venture may rapidly increase its market share in the downstream market.

To test this hypothesis, Table 3.3.2 demonstrates whether the intra-firm buyers at their incipient stages are more likely to receive loans. The regression equation (1) is tested after including interaction terms of the supply chain variables and an indicator variable showing whether the related party is in its nascent stage. I utilize two types of such indicator variables. Columns (1) and (2) utilize whether the related party has a below-median asset size within its 2-digit industry, and (3) and (4) use whether less than 3 calendar years have passed since the related party was born.

The results uniformly support this paper's hypothesis. Smaller firms are much more likely to receive loans and less likely to provide loans. Moreover, when the related party

Table 3.3.2: Net Providers and Receivers of Intra-group Loans, Extended Models

	(1) Provider	(2) Receiver	(3) Provider	(4) Receiver
Low CFR	-0.223^{***}	-0.129^{***}	-0.224^{***}	-0.125^{***}
	(0.036)	(0.036)	(0.036)	(0.036)
Δ_{ijt}Capital Intensity	-0.001^{**}	0.000	-0.002^{**}	0.000
	(0.000)	(0.001)	(0.000)	(0.001)
Δ_{ijt}log (Assets)	-0.134^{***}	-0.044^{***}	-0.144^{***}	-0.075^{***}
	(0.008)	(0.008)	(0.007)	(0.007)
Δ_{ijt}Profitability	-0.069^{*}	-0.037	-0.072^{*}	-0.035
	(0.034)	(0.035)	(0.035)	(0.035)
Δ_{ijt}Liquidity	0.803^{***}	-0.675^{***}	0.804^{***}	-0.679^{***}
	(0.026)	(0.028)	(0.026)	(0.028)
Direct Ownership	0.605^{***}	-0.054	0.589^{***}	-0.041
	(0.098)	(0.085)	(0.098)	(0.085)
Direction$_{ijt}$: *Below*	-0.036	0.209^{*}	-0.021	0.188^{*}
	(0.102)	(0.090)	(0.102)	(0.090)
Direction$_{ijt}$: *Above*	-0.245^{**}	0.675^{***}	-0.227^{*}	0.669^{***}
	(0.095)	(0.082)	(0.095)	(0.082)
Sales$_{ijt} > 0$	0.833^{***}	-0.102^{***}	0.841^{***}	-0.129^{***}
	(0.030)	(0.028)	(0.026)	(0.024)
Purchases$_{ijt} > 0$	-0.073^{*}	1.393^{***}	-0.115^{***}	1.308^{***}
	(0.029)	(0.029)	(0.024)	(0.025)
Sales$_{ijt} > 0$ * Small$_{jt}$	0.055	-0.093^{**}		
	(0.034)	(0.035)		
Purchases$_{ijt} > 0$ * Small$_{jt}$	-0.099^{**}	0.190^{***}		
	(0.036)	(0.034)		
Sales$_{ijt} > 0$ * Infant$_{jt}$			0.227^{***}	-0.240^{***}
			(0.057)	(0.066)
Purchases$_{ijt} > 0$ * Infant$_{jt}$			-0.034	0.041
			(0.071)	(0.070)
Num. obs.	87,466	84,878	87,402	84,830
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note.—This table reports logistic regressions that extend the analysis from Table 3.3.1. The dependent variables all utilize net non-trade credit loans from the reporting firm to the related party during the given fiscal year. In columns (1) and (2), $Small_{jt}$ is an indicator variable with a value of one if the related party's total asset is smaller than its 2-digit industry's median total assets in the given year. In Columns (3) and (4), $Infant_jt$ is an indicator variable with a value of one when it has been less than 3 calendar years since the related party was established. All regressions include the reporting firm fixed effects and year fixed effects, and the standard errors are clustered at the reporting firm level. The sample consists of South Korean firm pairs in business groups and extends from 2013 to 2021. Variations in sample size utilized in each regression come from the number of firms that are successfully matched to all independent variables.

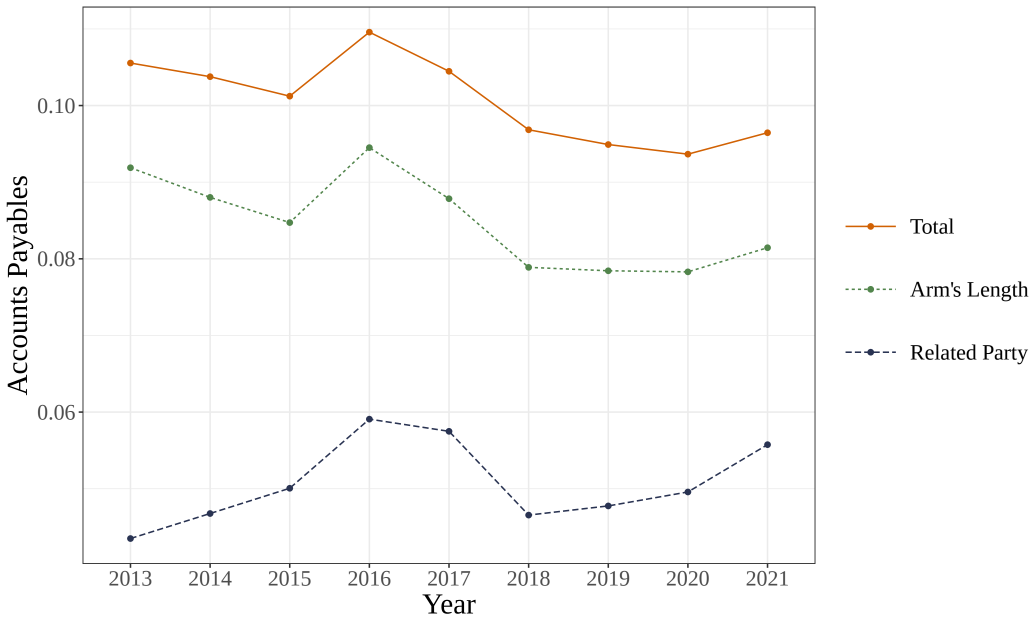
is in its nascent stage, the effect of being in supply chains is amplified in all specifications.

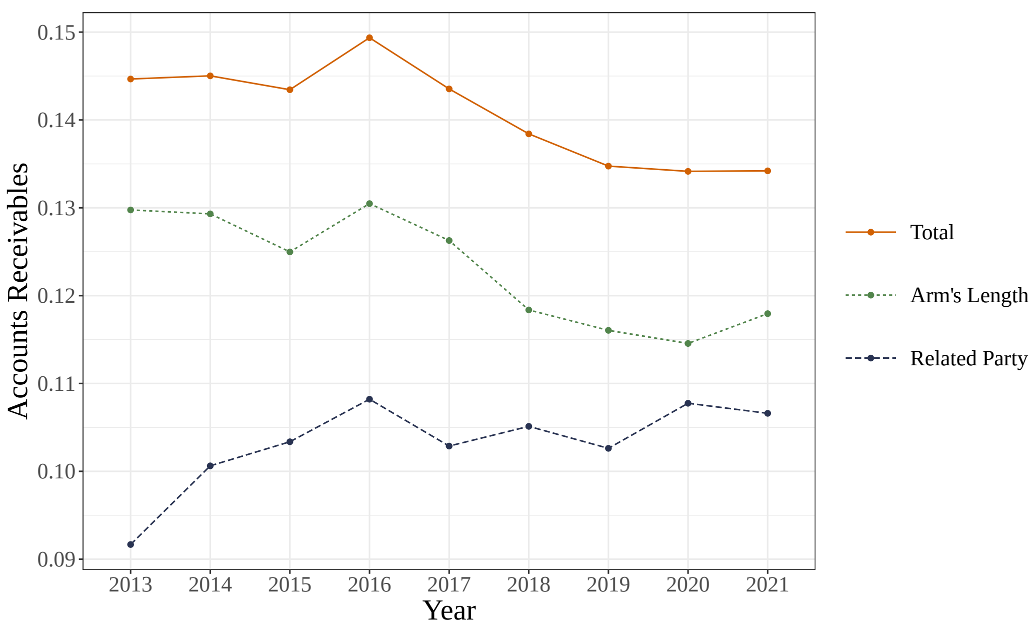
One caveat is on the sample size's dependency on how I define the dependent variable. When using net gross loans to define the dependent variable (columns (2) and (4) of Table 3.3.1), the sample size is larger than when using only the non-trade credit. This is because due to the fixed effects for the reporting firm and the fiscal year, the data loses more observations in non-trade credit terms as having only net providing relationships or net receiving relationships. This provides one evidence showing that differences between trade credits and non-trade credit loans and debts may be an important factor in analyzing intra-group capital flows. However, literature in most cases have not paid attention on the two items (Buchuk et al., 2014). This paper delves a step further into this issue in the following section.

3.4 Intra-Group Trade Credits

This paper brings forward novel stylized facts about a set of intra-group transactions that have hitherto been missing from the discussion: trade credit. Trade credits merit separate analysis from the non-trade credit loans (hereinafter 'loans' for brevity) for the following reasons. First, trade credits assume a large part of firms' working capital, and are susceptible to large changes when in sharp short-term economic shocks (Love et al., 2007; Garcia-Appendini and Montoriol-Garriga, 2013). Second, sales and purchases of intermediate goods have to precede trade credits; in other words, they only take place upon the pre-existing relationships of supply chains. Hence, incentives other than the standard incentives behind the provision of loans may be important.

Figure 3.4.1: Median Accounts Payables and Receivables, Annual Financial Statements





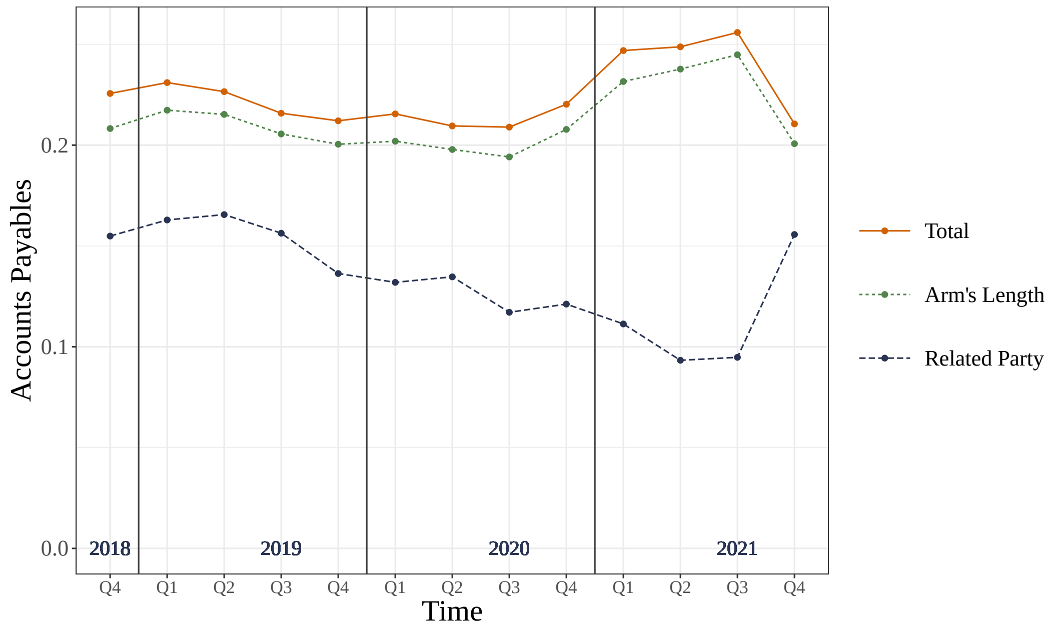
Note.—The figures show median values of South Korean firms' trade credit uses in each fiscal year, divided by the types of trades. The vertical axis of the top figure is median Accounts Payables divided by Cost of Goods and Services (COGS). The vertical axis of the bottom figure is Accounts Receivables divided by Sales. *Total* shows firm's total economic activities, *Related Party* shows trades with related parties only, and *Arm's Length* shows non-related party trades only.

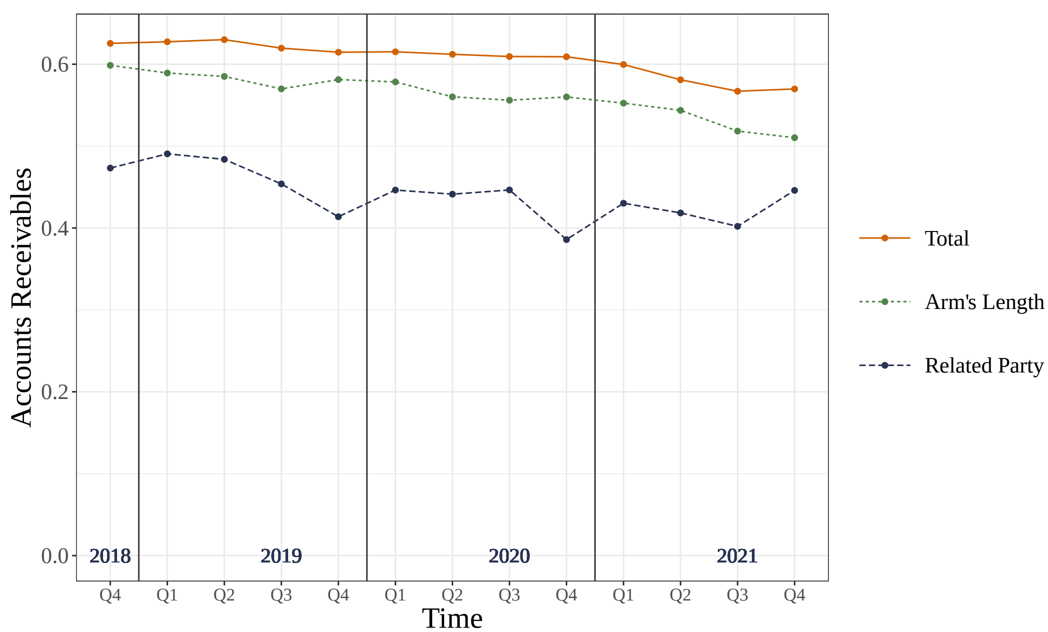
As a starting point toward filling the gap in the literature, this paper documents two novel stylized facts regarding trade credit practices between related parties. First, trade credits consistently appear to account for a much smaller portion of intra-group trades, compared to arm's length. Figure 3.4.1 shows the yearly median share of Accounts Payables among Cost of Goods Sold, and the share of Accounts Receivables in Total Sales, respectively. Compared to total sales, for related-party trades, the trade credits continuously assume a much smaller portion.

One possible hypothesis is that in intra-group trades, firms do not need to leverage trade credits to ensure complete adherence to the trade contracts. Research including and following Antras and Foley (2015) has demonstrated early-stage evidences showing that larger firms tolerate higher financing costs by forcing the smaller and less stable counterparties to provide trade credit, thereby ensuring that the trade partners will uphold the contracts fully. In the intra-group context, existing ownership chains may act as insurance that incentives are aligned between the buyer and the seller, and therefore there may be less need for trade credits to remain on the books.

Second, when faced with a short-term economic shock, there is early-stage evidence showing that intra-firm trade credit reacts differently compared to arm's length trade credits. Figure 3.4.2 demonstrates the quarterly changes in median trade credits previous to and entering the global pandemic of COVID-19. On the accounts payables side, the data shows a sharp divergence between the intra-group trade credits, while the change is less visible from the accounts receivables side.

Figure 3.4.2: Median Accounts Payables and Receivables, Quarterly Financial Statements





NOTE.—The figures show median values of South Korean firms' trade credit uses in each fiscal quarter, divided by the types of trades. The vertical axis of the top figure is median 3-quarter moving average Accounts Payables divided by Cost of Goods and Services (COGS). The vertical axis of the bottom figure is Accounts Receivables divided by Sales. *Total* shows firm's total economic activities, *Related Party* shows trades with related parties only, and *Arm's Length* shows non-related party trades only.

This trend of intra-firm trade credits contrasts the literature's findings on arm's length trade credits. In fact, even in the South Korean data, accounts payables with non-related parties, which rise sharply then decrease, follows exactly what Love et al. (2007) have documented with the 1997 Asian financial crisis and the 1994 Mexico peso devaluation. Love et al. (2007) show that initially in a sharp economic downturn, the use of trade credit appears to increase sharply then gradually decreases. However, that the intra-group trade credits show the opposite movements, as well as the reason behind such divergence, has not been documented.

Such diverging trends of trade credits between related-party trades and arm's length transactions merit further investigation. What factors are behind such differences? Moreover, by limiting the analysis to only the non-trade credit loans, is the analysis missing an important network of capital flows intra-group, or will the analysis only show small differences, as was shown in Table 3.3.1? Lastly, do groups utilize trade credits and explicit loans interchangeably when reallocating resources, or do the two types of loans follow separate incentives? I intend to study this in subsequent papers.

3.5 Conclusion

A major discussion on business groups has focused on whether intra-group transactions are utilized for the benefits of the entire group, or for the private benefits of those in ultimate control of it. In such discussions, a long-standing lack of two aspects of data has impeded rigorous empirical research: an accurate measurement of the group-controlling shareholders' cash-flow rights, and detailed measurements of intra-group transactions that span from trades of intermediate goods to flow of funds through loans.

This paper contributes to the literature by substantively improving upon the latter. By web-scraping all publicly available financial statements of South Korean firms, this paper is able to visualize the intra-group dealings of more than 34,000 firms and 790,000 firm-year-related party triples, in various contexts spanning trades, trade credits, and loans.

By utilizing the better data, this paper firstly strengthens and supports the current literature and provides evidence that, at least in the South Korean context, the intra-group flow of loans seems to benefit the entire group. I show, in line with the literature, that the firms with higher marginal benefits from obtaining more funds are significantly more likely to be net receivers of such loans, and less likely to be net providers. The group-controlling shareholders' cash-flow rights, on the other hand, do not appear to affect the direction of loans as expected by the tunneling hypothesis.

Moreover, this paper extends the current literature by providing more context into *why* business groups find intra-group reallocation of funds beneficial to the entire group. This paper finds that proximity within ownership relationships, as well as supply chain relationships, are strong indicators of lending relationships. This finding emphasizes that concerns over *spillovers* from a member firm's performance to others may provide an incentive to reallocate capital. Also, I show evidence showing that firms support new ventures that extend the internal supply chain by providing loans.

Lastly, this paper provides a starting point in filling the gap in the literature on intra-group loans: trade credits. Even though trade credits are widely used and are vital components of firms' working capital, they have not been studied in a business group setting. I document two novel stylized facts that attest to a possible divergence of firms' trade credit usage within and across the firm border.

One caveat that should follow is that the results covered in this paper are still of exploratory nature, and that speaks to one specific economy of one data period. For one thing, due to data limitations, this paper cannot speak to whether *prices* within intra-group transactions are affected by the tunneling or the efficiency hypothesis. Also, this paper's results may not necessarily extend to other economies or eras—as literature has repeatedly pointed out, regulations play a large role in shaping and limiting the private profit-seeking of business groups' controlling shareholders. Further investigations into the more rigorous causal inferences, as well as extensions into how regulations have shaped such behaviors, are in order: this paper provides a small step in a better direction.

Bibliography

Acemoglu, D., S. Johnson, and T. Mitton (2009). Determinants of vertical integration: financial development and contracting costs. *The Journal of Finance 64*(3), 1251–1290.

Aghion, P., R. Griffith, and P. Howitt (2006). Vertical integration and competition. *American Economic Review 96*(2), 97–102.

Alfaro, L. and A. Charlton (2009). Intra-industry foreign direct investment. *American Economic Review 99*(5), 2096–2119.

Alfaro, L., D. Chor, P. Antras, and P. Conconi (2019). Internalizing global value chains: A firm-level analysis. *Journal of Political Economy 127*(2), 508–559.

Alfaro, L., P. Conconi, H. Fadinger, and A. F. Newman (2016). Do prices determine vertical integration? *The Review of Economic Studies 83*(3), 855–888.

Alfaro-Ureña, A., M. Fuentes, I. Manelici, and J. Vasquez (2018). The costa rican production network: Stylized facts. *Research Paper Series*.

Alfaro-Ureña, A., I. Manelici, and J. P. Vasquez (2020). The effects of joining multinational supply chains: New evidence from firm-to-firm linkages. *Available at SSRN 3376129*.

Almeida, H., C.-S. Kim, and H. B. Kim (2015). Internal capital markets in business groups: Evidence from the asian financial crisis. *The Journal of Finance 70*(6), 2539–2586.

Altomonte, C., G. Ottaviano, and A. Rungi (2018). Business groups as hierarchies of firms: Determinants of vertical integration and performance. Technical report, Fondazione Eni Enrico Mattei.

Antràs, P. and D. Chor (2013). Organizing the global value chain. *Econometrica 81*(6), 2127–2204.

Antras, P. and C. F. Foley (2015). Poultry in motion: a study of international trade finance practices. *Journal of Political Economy 123*(4), 853–901.

Atalay, E., A. Hortaçsu, M. J. Li, and C. Syverson (2019). How wide is the firm border? *The Quarterly Journal of Economics 134*(4), 1845–1882.

Atalay, E., A. Hortaçsu, and C. Syverson (2014). Vertical integration and input flows. *American Economic Review 104*(4), 1120–48.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*, 685–725.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2010). Intrafirm trade and product contractibility. *American Economic Review 100*(2), 444–48.

Blyde, J. and D. Molina (2015). Logistic infrastructure and the international location of fragmented production. *Journal of International Economics 95*(2), 319–332.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Buchuk, D., B. Larrain, F. Muñoz, and F. Urzúa (2014). The internal capital markets of business groups: Evidence from intra-group loans. *Journal of Financial Economics 112*(2), 190–212.

Davis, J. and M. Goadrich (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pp. 233–240.

Efron, B. (2020). Prediction, estimation, and attribution. *International Statistical Review 88*, S28–S59.

Fajgelbaum, P., G. M. Grossman, and E. Helpman (2015). A linder hypothesis for foreign direct investment. *The Review of Economic Studies 82*(1), 83–121.

Fan, J. P. and L. H. Lang (2000). The measurement of relatedness: An application to corporate diversification. *The Journal of Business 73*(4), 629–660.

Fort, T. C. (2017). Technology and production fragmentation: Domestic versus foreign sourcing. *The Review of Economic Studies 84*(2), 650–687.

FTC (2021, September 15). Statement of chair lina m. khan, commissioner rohit chopra, and commissioner rebecca kelly slaughter on the withdrawal of the vertical merger guidelines. *Commission File No. P810034*.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2021). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance*, Forthcoming.

Garcia-Appendini, E. and J. Montoriol-Garriga (2013). Firms as liquidity providers: Evidence from the 2007–2008 financial crisis. *Journal of financial economics 109*(1), 272–291.

Garg, S., B. Tan, and P. Ghosh (2021). Within firm supply chains: Evidence from india. *Working Paper*.

Gopalan, R., V. Nanda, and A. Seru (2007). Affiliated firms and financial support: Evidence from indian business groups. *Journal of Financial Economics 86*(3), 759–795.

Grossman, S. J. and O. D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy 94*(4), 691–719.

Hart, O. and J. Moore (1990). Property rights and the nature of the firm. *Journal of political economy 98*(6), 1119–1158.

Hoberg, G. and V. Maksimovic (2015). Redefining financial constraints: A text-based analysis. *The Review of Financial Studies 28*(5), 1312–1352.

Hoberg, G. and G. Phillips (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy 124*(5), 1423–1465.

Hong, Y. (2022). Related-party trades in vertical integration. *SSRN*.

Hosmer Jr, D. W., S. Lemeshow, and R. X. Sturdivant (2013). *Applied logistic regression*, Volume 398. John Wiley & Sons.

Jeong, E.-Y. and T. W. Martin (2018, June 12). South korea regulator says samsung biologics violated accounting rules. *The Wall Street Journal*.

Johnson, S., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2000). Tunneling. *American economic review 90*(2), 22–27.

Klapper, L., L. Laeven, and R. Rajan (2012). Trade credit contracts. *The Review of Financial Studies 25*(3), 838–867.

Klein, B., R. G. Crawford, and A. A. Alchian (1978). Vertical integration, appropriable rents, and the competitive contracting process. *The journal of Law and Economics 21*(2), 297–326.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018). Human decisions and machine predictions. *The quarterly journal of economics 133*(1), 237–293.

Lafontaine, F. and M. Slade (2007). Vertical integration and firm boundaries: The evidence. *Journal of Economic Literature 45*(3), 629–685.

Li, K., F. Mai, R. Shen, and X. Yan (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies 34*(7), 3265–3315.

Love, I., L. A. Preve, and V. Sarria-Allende (2007). Trade credit and bank credit: Evidence from recent financial crises. *Journal of Financial Economics 83*(2), 453–469.

Monarch, R., J. Park, and J. Sivadasan (2017). Domestic gains from offshoring? evidence from taa-linked us microdata. *Journal of International Economics 105*, 150–173.

Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics 122*(2), 569–600.

Nunn, N. and D. Trefler (2013). Incomplete contracts and the boundaries of the multinational firm. *Journal of Economic Behavior & Organization 94*, 330–344.

Ramondo, N., V. Rappoport, and K. J. Ruhl (2016). Intrafirm trade and vertical fragmentation in us multinational corporations. *Journal of International Economics 98*, 51–59.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of international Economics 48*(1), 7–35.

Ruhl, K. J. (2015). How well is us intrafirm trade measured? *American Economic Review 105*(5), 524–29.

Santioni, R., F. Schiantarelli, and P. E. Strahan (2020). Internal capital markets in times of crisis: The benefit of group affiliation. *Review of Finance 24*(4), 773–811.

Williamson, O. E. (1971). The vertical integration of production: market failure considerations. *The American Economic Review 61*(2), 112–123.

Williamson, O. E. (1979). Transaction-cost economics: the governance of contractual relations. *The journal of Law and Economics 22*(2), 233–261.