Product Bundling, Joint Markups and Trade Liberalization

Ji Hye Heo*

October 2022 Version

Abstract

Product bundling is a frequent practice of multi-product firms used to increase firm-level profits. This paper examines how product bundling affects a firm's markups at various levels in international trade. Joint pricing decisions for multi-product firms with product bundling entail information about joint markups in their profit maximization problem. Utilizing this information, I propose a method to estimate transaction-level markups incorporating multi-product firms with product bundling. Focusing on Chinese exporters, multi-product firms that bundle products enjoy markups that are roughly 30% higher than firms with independent pricing. Analysis of China's WTO accession shows that although trade liberalization increased markups for bundling firms, tariff reductions reduced markup differences across all firms through greater competition.

Keywords: Joint markups, multi-product firms, product bundling, pass-through, consumer valuation estimation, trade liberalization

JEL Codes: D22, D24, L11, F13, L60

*I am immensely grateful to my advisor Joel Rodrigue for his continuous support and to my committee members Tong Li, Yuya Sasaki, and Luke Froeb for their advice and support on this project. I would also like to thank seminar participants at Vanderbilt University and conference participants at the WEAI, Young Economists Symposium 2022 for their comments and helpful suggestions. Contact: ji.hye.heo@vanderbilt.edu. Vanderbilt University, Department of Economics

1 Introduction

Multi-product firms are generally larger and hold a significant share of transactions in international trade, as shown by Bernard et al. (2012). As such, understanding the behavior of multi-product firms is central to characterizing the costs and benefits of trade policy. Naturally, there have been many studies on multi-product firms, such as their entry/exit decisions, product scope, and product quality (See Bernard et al. (2011), Lopresti (2016), and Manova and Yu (2017)). However, previous literature on multi-product firms and trade has overlooked one of the main advantages multi-product firms have over single-product firms; the ability to engage in product bundling, a practice whereby firms sell multiple goods in a single package.[1] This paper seeks to study (i) how product bundling impacts a firm's markups and (ii) how the impacts of trade liberalization may differ for bundling firms.

Multi-product firms that engage in product bundling make pricing decisions jointly across products they decide to bundle. Since previous methods to recover markups rely on an implicit assumption that firms price goods independently, markups for bundling firms are not correctly captured by methods from the literature. Thus, I first provide an alternative methodology to flexibly identify markups at the transaction level for both bundling and non-bundling firms. The multi-product firm with mixed bundling practices makes joint pricing decisions for all of its single-product goods and bundles to maximize its profit at the firm-level. Thus, the firm's first-order conditions from the profit maximization problem reflect the firm's joint pricing decision, which creates a markup linkage across goods. These expressions for markups are in terms of prices and consumer tastes across products, captured by the distribution of consumer valuations. After estimating consumer tastes with transaction data, markups are recovered using the information from the first-order conditions.

This paper proposes a methodology that uses demand-side information based on a firm's profit maximization problem and consumer rationality, similar to the traditional structural

[1] As a multiple goods provider, multi-product firms can sell their products independently with separate pricing or jointly with product bundling. The option to buy products separately or as a bundle is referred to as *mixed bundling* practices. If buyers can only purchase products as a bundle, it is called *pure bundling* practices. Depending on the market structure and the firm's market power, a firm engaging in bundling practices could price a bundle at a higher (*bundling premium*) or a lower (*bundling discount*) price. This paper focuses on mixed product bundling practices with price discounts, the most prominent case in product bundling.

approach from Berry et al. (1995). This significantly departs from the widely used method to recover markups using production-side information and a firm's cost minimization problem proposed by De Loecker and Warzynski (2012) and De Loecker et al. (2016).[2] This paper's method also departs from the approach from Berry et al. (1995) because consumer tastes across products are estimated using transaction data, which eases the burden on the data compared Berry et al. (1995), where sales, product characteristics, and market share data are required. By incorporating product bundling into the framework, this paper characterizes the difference in markup strategy across independent pricing firms and bundling firms with joint pricing decisions, which previous methods cannot address.

This paper is also among the first to study strategic bundling practices by multi-product firms in an international trade context. After the basic framework of Stigler (1963), Adams and Yellen (1976), and McAfee et al. (1989), the bundling literature has focused on either theoretically extending the basic framework[3] or analyzing bundling practices in retail, telecommunication, and software product markets[4]. However, Iyoha et al. (2022) document that product bundling is also prevalent in international markets. Specifically, they find that 37.76% of transactions are for bundles, accounting for 43.49% of import values in Columbia between 2015 and 2019. These findings are also present in the data sample I use for the main analysis. Specifically, out of 7630 firm-market-year pairs, most firms (96.73%) are multi-product firms for a given market and year. Among 222 firms that sell both CPUs (central processing unit) and GPUs (graphics processing unit) for a

[2]While the production side approach is widely used due to its simplicity and ease of data restrictions, there have been many challenges. For example, if the production data does not contain price information but only revenue, only revenue elasticities can be obtained instead of output elasticities. With the revenue elasticity, the expression for markups collapses to one; hence does not entail any information about markups. See Klette and Griliches (1996) and Bond et al. (2021). Also, there are discussions of identification issues where the markup is not identified using the proxy model (See Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015)) to estimate a production function and hence the output elasticity. See Flynn et al. (2019), Doraszelski and Jaumandreu (2019), and Jaumandreu (2018) for relevant discussions.

[3]See Zhou (2017), and Zhou (2021) for pure and mixed bundling practice in a competitive setting where there is an arbitrary number of firms and Chen and Riordan (2013) for general conditions for the profitability of product bundling where copula is used to model the stochastic dependence of consumer values.

[4]Regarding software, see *United States v. Microsoft Corporation, 253 F.3d 34* court case where the U.S government accused Microsoft of illegally maintaining its monopoly position primarily through bundling PC with Internet Explorer. Also, Crawford and Yurukoglu (2012) study the short-run welfare in the television channel market when á la carte policies which require distributors to offer individual channels for sale to consumers are introduced. Simulation results show increased input costs offset consumer benefits from purchasing individual channels.

given market in a year, 111 are engaged in bundling for CPU and GPU products. By recovering markups jointly for multi-product firms with mixed bundling practices, I recover markups at the transaction level and examine systematic differences in markups between firms with different pricing decisions. The empirical analysis for computer parts (CPU and GPU) for Chinese exporters from 2000 to 2006 shows that multi-product firms with product bundling enjoy markups which are roughly 30.6% higher compared to their counterparts without bundling practices. These differences in markups across multi-product firms with and without bundling practices may plausibly reveal how multi-product firms use bundling practices to retain their market power. Nevertheless, to my knowledge, there has yet to be a study of product bundling in an international trade setting to date.

Lastly, this paper adds to the literature studying the relationship between markups and competition in response to trade reform. Changes in market competitiveness forces firms to revisit pricing decisions, particularly when firms exert market power. De Loecker et al. (2016) study the impact of India's trade liberalization on markups, prices, and costs and found (i) incomplete pass-through of input cost declines to prices, and (ii) a pro-competitive effects on markups. However, in their setting, each product's markup was assumed to be *independent* of other products' markups even though most production occurred within multi-product firms. By recovering markups *jointly* for multi-product firms with product bundling, I determine how joint pricing affects firm profitability after trade liberalization. Empirical analysis shows that for computer parts, trade liberalization induced an increase in markups for all types of firms, but the magnitude was the highest for bundling firms. This shows that product bundling impacts trade policy changes and the pro-competitive effects of trade liberalization.

The structure of this paper is as follows. Section 2 describes the data sets used in the empirical analysis and China's WTO accession features for products of interest. Section 3 presents an empirical framework to recover markups using information from transactions and firm pricing decisions for both bundling and non-bundling firms. In section 4, empirical results are presented, and section 5 concludes.

2 Data and Trade Policy Background

I first describe the Chinese Customs data (CCD) in section 2.1, because it determines the base unit in which markups are recovered, how firms are classified into different types, and the product choice for the empirical analysis. Basic features of China's WTO accession, such as tariff changes are summarized in section 2.2.

2.1 *Data*

I take advantage of the Chinese Customs data that the Chinese Customs Office collects to explore the markup behavior across firms, time, and international markets. The CCD records Chinese firm-level exports and imports between 2000 and 2006 at the destination market-monthly level with corresponding HS6 codes, quantities, values, and firm characteristics such as firm names, firm ownership, addresses, and cities.

There are a few things to note about this data set. Firstly, because this is customs data, all the empirical analysis is focused on exporter firms and their export transactions.[5] Secondly, the framework to recover markups requires transaction data where ideally, one will have transactions recorded between each seller and buyer firm in a short period of time. While the frequency of CCD is at monthly level which is a good measure for international trade, there is an aggregation at the buyer side. This buyer side aggregation may lead to misclassifying multiple single goods transactions across different firms in a market into bundled good transactions from one buyer. To check this, I introduce additional data for capturing individual transactions between China and USA for the years 2004 and 2005.[6] Lastly, unlike production data where domestic and foreign quantities are aggregated, market-level transaction records allow me to incorporate demand-side characteristics into the framework and carry out the analysis by market. Hence, in the main empirical analysis, the markups are recovered at the firm-market-product-monthly

[5] De Loecker and Warzynski (2012) show that exporter firms, on average, have higher markups compared to domestic firms. However, if the difference is not systematically different across firms with different pricing strategies, then focusing on exporters will not create significantly different results compared to domestic firms.

[6] This data shows that most (81%) of transactions remain the same when seller firm-to-buyer transactions are aggregated to the seller firm-to-buyer market.

level and aggregated to various levels such as the firm-yearly level.

2.1.1 Price Imputation for Multi-product Firms

The framework to recover markups from the transaction side requires price information available to the buyer at the moment of the transaction. That is, while the price is observed for only products sold at a given transaction, prices of unsold products (including the bundle) for multi-product firms need to be imputed. These unobserved prices are imputed based on the firm's actual behaviors using the monthly feature of the Customs data. The key intuition is to impute the unobserved prices using the observed price data from the closest month.

Consider a benchmark case with two products, product 1 and product 2. For a firm f in market d, let (y_1, y_2) denote dummy variables for selling product 1 and product 2 and let (p_1, p_2) be the corresponding *observed* price for a transaction. Let (x_1, x_2, x_b, d) be the final *imputed* prices for product 1, product 2, both products combined, and bundling discount for the transaction that will be used to estimate consumer valuations and recover markups. The bundling discount is calculated as $d = (x_1 + x_2) - x_b$. If a transaction is a multi-product transaction, $(y_1, y_2) = (1, 1)$, with $d > 0$, then that transaction is classified as a transaction with product bundling. Transactions with either $(y_1, y_2) \neq (1, 1)$ or $d = 0$ are not classified as bundling transactions.[7]

Firstly, for a given transaction, if the price is observed, the imputed price is simply observed price itself, i.e., $x_j = p_j$. For example, in the case of $(y_1, y_2) = (1, 0)$, $x_1 = p_1$ and for $(y_1, y_2) = (1, 1)$, $x_b = p_1 + p_2$. If price is not observed for product j, this means $y_j = 0$ for $j = 1, 2$. Then for a given firm-market-year, I find the closest transaction where only $y_j = 1$ and $y_{-j} = 0$ where $-j$ denotes the other good. If there is no such transaction, I find the closest transaction with $(y_1, y_2) = (1, 1)$. I use the price from the closest transaction as the imputed price, i.e, $x_j = p_j^c$ where p_j^c denotes the price from the closest transaction. For $j = b$, I find the closest transaction where $(y_1, y_2) = (1, 1)$ and impute it as $x_b = p_1^b + p_2^b$ where p_j^b are prices from the closest multi-product transaction. If there are no transactions

[7]For example, a transaction where $(y_1, y_2) = (1, 0)$ with $d > 0$ is transaction where the buyer only buys product 1 even though there is a discount for a bundled product. On the other hand, a transaction where $(y_1, y_2) = (1, 1)$ but $d = 0$ is simply a transaction with multiple products and is not classified as a bundled transaction.

with $(y_1, y_2) = (1,1)$, simply set as $x_b = x_1 + x_2$.

Table 1 presents a basic example of the imputation procedure. To show the process more clearly, multi-product firms that sold both product 1 and product 2 to market d in a given year t are divided into five groups. Firms that have sold only good j or both goods are classified into group j with $j = 1, 2$. That is, firms in group 1 have the following; $(y_1, y_2) = \{(1,0), (1,1)\}$. Firms that have only single-product transactions, i.e., $(y_1, y_2) = \{(1,0), (0,1)\}$ are in group 3. Firms that only have multi-product transactions, $(y_1, y_2) = \{(1,1)\}$, are in group 4. Lastly, firms that have sold all composition of goods are classified into group 5, that is, they have $(y_1, y_2) = \{(1,0), (0,1), (1,1)\}$. Note that by construction, firms in group 3 and 4 can never be classified into bundling firms by design.[8] Out of five multi-product transactions, in this example only three are classified as a bundling transaction.

Table 1: Price Imputation Example

	Observed				*Imputed*				
	y_1	y_2	p_1	p_2	x_1	x_2	x_b	d	Bundle?
Group 1	1	0	80	-	80	100	170	10	No
	1	1	70	100	80	100	170	10	Yes
Group 2	0	1	-	120	70	120	170	20	No
	1	1	70	100	70	120	170	20	Yes
Group 3	1	0	80	-	80	120	200	0	No
	0	1	-	120	80	120	200	0	No
Group 4	1	1	60	100	60	100	160	0	No
	1	1	70	120	70	120	190	0	No
Group 5	1	0	80	-	80	120	170	30	No
	0	1	-	120	80	120	170	30	No
	1	1	70	100	80	120	170	30	Yes

Note: This table shows how price imputation for unobserved prices are carried out with a simple example. Firm-market-year pairs are grouped in to five different groups based on their transaction behavior, i.e, (y_1, y_2). A transaction is classified as a bundling if it have $(y_1, y_2) = (1,1)$ and $d > 0$.

2.1.2 Firm Type Definition and Data Descriptive

After unobserved prices are imputed, firms can be classified into single- and multi-product firms with and without bundling practices. This classification is based on the sales behavior

[8] Thus, this imputation constructs bundling transactions in a conservative way. Group 4 account for about 13% of firm-market pairs.

rather than the production behavior. Firstly, firms are categorized into single-, or multi-product firms depending on how many goods they sell to each destination market in a given year. For example, if a firm f produced multiple products but sold only a single HS6 code product to destination market d in year t, then firm f is classified as single-product firm in market d in year t. Once unobserved prices are imputed for multi-product firms that sold products of interest, multi-product firms are further divided into bundling firms and non-bundling firms based on whether there is a bundling transaction in market d in year t.

In this paper, a CPU and a GPU are selected for the analysis.[9] The products are selected based on the following criteria. Firstly, there have to be enough observations. Electrical machines are one of the most exported goods from China during the sample period. Also, the relationship between goods must be considered. Goods are chosen where firms are likely to produce both and sell them as a bundle. CPU and GPU are both processing units that are essential parts of a computer and frequently produced from the same manufacturers. A basic description of the data for CPU and GPU is summarized in Table 2.

Table 2: Summary Statistics

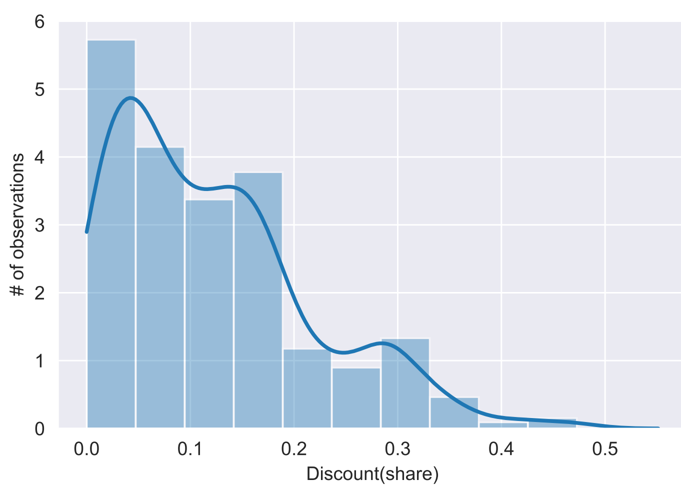
	Observation	(%)
Firm Characteristics(fdt)	7,630	100.0%
Single-product Firms	849	11.13%
Multi-product Firms	6,781	88.87%
selling either CPU or GPU	6,559	96.73%
selling both CPU and GPU	222	3.274%
with Product bundling	111	50.0%
without Product bundling	111	50.0%
Transaction Characteristics(fdm)		
Number of Transactions	51,882	100.0%
$MPT_{fdm} = 0$	49,787	97.06%
$MPT_{fdm} = 1$	2,095	2.94%
$Bundling_{fdm} = 1$ in $MPT_{fdm} = 1$	932	30.79%

Note: Following are explanations for the subscripts: f is for firms, d is for destination markets, j is for products(CPU and GPU), and m and t are time subscripts that each stand for month and year. MPT_{fdm} is a dummy that refers to multi-product transactions that consist of both CPU and GPU between firm f and market d for month m.

[9] The products are classified at the HS6 code level. Specifically, CPUs are {847130, 847141, 847149, 847150, 847160, 847170} and GPUs are {847180}.

In the upper panel, Table 2 shows that based on the transaction side classification, multi-product firms are the majority firm types in the CPU and GPU product market which aligns with findings from the multi-product firm literature. Specifically, out of 7,630 firm-market-year pairs, multi-product firms account for roughly 89% of the observations.[10] Out of those multi-product firms, 222 firms sold both CPUs and GPUs to market d in year t which is about 3% of the multi-product firms. As we increase the size of products of interest, the ratio of multi-product firms selling those goods will increase. Out of 222 firms, exactly 50% of firms engaged in product bundling with CPUs and GPUs. The bottom panel describes the baseline transactions, defined at firm-market-month level. There are total 51,882 transactions and out of those, roughly 3% transactions sold both CPUs and GPUs. Among transactions that sold both CPUs and GPUs, 30.79% of them are bundled transactions. Figure 1 plots the share of the bundling discounts for those bundling transactions. It displays a power law feature for the bundling discount where most of the discount is in between 5% to 20% of the original price.

Figure 1: Bundling discount share for bundled transactions



Note: This figure plots the bundling discount share which is bundling discount over the sum of each component product prices on the x-axis, i.e, $\frac{d_{fdbt}}{p_{fd1t}+p_{fd2t}}$. The y-axis shows normalized density for the number of observations.

[10] Out of 4,591 possible firm-importer pairs, only 4% of firms changed their status during the sample period for a given market, and majority of firms remained with their original type for a given market.

2.2 *WTO Accession and Tariff Reductions*

China's WTO accession, which took place in 2001, has included substantial tariff reductions (see Lu et al. (2015)). In this section, I document the impact of China's trade liberalization on CPU and GPU markups using tariff data from WITS database and trade values from UN COMTRADE. To examine the impact of tariff reductions and improved overall market access, I focus on the top 30 markets where China had the most transactions for CPUs and GPUs. They account for 95.5% of the quantity exported and 96.6% of trade value.[11]

Figure 2 displays the evolution of China's aggregated output and input tariffs.[12] For each market, output tariffs are at the HS2 code level. These output tariffs are aggregated using each market's trade value as weights to construct the aggregated output tariffs. For input tariffs, I follow De Loecker et al. (2016) and construct input tariffs for each market by passing the tariff data at ISIC Rev3 level to China's input-output matrix table for 1995-2010 and use value as weights to create the aggregated input tariff.[13] Figure 2 shows that trade liberalization brought a sharp decline in both output and input tariffs for CPU and GPU products. Specifically, the output tariff declined from roughly 2% to 0.8%, and the input tariff declined significantly from roughly 20% to 7%.

3 A Framework to Estimate Markups

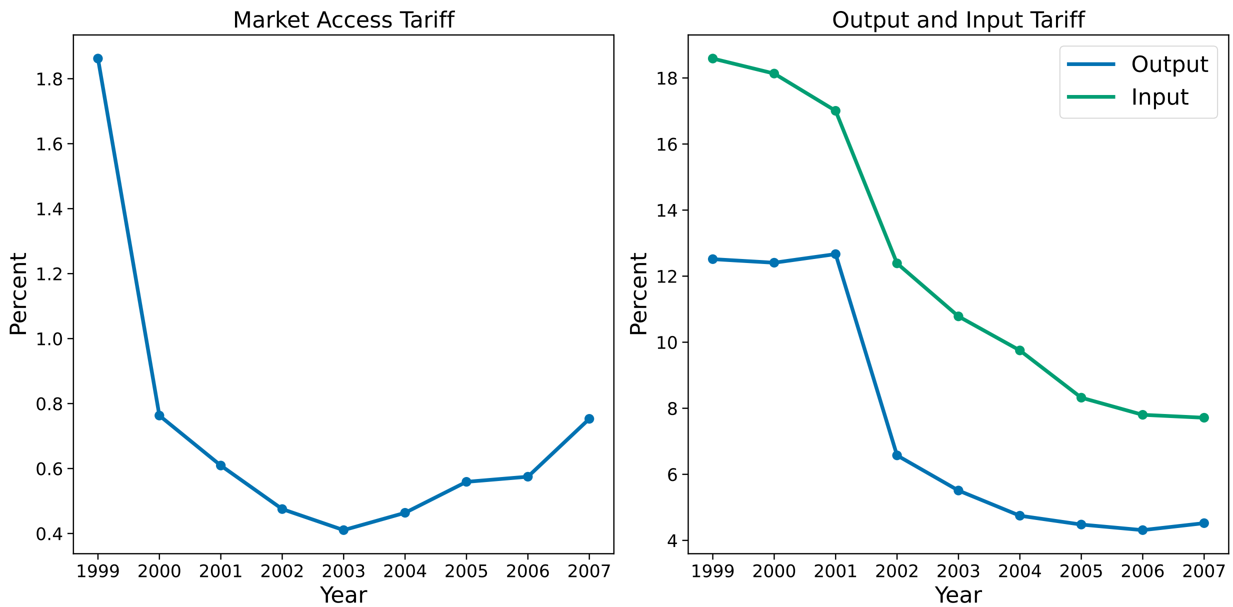
To incorporate joint markups and product bundling, I introduce an empirical model from the bundling literature. While non-bundling firms price goods independently, the bundling firm will choose prices for all of its single-product goods and bundles jointly to maximize its profit across all products simultaneously. Therefore, the firm's first-order conditions from

[11]These markets are Hong Kong, USA, Japan, Taiwan, Netherlands, Singapore, Germany, UK, South Korea, Australia, Malaysia, France, India, Thailand, UAE, Canada, Italy, Spain, Philippines, Brazil, Mexico, Belgium, South Africa, Israel, Turkey, Finland, New Zealand, Ireland, Indonesia, Poland in order of frequency. Aside from Taiwan who joined WTO alongside China, rest of 29 markets were all WTO members before China's WTO accession.

[12]Output and input tariffs for each market are displayed in the appendix.

[13]The formal definition of the input tariff is as follows. $\tau_{idt}^{input} = \sum_k a_{ki}\tau_{kdt}^{output}$ where τ_{kdt}^{output} is the export tariff for market d to China on industry k at time t and a_{ki} is the share of industry k in the value of industry i from the input-output table.

Figure 2: Tariffs for CPU and GPU from 1998 to 2007



Note: This figure plots aggregated market access, output and input tariffs for China from 1998 to 2007 for HS2 level 84, which contains CPU and GPU.

the profit-maximizing problem reflect information on independent markups for non-bundling firms and a markup linkage across all goods for bundling firms. This information from the FOCs is expressed in terms of consumers' valuations for the individual goods and the optimal price levels that the firm chose.

Firms choose optimal prices based on their marginal cost and demand. Using monthly transaction data and assumptions on the parametric structure for the consumer's valuations, marginal costs and consumer's valuations across goods are recovered from revealed information in the data. Once consumer valuations across goods are obtained, markups among firms that sell goods separately can be explicitly calculated, while joint markups among bundling firms are solved numerically.

For the rest of the paper, denote the set of individual goods (bundles) for firm f in year t as $\mathcal{G}_{ft}(\mathcal{B}_{ft})$ and let the number of components in the set be $G_{ft}(B_{ft})$ respectively. Let J_{ft} be the total number of products the multi-product firm sells, either as individual products or as a product bundle. For example, when firm f produces two discrete products and sells three products - both individual products and one product bundle of both single-products - we have the following : $J_{ft} = 3$, $\{1,2\} \in \mathcal{G}_{ft}$ and $\{b\} \in \mathcal{B}_{ft}$. Theoretically, for a total number of individual products G_{ft}, the number of possible bundles is at most, $\sum_{b=2}^{G_{ft}} \binom{G_{ft}}{b}$.

Let c_{fdjt} be a firm f's constant marginal cost for single product $j \in \mathcal{G}_{ft}$ in market d and year t.[14] The marginal cost of a bundle is the sum of the marginal costs of its single product components. The price of a bundle is potentially offered at a discount relative to the sum of its components.[15] For example, in the case of $G_{ft} = 2$, $c_{fdbt} = c_{fd1t} + c_{fd2t}$ and $P_{fdbt} = P_{fd1t} + P_{fd2t} - d_{fdbt}$ with $d_{fdbt} > 0$ where subscript b refers to a bundled product made of product 1 and product 2. Multi-product firms that do not engage in bundling practices could be interpreted as having $d_{fdbt} = 0$, effectively selling both goods simultaneously. Thus, while the subsequent discussion assumes multi-product firms bundle individual products, it could easily be applied to multi-product firms without bundling by setting $d_{fdbt} = 0$.

Assume consumers for each firm desire at most one unit of each good and demand each good independently of their consumption of the other goods.[16] For these consumers, consider the consumer valuations for G_{ft} goods $\mathbf{v}_{fdt} = (v_{fd1t}, ..., v_{fdG_{ft}t})$ which are distributed according to the unknown distribution function $\Psi_{fdt}(\mathbf{v}_{fdt})$.[17] Let $\psi_{fdt}(\mathbf{v}_{fdt})$ and $\psi_{fdkt}(v_{fdkt})$ be the probability density function and marginal density functions for product k for $\Psi_{fdt}(\mathbf{v}_{fdt})$. To avoid trivial cases, a positive measure of consumers exists such that $v_{fdjt} \geq c_{fdjt}$ for all j, and resale by consumers is not possible.

This paper will focus on a benchmark case where $G_{ft} = 2$ to build on key intuition as transparently as possible. Then, I outline how to generalize the estimation procedure for cases where $G_{ft} > 2$. Generalizing the estimation process for an arbitrary number of single products and bundles is a straightforward extension of the $G_{ft} = 2$ setting, albeit with

[14]The assumption that marginal costs are constant is needed to construct the marginal cost for the bundle. This assumption can be relaxed for firms that price goods separately to incorporate non-constant returns to scale.

[15]In this framework, bundling premium in which a bundle is offered at a higher price compared to the sum of its component goods is not considered. Intuitively, consumers always have the option to buy single product goods together rather than a bundle when there are mixed bundling practices.

[16]The unit demand assumption is relaxed in section 3.3 by utilizing quantity information from the transaction data.

[17]Consumer valuation distribution function Ψ can vary along various dimensions. The choice heavily depends on the number of observations in the data. In this paper, Ψ varies by firm, market, and year to capture demand characteristics at the firm, market, and year level.

substantially more derivations. In practice, bundled products do not typically contain many individual products, which eases the burden of derivation and any data restrictions.[18].

The subsequent section describes the framework for recovering markups with the transaction unit based on the Chinese Customs data; thus, the transaction is at the firm-market-monthly level. However, the transaction can be defined based on the available data.[19]

3.1 *Recovering Markups for Non-bundling Firms*

I first describe how to recover markups for firms that do not engage in bundling practices. Firms with independent pricing decisions maximize market-level profit by maximizing profits from each product independently.[20] Thus, the first-order conditions for each product-level profit entails information about *independent* product markups. The profit maximization problem for product j in market d for year t is

$$\underset{\mathbf{P}_{fdjm}}{\operatorname{argmax}} \Pi_{fdjt} = \underset{\mathbf{P}_{fdjm}}{\operatorname{argmax}} \sum_{m\in t} \Pi_{fdjm} = \underset{\mathbf{P}_{fdjm}}{\operatorname{argmax}} \sum_{m\in t} (P_{fdjm} - c_{fdjt}) Q^D_{fdjm} \tag{1}$$

where $Q^D_{fdjm}(P_{fdjm})$ is the quantity demanded for product j in market d at month m. Given the consumer valuations for product j in market d and year t, consumers whose valuation is higher than the price will purchase the good. Thus, $Q^D_{fdjm}(P_{fdjm}) = \int_{P_{fdjt}}^{\infty} \psi_{fdjt}(x)dx$.

Note that the quantity demanded for good j is only a function of good j characteristics such as price P_{fdjm} and its marginal distribution ψ_{fdjt} and does not depend on other products' characteristics. Then the first-order condition (2) gives the following equation in terms of the marginal density of valuations for product j, ψ_{fdjt}, monthly prices in year t, $P_{fjdm\in t}$, and

[18]Iyoha et al. (2022) find that most multi-product transactions have fewer than four products.

[19]The assumptions on the unit of the marginal costs, consumer valuations, and prices and hence the markups could be chosen appropriately depending on how detailed the 'transaction' is. For example, this paper defines the transaction as an exporter firm to the destination market at the monthly level. Thus, in this paper, the marginal costs can differ by destination market, i.e., c_{fdjt} to reflect shipping or market-specific marketing fees. However, it is not reasonable to assume that the marginal costs will differ at a monthly level; hence the time unit remains at a yearly level. Consumer valuations for each firm's residual demand also differs by destination market and year, i.e., Ψ_{fdt}, to reflect market-specific demand characteristics. The unit price for each product follows the unit of transactions, i.e., P_{fdjm} where m is for a month.

[20]In this paper, independent pricing firms include single- and multi-product firms without bundling practices. Markups for multi-product firms without bundling can be recovered using (1) the separate pricing method and (2) the product bundling method with $d_{fbt} = 0$, setting a discount equal to zero.

the marginal cost c_{fdjt}.[21] After the distribution for the consumer's valuations is estimated, equation (2) is used to recover the marginal cost c_{fdjt} and markups μ_{fjdm} for firms without bundling practices.

$$\sum_{m \in t} \Big(Q^D_{fdjm}(P_{fdjm}) - (P_{fdjm} - c_{fdjt})\psi_{fdjt}(P_{fdjm}) \Big) = 0 \tag{2}$$

Equation (2) shows the identification problem of previous methods in recovering joint markups for bundling firms. Once joint pricing decisions are incorporated, the number of unknown parameters (marginal costs and markups) increases with product size while the information (one first-order condition) stays the same.

3.2 *Recovering Joint Markups with Product bundling*

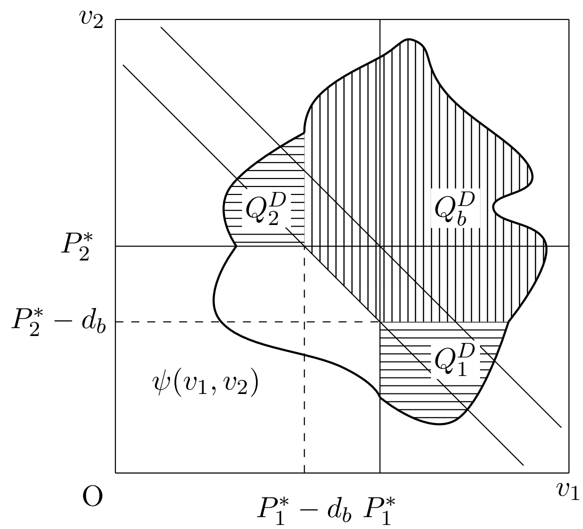
To identify joint markups, I introduce a framework from the bundling literature. The approach employs a model setting similar to McAfee et al. (1989) and Chen and Riordan (2013) in that consumer valuations are introduced to capture demand-side information. While their work focuses on finding the theoretical conditions in which it is more profitable for the firm to engage in product bundling, I focus on the joint pricing behavior of bundling firms and hence joint markups. Information regarding consumer taste is required to recover markups. I borrow the strategy to estimate consumer valuations from Letham et al. (2014), where variations in purchase behavior and prices are used.

3.2.1 $G_{ft} = 2$ *Case*

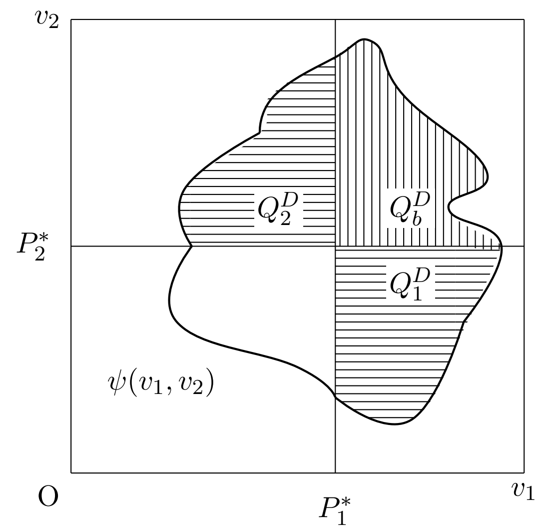
Since consumers are rational, a given consumer will purchase product k from a firm f only if it gives her the highest utility among all other options. This enables me to write the quantity demanded for each good j (Q^D_{fdjm}) in terms of prices and the distribution of

[21] The intuition for equation (2) is simple. A firm should choose a price such that marginal revenue of increasing 1 unit of price is equal to marginal cost of increasing 1 unit of price. If a firm increases the price by 1 unit, the firm will gain additional profit from existing customers ($1 \times Q^D_{fdjm}(P_{fdjm})$) and lose profits from customers who were on the margin ($(P_{fdjm} - c_{fdjt})\psi_{fdjt}(P_{fdjm})$).

Figure 3: Graphical illustrations : Joint Density Functions



(a) $G_{ft} = 2$ with bundling



(b) $G_{ft} = 2$ without bundling

Note: This figure graphically depicts the quantity demanded for each good j which depends on the joint density ψ and the price variables. The left panel depicts the case for a multi-product firm with product bundling and the right panel shows the quantity demanded for a multi-product firm without bundling. For a multi-product firm without bundling, Q^D_{fd1m} and Q^D_{fd2m} refer to the quantities demanded for only products 1 and 2, respectively, and Q^D_{fdbm} refers to the quantity demanded for both goods without discount. For both figures, subscripts f, d, and m or t are dropped for parsimony.

consumer valuations. For example, when $G_{ft} = 2$,

$$\begin{aligned}
&[Q^D_{fd1m} \text{ when}] \qquad v_{fd1t} - P_{fd1m} \geq max\{0, v_{fd1t} + v_{fd2t} - P_{fdbm}\} \\
&[Q^D_{fd2m} \text{ when}] \qquad v_{fd2t} - P_{fd2m} \geq max\{0, v_{fd1t} + v_{fd2t} - P_{fdbm}\} \qquad (3) \\
&[Q^D_{fdbm} \text{ when}] \qquad v_{fd1t} + v_{fd2t} - P_{fdbtm} \geq max\{0, v_{fd1t} - P_{fd1m}, v_{fd2t} - P_{fd2m}\}
\end{aligned}$$

Denote the vector of prices as $\mathbf{P}_{fdm} = (P_{fd1m}, P_{fd2m}, P_{fdbm})$ and combining each inequality and applying the definition of P_{fdbm} gives the following expressions for the quantity demanded in equation (4), which are graphically illustrated in the left panel of Figure 3. The consumer whose valuation falls in the area marked as Q^D_1, Q^D_2, or Q^D_b is going to buy good 1, good 2, or the bundled good, respectively. Note that the quantity demanded for single goods $j = 1, 2$ are functions of not only its price but also the price of a bundled good b, and hence the price of the other good, explicitly showing the linkage across goods for firms with joint pricing.

$$
\begin{aligned}
Q^D_{fd1m}(\mathbf{P}_{fdm}) &= \int_{P_{fd1m}}^{\infty} \int_{0}^{P_{fdbm}-P_{fd1m}} \psi_{fdt}(x,y)dydx \\
Q^D_{fd2m}(\mathbf{P}_{fdm}) &= \int_{0}^{P_{fdbm}-P_{fd2m}} \int_{P_{fd2m}}^{\infty} \psi_{fdt}(x,y)dydx \\
Q^D_{fdbm}(\mathbf{P}_{fdm}) &= \int_{P_{fd1m}}^{\infty} \int_{P_{fdbm}-P_{fd1m}}^{\infty} \psi_{fdt}(x,y)dydx + \int_{P_{fdbm}-P_{fd2m}}^{P_{fd1m}} \int_{P_{fdbm}-x}^{\infty} \psi_{fdt}(x,y)dydx
\end{aligned} \tag{4}
$$

These expressions for the quantities demanded can be plugged into the firm's profit maximization problem. The profit-maximizing firm will simultaneously choose all prices $\mathbf{P}_{fdm}$ to maximize its profit:

$$
\underset{\mathbf{P}_{fdm}}{\operatorname{argmax}} \, \Pi_{fdt} = \underset{\mathbf{P}_{fdm}}{\operatorname{argmax}} \sum_{m \in t} \Pi_{fdm} = \underset{\mathbf{P}_{fdm}}{\operatorname{argmax}} \sum_{m \in t} \Big(\Pi_{fd1m} + \Pi_{fd2m} + \Pi_{fdbm} \Big)
$$

$$
\text{where} \quad \Pi_{fdkm} = (P_{fdkm} - c_{fdkt}) Q^D_{fdkm}, \quad \text{for all } k \in \{1, 2, b\}
$$

and the analytical expression for $Q^D_{fdkm}(\mathbf{P}_{fdm})$ in terms of prices is derived from the rational consumer assumption as above. Thus, the profit function is as follows:

$$
\begin{aligned}
\Pi_{fdt} = &\sum_{m \in t} \Big(\Pi_{fd1m} + \Pi_{fd2m} + \Pi_{fdbm} \Big) \\
= &(P_{fd1t} - c_{fd1t}) \int_{P_{fd1t}}^{\infty} \int_{0}^{P_{fdbt}-P_{fd1t}} \psi_{fdt}(x,y)dydx + (P_{fd2t} - c_{fd2t}) \int_{0}^{P_{fdbt}-P_{fd2t}} \int_{P_{fd2t}}^{\infty} \psi_{fdt}(x,y)dydx \\
+ &(P_{fdbt} - c_{fd1t} - c_{fd2t}) \left[\int_{P_{fd1t}}^{\infty} \int_{P_{fdbt}-P_{fd1t}}^{\infty} \psi_{fdt}(x,y)dydx + \int_{P_{fdbt}-P_{fd2t}}^{P_{fd1t}} \int_{P_{fdbt}-s}^{\infty} \psi_{fdt}(x,y)dydx \right].
\end{aligned}
$$

The first-order conditions for price variables give the following three equations that express the relationship between marginal costs (hence markups) across products in terms of consumer valuation $\psi_{fdt}(v_{fd1t}, v_{fd2t})$ and price variables.[22]

$$
\begin{aligned}
&\sum_{m \in t} \Big(Q^D_{fd1m}(\cdot) - (P_{fd1m} - c_{fd1t}) \mathcal{A}_{fdm} + (P_{fd2m} - c_{fd2t} - d_{fdbm}) \mathcal{B}_{fdm} \Big) = 0 \\
&\sum_{m \in t} \Big(Q^D_{fd2m}(\cdot) - (P_{fd2m} - c_{fd2t}) \mathcal{C}_{fdm} + (P_{fd1m} - c_{fd1t} - d_{fdbm}) \mathcal{D}_{fdm} \Big) = 0 \qquad (5) \\
&\sum_{m \in t} \Big(Q^D_{fdbm}(\cdot) - (P_{fd1m} - c_{fd1t})(\mathcal{D}_{fdm} + \mathcal{E}_{fdm}) - (P_{fd2m} - c_{fd2t})(\mathcal{B}_{fdm} + \mathcal{E}_{fdm})
\end{aligned}
$$

[22] The derivation of these equations is included in the Appendix.

$$+d_{fbm}(\mathcal{B}_{fdm} + \mathcal{D}_{fdm} + \mathcal{E}_{fdm})\Big) = 0$$

where $\mathcal{A}_{fdm} = \int_0^{P_{fdbm}-P_{fd1m}} \psi_{fdt}(P_{fd1m}, y)dy$, $\mathcal{B}_{fdm} = \int_{P_{fd1m}}^{\infty} \psi_{fdt}(x, P_{fdbm} - P_{fd1m})dx$, $\mathcal{C}_{fdm} = \int_0^{P_{fdbm}-P_{fd2m}} \psi_{fdt}(x, P_{fd2m})dx$, $\mathcal{D}_{fdm} = \int_{P_{fd2m}}^{\infty} \psi_{fdt}(P_{fdbm} - P_{fd2m}, y)dy$, and lastly $\mathcal{E}_{fdm} = \int_{P_{fdbm}-P_{fd2m}}^{P_{fd1m}} \psi_{fdt}(x, P_{fdbm} - x)dx$.[23] Note that after estimating the consumer's valuation distribution $\psi_{fdt}(x, y)$, the first-order conditions provide the expression needed to identify joint markups.[24]

Denote equations from they system (5) as $\mathbf{\Gamma}(\mathbf{P}_{fdm}, \mathbf{Q}^D_{fdm}, \psi_{fdt}(\mathbf{v}_{fdt}); \mu_{fdj\in\{1,2\}m}) = 0$. Note $\mathbf{\Gamma}(\mu_{fdj\in\mathcal{G}_{ft}m}) = 0$, is a three ($J_{ft}$) by one vector of equations. Because we have two (G_{ft}) unknown joint markup parameters and three (J_{ft}) individual equations, it is over-determined. I propose to recover joint markups by solving $\mathbf{\Gamma}(\mu_{fdj\in\mathcal{G}_{ft}m}) = 0$ numerically and choosing the set of $\mu_{fdj\in\mathcal{G}_{ft}m}$ that minimizes the error below a given threshold level. The existence of a sufficiently small threshold level will filter out any cases where there is no solution for $\mu_{fdj\in\mathcal{G}_{ft}m}$.

3.2.2 *General Case with* $G_{ft} > 2$

Here I provide a general approach for deriving markup expressions across goods for cases where $G_{ft} > 2$. As noted before, once the number of single product goods exceeds two, the total number of possible combinations of single goods to make a bundled product becomes $\sum_{b=2}^{G_{ft}-1} \binom{G_{ft}}{b}$. This means even if firms have identical $\mathcal{G}_{ft}$, i.e., the same individual goods, they might have different bundled goods, i.e., different $\mathcal{B}_{ft}$. Thus, when G_{ft} exceeds two, I treat it as if all firms offer all possible combinations of a bundle. That is $B_{ft} = \sum_{b=2}^{G_{ft}-1} \binom{G_{ft}}{b}$

[23] For multi-product firms without bundling, taking first-order conditions with respect to only (P_{fd1t}, P_{fd2t}) or plugging in $d_{fdbt} = 0$ to equation (5) results in an identical result. In this case, joint pricing from product bundling is removed; hence, product markups are independent of one another as in previous literature.

$$Q^D_{fd1m}(P_{fd1m}, P_{fd2m}) + Q^D_{fdbm}(P_{fd1m}, P_{fd2m}) - (P_{fd1m} - c_{fdjt})(\mathcal{A}_{fdt} + \mathcal{D}_{fdt}) = 0$$
$$Q^D_{fd2m}(P_{fd1m}, P_{fd2m}) + Q^D_{fdbm}(P_{fd1m}, P_{fd2m}) - (P_{fd2m} - c_{fdjt})(\mathcal{B}_{fdt} + \mathcal{C}_{fdt}) = 0$$

The regression results in section 4 is robust to numerically estimating markups of multi-product firms without bundling by plugging in $d_{fdbt} = 0$.

[24] The intuition for the first-order conditions still holds just like it did for equation (2). If the firm increases the price for product 1 by 1 unit, it will gain additional profit from existing consumers from Q_{fd1m} and from consumers who were on the margin between Q_{fd1m} and Q_{fdbm} which is captured by the first and third terms. However, with the price increase, it will lose profit from consumers who were on the margin between Q_{fd1m} and not buying.

for all firms. Then the discount value for combinations of goods that are not bundled can be set at zero as in the case of multi-product firms without bundling practices.

Thus, if $G_{ft} > 2$, we follow the same steps as in the $G_{ft} = 2$ case. First, construct the following profit-maximizing problem for a firm f.

$$\underset{\mathbf{P}_{fdm}}{\text{argmax}}\, \Pi_{fdt}(\mathbf{P}_{fdm}) = \underset{\mathbf{P}_{fdm}}{\text{argmax}} \sum_{k \in \mathcal{G}_{ft} \cup \mathcal{B}_{ft}} \Pi_{fdkt} \tag{6}$$

where $\Pi_{fdkt} = \sum_{m \in t} (P_{fdkm} - c_{fdkt}) Q^D_{fdkm}$, for all $k \in \mathcal{G}_{ft} \cup \mathcal{B}_{ft}$. Second, using the rational consumer assumption, derive expressions for the quantity demanded, i.e., $Q^D_{fdkm}(\mathbf{P}_{fdm})$ for $k \in \mathcal{G}_{ft} \cup \mathcal{B}_{ft}$. Note that for $k \in \mathcal{G}_{ft}$, $Q^D_{fdkm}(\mathbf{P}_{fdm})$ should be expressed in terms of its price, and the price of bundled goods where k is a component of. For $k \in \mathcal{B}_{ft}$, $Q^D_{fdkm}(\mathbf{P}_{fdm})$ should be a function of its price and the price of all of the individual products of which bundle k is composed. After deriving expressions for the quantity demanded, plug it into the profit function to derive J_{ft} first-order conditions with joint markups. Denote it as $\mathbf{\Gamma}(\mathbf{P}_{fdm}, \mathbf{Q}_{fdm}, \psi_{fdt}(\mathbf{v}_{fdt}); \mu_{fdj \in \mathcal{G}_{ft} m}) = 0$ and recover joint markups numerically as in the $G_{ft} = 2$ case.

3.3 *Consumer Valuation Estimation*

This section describes how the consumer valuation distribution ψ_{fdt} is estimated using the approached proposed by Letham et al. (2014). The joint probability density function, $\psi_{ft}(\mathbf{v}_{fdt})$, describes how the consumer's valuation for each product is distributed as well as how it is correlated with the valuations of other products at the firm-year level. Because it differs at the firm-level, it can capture cross-sectional differences in firms, such as consumer types and quality (hence price). While making specific assumptions on the correlation structure across goods for $\psi_{fdt}(\mathbf{v}_{fdt})$ is possible[25], if the relationship between goods affects markups in a meaningful way, specific assumptions will likely distort the estimation of markups.[26] Using transaction data, Letham et al. (2014) propose a statistically consistent

[25] See Letham et al. (2014) for a survey of studies that made either independent or perfectly correlated assumptions on the correlation structure across goods.

[26] While it is described in terms of profits rather than markups, Letham et al. (2014) shows how imposing an independent correlation assumption could lead to very different predictions on possible profit when a bundled product is introduced.

inference procedure using copulas to recover correlated consumer valuations. The key intuition is to put a parametric assumption on the joint density function's marginal distributions and choose a specific copula function that will fit the overall correlation structure well. The marginal distribution will contain information on the valuation's marginal structure; hence, the demand for each product can be recovered from the marginal distribution afterward. After parameters for the marginal distributions are estimated, the copula parameter is estimated using these marginal parameters to fit the data in a maximum likelihood sense.

Define a transaction as a deal between seller and buyer during a certain period of time. In a retail setting where consumers buy goods often in small amounts, each day would be a good choice for the period of time. In trade, where buyer firms purchase goods in large amounts from specific sellers, monthly or yearly may be an adequate choice depending on the goods of interest. Consider a set of transaction data that consists of two components. One component is purchase data, $\mathbf{y}^s = [y_1^s, ..., y_{G_{ft}}^s]$, where y_j^s is 1 if item j is sold in transaction s and 0 otherwise.[27] The other component is the price data for individual products at transaction s, $\mathbf{P}^s = [P_1^s, ..., P_{G_{ft}}^s]$. Let S denote the total number of transactions. Since consumers maximize utility, $y_j^s = 1$ if and only if $v_j^s \geq P_j^s$. This relationship provides a model for the relationship between the latent variable valuations v_j^s and transaction data (y_j^s, P_j^s).

The copula $\mathbb{C}_{ft}(\cdot)$ for $\Psi_{ft}(\cdot)$ is a distribution function over $[0,1]^{G_{ft}}$ with uniform margins such that $\Psi_{ft}(v_{f1t}, ..., v_{fG_{ft}t}) = \mathbb{C}_{ft}(\Psi_{f1t}(v_{f1t}), ..., \Psi_{fG_{ft}t}(v_{fG_{ft}t}))$. The copula $\mathbb{C}_{ft}$ contains all information on the dependence structure between the components of $(v_{f1t}, ..., v_{fG_{ft}t})$ and combines each marginal distribution Ψ_{fkt} in a way to return the joint distribution Ψ_{ft}. Suppose each marginal distribution is a function of parameters $\boldsymbol{\theta}_{fjt}$, i.e, $\Psi_{fjt}(v_{fjt}; \boldsymbol{\theta}_{fjt})$ and the copula distribution belongs to a family with parameters $\boldsymbol{\phi}_{ft}$, i.e, $\Psi_{ft}(\boldsymbol{v}_{ft}; \boldsymbol{\theta}_{ft}, \boldsymbol{\phi}_{ft}) = \mathbb{C}_{ft}(\Psi_{f1t}(v_{f1t}; \boldsymbol{\theta}_{f1t}), ...\Psi_{fG_{ft}t}(v_{fG_{ft}t}; \boldsymbol{\theta}_{fG_{ft}t}); \boldsymbol{\phi}_{ft})$. Letham et al. (2014) propose an inference functions for margins (*IFM*) procedure that is similar to pseudo-maximum likelihood estimation where we choose parametric forms for the margins $\Psi_{fjt}(\cdot)$ and copula $\mathbb{C}_{ft}$, then find the parameters for which $\mathbb{C}_{ft}(\Psi_{f1t}(v_{f1t}), ..., \Psi_{fG_{ft}t}(v_{fG_{ft}t}))$

[27] The item j here is a unit product with a quantity equal to one. The unit demand assumption is relaxed by treating q units of a product sold as 1 unit of a product sold q times during the estimation procedure.

is the closest to $\Psi_{ft}(v_{f1t}, ...v_{G_{ft}t})$ in a likelihood sense.

The optimization can be performed in two steps. First, each marginal distribution is fit independently to recover $\hat{\theta}_{fjt}$. In the second step, the estimated marginal distributions are used to fit the correlation structure ϕ_{ft}.

$$\hat{\boldsymbol{\theta}}_{fjt} \in \underset{\boldsymbol{\theta}_{fjt}}{\operatorname{argmax}}\, l_{fjt}(\boldsymbol{\theta}_{fjt}) \quad j = 1, ..., G_{ft} \tag{7}$$

$$\hat{\boldsymbol{\phi}}_{ft} \in \underset{\boldsymbol{\phi}_{ft}}{\operatorname{argmax}}\, l_{ft}(\hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) \tag{8}$$

The likelihood function for each marginal distribution in equation (7) is derived from the observed purchase patterns of the utility-maximizing consumer. Let $\mathfrak{p}_{fj}(P_j^s)$ be the purchase probability for item j at price P_j^s which is equivalent to the demand model for item j. Then the demand and inverse marginal valuation distribution functions have the following relationship.

$$\mathfrak{p}_{fj}(P_j^s) = \mathbb{P}(y_j^s = 1) = \mathbb{P}(v_j^s > P_j^s) = 1 - \Psi_{fjt}(P_j^s; \boldsymbol{\theta}_{fjt})$$

Therefore, the likelihood function can be constructed employing the Bernoulli distribution for y_j^s such that $y_j^s \sim Bernoulli(1 - \Psi_{fjt}(P_j^s; \boldsymbol{\theta}_{fjt}))$ and resulting in the following likelihood function for given data $\{P_j^s, y_j^s\}_{s=1}^S$.

$$l_{fjt}(\boldsymbol{\theta}_{fjt}) = \sum_{s=1}^{S} (y_j^s \log(1 - \Psi_{fjt}(P_j^s; \boldsymbol{\theta}_{ft}))) + (1 - y_j^s) \log(\Psi_{fjt}(P_j^s; \boldsymbol{\theta}_{ft})) \tag{9}$$

The relationship between the marginal distribution and the demand model provides a natural selection criterion for the marginal distributions. For example, as Letham et al. (2014) has stated, if the demand model is linear, the corresponding valuation distribution is the uniform distribution. If the demand model follows the normal distribution function, the corresponding marginal valuation distribution also follows the normal distribution. For empirical analysis, I follow Letham et al. (2014) in using uniform distributions for the marginal distributions and a Gaussian copula function.

Once the marginal parameters $\boldsymbol{\theta}_{ft}$ are estimated by maximizing equation (9), these estimators are used to obtain an estimate of the copula parameters $\boldsymbol{\phi}_{ft}$ along with the data.

$$l_{ft}(\hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) = \sum_{s=1}^{S} \log \mathfrak{p}_f(\mathbf{y}^s | \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) \tag{10}$$

$$= \sum_{s=1}^{S} \log \int \mathfrak{p}_f(\mathbf{y}^s | \mathbf{v}^s, \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) \mathfrak{p}_f(\mathbf{v}^s | \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) d\mathbf{v}^s \tag{11}$$

$$= \sum_{s=1}^{S} \log \int_{v^{s,l}_{G_{ft}}}^{v^{s,u}_{G_{ft}}} \cdots \int_{v^{s,l}_{1}}^{v^{s,u}_{1}} \psi_{ft}(v^s_1, ..., v^s_{G_{ft}}; \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) dv^s_1 ... dv^s_{G_{ft}} \tag{12}$$

$$= \sum_{s=1}^{S} \log \sum_{k=0}^{G_{ft}} (-1)^k \sum_{I \subseteq \{1,...,G_{ft}\}, |I|=k} \Psi_{ft}(\mathbf{v}^s; \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) \tag{13}$$

where the equality from equations (11) to (12) uses $\mathfrak{p}_f(\mathbf{v}^s | \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) = \mathfrak{p}_f(\mathbf{v}^s | \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft}) = \psi_{ft}(\cdot; \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft})$ and make use of the lower and upper limits of the integration as follows:

$$v^{s,l}_j = \begin{cases} -\infty & \text{if} \quad y^s_j = 0 \\ P^s_j & \text{if} \quad y^s_j = 1 \end{cases} \qquad v^{s,u}_j = \begin{cases} P^s_j & \text{if} \quad y^s_j = 0 \\ \infty & \text{for} \quad y^s_j = 1 \end{cases}.$$

The representation of the likelihood formula in equation (12) is intractable due to multiple integrals. Letham et al. (2014) employ the rectangular integral of the probability density function to derive equation (13) where,

$$\tilde{v}^s_j(I) = \begin{cases} v^{s,l}_j & \text{if} \quad j \in I \\ v^{s,u}_j & \text{if} \quad j \notin I \end{cases}$$

Thus, the complete, statistically consistent inference procedure to estimate the consumer valuation distribution $\psi_{ft}(\mathbf{v})$ is,

$$\hat{\boldsymbol{\theta}}_{fjt} \in \underset{\boldsymbol{\theta}_{fjt}}{\text{argmax}} \sum_{s=1}^{S} (y^s_j \log(1 - \Psi_{fjt}(P^s_j; \boldsymbol{\theta}_f))) + (1 - y^s_j) \log(\Psi_{fjt}(P^s_j; \boldsymbol{\theta}_f))$$

$$\hat{\boldsymbol{\phi}}_{ft} \in \underset{\boldsymbol{\phi}}{\text{argmax}}_f \sum_{s=1}^{S} \log \sum_{k=0}^{G_{ft}} (-1)^k \sum_{I \subseteq \{1,...,G_{ft}\}, |I|=k} \Psi_{ft}(\tilde{\mathbf{v}}^s; \hat{\boldsymbol{\theta}}_{ft}, \boldsymbol{\phi}_{ft})$$

4 Empirical Analysis

In this section, I use the framework in section 3 to recover markups for Chinese exporters and test whether bundling firms, on average, have different markups. Also, I use China's

WTO accession and the accompanied reductions in tariffs to see whether trade liberalization differentially affects markups among bundling firms. To study the effect of bundling practices on markups, the important thing to note is that for a given firm, the decision to bundle or not purely depends on the dependence of consumer values summarized by the copula and not on marginal costs.[28]

After estimating the consumer's valuations that capture demand-side characteristics, markups for firms with and without bundling can be computed from the FOCs as described in the previous section. Recovered markup estimates and regression analysis reveal several major findings. First, I use the Chinese Manufacturing Data (*CMD*) to recover additional markups following De Loecker et al. (2016) method, and compare them to my markups and find that incorporating the bundling feature for multi-product firms may explain one important channel as to why multi-product firms have higher markups compared to single-product firms.[29] Second, I investigate the relationship between markups and firm types across markets and time. These analyses cannot be done using the previous methods where product bundling and joint pricing decisions were not incorporated into the estimation process.

4.1 Markup Descriptions

As described in section 3, markups for firms with independent pricing are calculated from first-order conditions derived from the profit maximization problem, while joint markups for the bundling firms are recovered numerically from the expressions for the first-order conditions.

Table 3 presents recovered markups across firm types at the firm-market-yearly level. Columns (1) and (2) report the mean and the standard deviations of markups for all firm types. On average, CPU and GPU products are priced about 70 to 80 % higher than their

[28]This is because the decision to bundle or not depends on a local perturbation of the optimal price from independent pricing firms. From this optimal independent price, which should already have taken marginal cost into account, the choice to bundle or not purely depends on whether the firm can attract additional purchases from the consumer by offering a slight discount, d_{fdbt}, on the bundled product. Thus, the marginal cost matters for the magnitude of additional profit from bundling but not for the decision to bundle. See Chen and Riordan (2013) for proof.

[29]Discussions on how markups from the production side using De Loecker et al. (2016) method are estimated and regression results are in the Appendix.

Table 3: Markup (μ_{fdjt}) Results

	Total		Single-		Multi- No Bundling		Multi- Bundling	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
CPU	0.74	0.64	0.73	0.57	0.74	0.65	0.82	0.49
GPU	0.80	0.67	0.78	0.65	0.80	0.69	0.87	0.62

Note: This table reports the average and median value of recovered markups by firm types. Markups are at the firm-market-product-year level. Here, the top and bottom 3% values are trimmed.

original cost.[30] Within the sample, mean values of recovered markups increase as we move from a single-product firm to a multi-product firm without bundling to a multi-product firm with bundling. This may indicate that product bundling is used by multi-product firms to increase their market power to price goods over their marginal costs, which was not captured by previous literature.

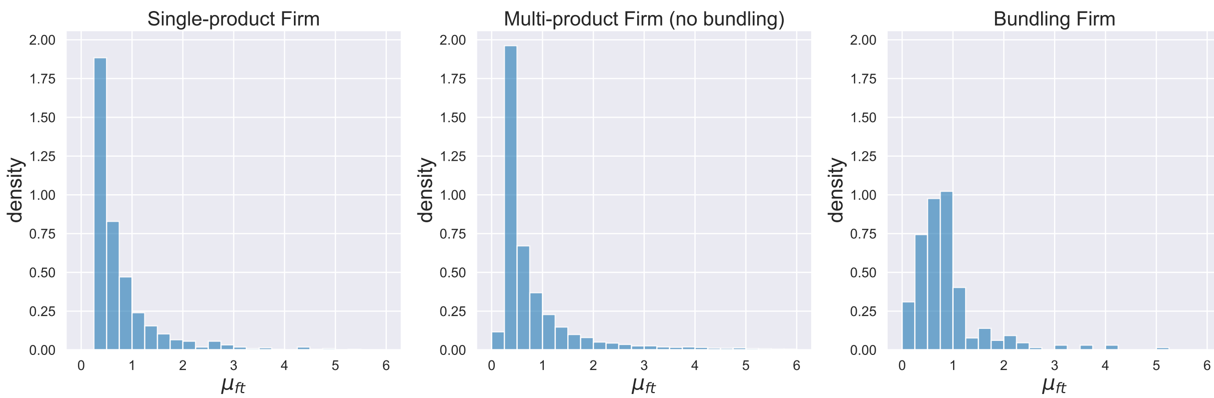
Figure 4 shows the histogram of recovered markups from the transaction side approach suggested by this paper and the production side approach from the previous literature.[31] Both approaches show some notable features. Firstly, the power law feature of the markup distribution can be seen across firm types and approaches. Also, for most cases, firm-level markups are below one, as shown in Table 3. However, Figure 4 shows that differences in markups across firm types are not fully captured on the production side where bundling firms are not present. Panel (a) shows that while the markup values of non-bundling firms present a high-peaked distribution at the lower levels of markups, the markup values of bundling firms are more dispersed and tilted toward the right, indicating that product bundling could affect markup distribution. These stark differences in markups across bundling and non-bundling firms cannot be seen from the production side in panel (b) but are attenuated to single- and multi-product firm differences.

[30]Note that a markup value of 0.7 means the firms are enjoying 70% of marginal cost as profit for each unit. For example, with a CPU of a marginal cost \$100, the price is set at \$170.

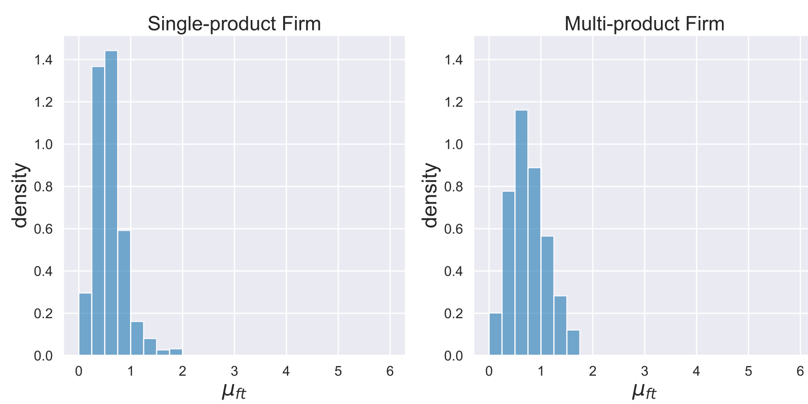
[31]Markups recovered using the lognormal distribution with the Gaussian copula instead of the uniform distribution with the Gaussian copula show similar results and are presented in the Appendix.

Figure 4: Markups (μ_{ft}) of CPU and GPU by firm-types

(a) Transaction Side



(b) Production Side



Note: Figure (a) plots the histogram of firm-year level markups for CPU and GPU that is recovered using the transaction side approach suggested by this paper. Figure (b) plots the histogram of firm-year level markups for CPU and GPU that is recovered using the production side approach. Note y-axis is normalized density and x-axis is $\mu_{ft} = \frac{P_{ft} - c_{ft}}{c_{ft}}$. Here, the top and bottom 3% values are trimmed for both cases.

4.2 Markups, Firm Heterogeneity and Trade Liberalization

The relationship between product bundling and markups depicted in Figure 4 may explain one additional channel as to why multi-product firms dominate international trade besides the productivity channel. To formally examine the effect of bundling on markups in international trade, I first analyze the effect of bundling on markups cross-sectionally, then across time, using China's WTO accession as a trade liberalization event. For the regression analyses, various levels of markups are used to see how not accounting for joint pricing decisions among bundling firms may lead to misleading or attenuated results.[32]

4.2.1 Markups and Firm Heterogeneity

I first study how markups differ across firm heterogeneity, such as single-, multi-, and multi-product firms with bundling using the following regression equation:

$$\log \mu_{fdjm} = \delta_{FE} + \delta_{MF_{fdt}} D_{MF_{fdt}} + \delta_{BF_{fdt}} D_{BF_{fdt}} + \varepsilon_{fdjm} \tag{14}$$

where $D_{MF_{fdt}}$ is a multi-product firm dummy, and $D_{BF_{fdt}}$ is a dummy variable for multi-product firms that engage in bundling practices. Both $D_{MF_{fdt}}$ and $D_{BF_{fdt}}$ vary by market-year, indicating that firm type is based on 'sales' side rather than 'production'. To capture any market specific, time, or ownership[33] trends, market, year, and firm ownership fixed effects are included. In this regression, $\delta_{MF_{fdt}}$ measures the percentage markup premium that a multi-product firm that does not engage in bundling has relative to single-product firms (i.e., "multi-product premium"). The percentage premium that the multi-product firm with bundling has over multi-product firms that do not engage in bundling (i.e., "bundling premium") will be captured by $\delta_{BF_{fdt}}$. Thus, $\delta_{MF_{fdt}} + \delta_{BF_{fdt}}$ measures the percentage premium of multi-product firms with bundling over single-product firms.

Table 4 shows the result of equation (14) at various levels of markups, and results align with economic intuition. Firstly, the multi-product firm dummy coefficients are significant

[32] The baseline markups are recovered at the firm-market-product-monthly level and aggregated to various levels of markups using values as weights. The regression results do not change much when quantities or simple averages are used as weights.

[33] Such as SOEs and private companies.

Table 4: Markups and Firm Heterogeneity

	(1) $log\mu_{fdjm}$	(2) $log\mu_{fdjt}$	(3) $log\mu_{fdt}$
$D_{MF_{fdt}}$	0.0921^{***}	0.0462^{**}	0.0859^{***}
(*Multi-product Firm Premium*)	(0.0092)	(0.0229)	(0.0228)
$D_{BF_{fdt}}$	0.0424^{***}	0.1229^{***}	0.2671^{***}
(*Product Bundling Premium*)	(0.0183)	(0.0436)	(0.0522)
Market FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Firm Type FE	Yes	Yes	Yes
F-statistic	27.74	7.701	11.98
Observation	36,076	7,852	7,459

Note: This table reports the coefficients from the regression (14). The dependent variable is (log) markup. Each column is an OLS regression result of log markup on firm heterogeneity for observations for CPU and GPU with various levels. Column (1) shows the results for baseline transactions, which is firm-market-product-monthly level, and column (2) and (3) show the results at the firm-market-product-yearly level and firm-market-yearly level. Finally column (4) shows the results at the firm-yearly level. The standard errors are in parenthesis and are bootstrapped. Significance : * 10 percent, ** 5 percent, *** 1 percent.

and positive at the firm-market-product-monthly level (column (1)), at the firm-market-product-yearly level (column (2)), and at the firm-market-yearly level (column (3)). While it is negative in column (4), where firm-yearly level markups are used, it is insignificant. This may partly be because the analysis is done with only the top 30 export markets, excluding other export markets and domestic market. Specifically, for baseline transactions in column (1), the multi-product firms without bundling practice have 9.65% higher markups than single-product firms on average and 8.97% higher markups at firm-market-yearly level in column (3).

The bundling dummy coefficients are all significant and positive across all specifications. For baseline transactions in column (1), bundling firms have, on average, 4.34% higher markups than multi-product firms without bundling practices and 14.4% higher markups than single-product firms. When monthly transactions are aggregated to the yearly level, the bundling dummy coefficient increases and shows that bundling firms, on average, have 13.1% higher markups than non-bundling multi-product firms. Similarly, at the firm-market-yearly level, multi-product firms with bundling practice enjoyed 30.6% higher markups compared to multi-product firms without bundling practice and 42.3% higher markups compared to single-product firms on average. At the firm-yearly level, multi-product firms with bundling goods have 21.4% higher markups than multi-product firms without bundling. In short, these results show that firms could potentially utilize product bundling to exercise market power and retain higher markups compared to others and highlight the importance of incorporating firm heterogeneity into consideration when examining markups.

Firms that engage in bundling practices price goods jointly; hence for these firms, studying markups at the product level, such as in columns (1) and (2), will not capture true market power. In columns (1) and (2), the *product bundling premium* associates product bundling practice with firms with market power (high markups). However, once we move from column (2) to (3), where markups are aggregated to firm-level and capture firm-level decisions, the large difference in *product bundling premium* from column (2) to column (3) shows that for multi-product firms, firm-level joint decisions such as product bundling may require analysis at both product and firm level to fully characterize their market power.

4.2.2 Bundling and Additional Sales

Findings from bundling literature indicate that firms engage in bundling practices to increase their profit by increasing the probability of selling additional products via a small discount on bundled products. To investigate if this is how bundling firms have higher markups than their counterparts, I follow the literature and run the following probit regression for only firms that sell both CPU and GPU.

$$MPT_{fdm} = \delta_{BF_{fdt}} D_{BF_{fdt}} + \delta_{\phi_{fdt}} \phi_{fdt} + X^T \beta + \epsilon_{fdm} \tag{15}$$

where MPT_{fdm} is a dummy equal to one when a transaction between firm f and market d in month m is a multi-product transaction with CPU and GPU. A positive value for coefficient $\delta_{BF_{fdt}}$ indicates that bundling firms do have a higher probability of selling goods together than others. I expect a positive value for $\delta_{\phi_{fdt}}$ since consumers who value CPUs are more likely value GPUs more and thus have a higher chance of buying both goods. Vector X includes parameters related to consumer valuations such as maximum and minimum ($\theta_{\mathbf{fdt}}$), and prices ($\mathbf{p_{fdm}}$).

Table 5: Probit Regression for Multi-product purchases

	MPT_{fdm}
$\delta_{BF_{fdt}}$	0.0999^{*}
	(0.052)
$\delta_{\phi_{fdt}}$	0.2147^{***}
	(0.0067)
Other Covariates	Yes
Log-Likelihood	-1667.5
Observation	2753

Note: The dependent variable is a dummy MPT_{fdm} which equals one when a given transaction have both CPU and GPU. Other covariates such as marginal parameters ($\boldsymbol{\theta_{fdt}}$) and prices ($\mathbf{p_{fdm}}$) are included. The standard errors are in parenthesis. Significance : * 10 percent, ** 5 percent, *** 1 percent.

Table 5 displays the results for coefficients of interest. First, the bundling firm coefficient is positive and significant at the 10% level. As indicated by the literature, bundling firms have a higher probability of selling more goods (both CPU and GPU) than others. Also, the correlation between the consumer's value for CPU and GPU captured by ϕ_{fdt} is positive and significant at the 1% level, as expected. This result shows that bundling firms with CPU and GPU are more likely to sell both goods than other multi-product firms, increasing profit. If markup differences are high between products, then bundling firms can increase profit and average firm-level markups by attracting more sales on the product with higher markups.

4.2.3 Markups and Trade Liberalization

As with static differences in markups, firm heterogeneity may also impact how firms and their markups react to trade liberalization. In this section, I inspect how firms' react to changes in market competitiveness induced by trade policy across firm types. Overall, we would expect to see markups for Chinese exporters increase after trade liberalization due to a decrease in tariffs, but these upward trends may differ across firm types (single- vs. multi- vs. multi- with bundling). To analyze this, I study the evolution of markups in response to changes in tariffs with following equation:

$$\begin{aligned}\log \mu_{fdjm} = \delta_{FE} + \delta_{to}\tau_{dt}^{to} + \delta_{to*MF_{fdt}}\tau_{dt}^{to}D_{MF_{fdt}} + \delta_{to*BF_{fdt}}\tau_{dt}^{to}D_{BF_{fdt}} \\ + \delta_{ti}\tau_{dt}^{ti} + \delta_{ti*MF_{fdt}}\tau_{dt}^{ti}D_{MF_{fdt}} + \delta_{ti*BF_{fdt}}\tau_{dt}^{ti}D_{BF_{fdt}} + \mathbf{x}'\beta + \varepsilon_{fdjm} \quad (16)\end{aligned}$$

where τ_{dt}^{to}, τ_{dt}^{ti} are output and input tariffs for each market at the yearly level, and δ_{FE} includes appropriate fixed effects for each level of analysis. The δ_{to} will capture the effect of a one unit change in output tariffs on markups for single-product firms, $\delta_{to}+\delta_{to*MF_{fdt}}$ on markups of multi-product firms without bundling, and $\delta_{to}+\delta_{to*MF_{fdt}}+\delta_{to*BF_{fdt}}$ on multi-product firms with bundling. A similar interpretation holds for δ_{ti}.

Table 6: Markups and Trade Liberaliztion : Tariff Changes

	(1)	(2)	(3)
	$log\mu_{fdjm}$	$log\mu_{fdjt}$	$log\mu_{fdt}$
τ_{dt}^{to}	-0.0168^{***}	-0.0040	-0.0023
	(0.0046)	(0.0111)	(0.0110)
$\tau_{dt}^{to}D_{MF_{fdt}}$	0.0037	-0.0032	-0.0056
	(0.0032)	(0.0081)	(0.0080)
$\tau_{dt}^{to}D_{BF_{fdt}}$	-0.0128^{***}	-0.0273	-0.0259
	(0.0045)	(0.0155)	(0.0270)
τ_{dt}^{ti}	-0.0264^{***}	-0.0058^{*}	-0.0137^{***}
	(0.0012)	(0.0032)	(0.0032)
$\tau_{dt}^{ti}D_{MF_{fdt}}$	0.0084^{***}	0.0064^{***}	0.0093^{***}
	(0.0008)	(0.0023)	(0.0022)
$\tau_{dt}^{ti}D_{BF_{fdt}}$	0.0116^{***}	0.0208^{***}	0.0400^{***}
	(0.0021)	(0.0053)	(0.0062)
Market FE	Yes	Yes	Yes
Firm Type FE	Yes	Yes	Yes
F-statistic	79.06	3.834	9.979
Observation	36,076	7,852	7,459

Note: This table reports the coefficients from the regression (16). The dependent variable is (log) markup. Each column is an OLS regression result of log markup on firm heterogeneity for observations for CPU and GPU with various levels. Column (1) shows the results for baseline transactions, and column (2) shows the results at the firm-destination market-yearly level. Finally, column (3) shows the results at the firm-yearly level. The standard errors are in parenthesis and are bootstrapped. Significance : * 10 percent, ** 5 percent, *** 1 percent.

Table 6 presents results at various levels of markups. Because both output and input tariffs decreased, the negative sign on the coefficient corresponds to an increase in markups. The analysis with baseline transaction level in column (1) shows all coefficients to be significant except for the output tariff and multi-product firm without bundling cross-term. Specifically, when output tariffs decreased by 1 unit, markups increased by 1.69% for single-product firms and 2.61% for multi-product firms with product bundling on average.

When input tariffs decreased by 1 unit, markups increased by 2.67% for single-product firms, 1.81% for multi-product firms without bundling, and 0.64% for bundling firms.

Note that output tariff changes exacerbate markup differences across firm types. In contrast, input tariff changes mitigate the markup differences across firm types. Since the magnitude of tariff reduction is much higher for input tariffs than output tariffs, tariff reductions reduce markup dispersion across firm types, as Lu et al. (2015) find. In column (1), the pro-competitive effect happens by less cost-efficient firms becoming more cost-efficient rather than higher markup firms losing their markups. However, once the unit of analysis becomes more aggregated, output tariff reductions do not affect markups, but input tariffs continue to have a significant impact on markups, as De Loecker et al. (2016) suggest. At the firm-importer-yearly level, a decrease in input tariff by 1 unit results in 1.38% higher markups for single-product firms and 0.44% for multi-product firms without bundling. However, markups for bundling firms decrease by 3.49%. All tariff coefficients become insignificant when markups and tariff levels are aggregated into firm-yearly and yearly levels due to variation loss.

5 Conclusion

Recently, firm-level analysis has been a central focus of understanding international trade, e.g., multi-product firms, productivity, networks, and markups. In this paper, I look at an important source of firm heterogeneity that has been overlooked-a multi-product firm's ability to offer product bundles-and investigate whether the effects of the trade liberalization on markups differ across firm types.

In the empirical estimation, I estimate markups using transaction data and a framework that explicitly incorporates multi-product firms' joint pricing decisions, which is missing in the previous literature. Comparing estimated markups for Chinese exporters to markups recovered using De Loecker and Warzynski (2012) method shows that the production-side approach may miss one important channel (*bundling*) why multi-product firms have higher markups than single-product firms. While the main analysis focuses on two product cases for Chinese exporters, the method can be generalized to many products and individual firms in any market with market power.

My study also contributes to the literature on the relationship between markups and trade characteristics. While the previous study has focused on gains from trade at the aggregate level, I study how these effects may differ across firms depending on their type. Different magnitudes of the trade liberalization effect on markups show that policy changes such as tariffs and analysis on the pro-competitive effects of trade liberalization should account for firm heterogeneity.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. (2015) "Identification Properties of Recent Production Function Estimators." *Econometrica*, 83, 2411–2451.

Adams, William James, and Janet L. Yellen. (1976) "Commodity Bundling and the Burden of Monopoly." *The Quarterly Journal of Economics*, 90, 475–498.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott. (2012) "The Empirics of Firm Heterogeneity and International Trade." *Annual Review of Economics*, 4, 283–313.

Bernard, A. B., S. J. Redding, and P. K. Schott. (2011) "Multiproduct Firms and Trade Liberalization." *The Quarterly Journal of Economics*, 126, 1271–1318.

Berry, Steven, James Levinsohn, and Ariel Pakes. (1995) "Automobile Prices in Market Equilibrium." *Econometrica*, 63, 841–890.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. (2021) "Some Unpleasant Markup Arithmetic: Production Function Elasticities and Their Estimation from Production Data." *Journal of Monetary Economics*, 121, 1–14.

Chen, Yongmin, and Michael H. Riordan. (2013) "Profitability Of Product Bundling." *International Economic Review*, 54, 35–57.

Crawford, Gregory S, and Ali Yurukoglu. (2012) "The Welfare Effects of Bundling in Multichannel Television Markets." *American Economic Review*, 102, 643–685.

De Loecker, Jan, Pinelopi Koujianou Goldberg, Amit Khandelwal, and Nina Pavcnik. (2016) "Prices, Markups and Trade Reform." *Econometrica*, 84, 445–510.

De Loecker, Jan, and Frederic Warzynski. (2012) "Markups and Firm-Level Export Status." *The American Economic Review*, 102, 2437–2471.

Doraszelski, Ulrich, and Jordi Jaumandreu. (2019) "Using Cost Minimization to Estimate Markups." *mimeo.*

Flynn, Zach, Amit Gandhi, and James Traina. (2019) "Measuring Markups with Production Data." *mimeo.*

Iyoha, Ebehi, Ji Hye Heo, and Joel Rodrigue. (2022) "Product Bundling in International Markets." *mimeo.*

Jaumandreu, Jordi. (2018) "Dangerous Shortcuts: Ignoring Marginal Cost Determinants in Markup Estimation." *mimeo.*

Klette, Tor Jakob, and Zvi Griliches. (1996) "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous." *Journal of Applied Econometrics*, 11, 343–361.

Letham, Benjamin, Wei Sun, and Anshul Sheopuri. (2014) *Latent Variable Copula Inference for Bundle Pricing from Retail Transaction Data.* PMLR.

Levinsohn, James, and Amil Petrin. (2003) "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies*, 70, 317–341.

Lopresti, John. (2016) "Multiproduct firms and product scope adjustment in trade." *Journal of International Economics*, 100, 160–173.

Lu, Yi, Linhui Yu, and Linhui Yu. (2015) "Trade Liberalization and Markup Dispersion: Evidence from China's WTO Accession." *American Economic Journal: Applied Economics*, 7, 221–253.

Manova, Kalina, and Zhihong Yu. (2017) "Multi-product firms and product quality." *Journal of International Economics*, 109, 116–137.

McAfee, R. Preston, John McMillan, and Michael D. Whinston. (1989) "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values." *The Quarterly Journal of Economics*, 104, 371–383.

Olley, G. Steven, and Ariel Pakes. (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64, 1263–1297.

Stigler, George J.. (1963) “United States v. Loew’s Inc.: A Note on Block-Booking.” *The Supreme Court Review*, 1963, 152–157.

Zhou, Jidong. (2017) “Competitive Bundling.” *Econometrica*, 85, 145–172.

Zhou, Jidong. (2021) “Mixed Bundling in Oligopoly Markets.” *Journal of Economic Theory*, 194, 105257.

Appendix

A. First Order Conditions from the Profit Maximization

A.1 Derivation of the expressions

Here I demonstrate steps for calculating the first order conditions in Section 2. Recall the profit function was

$$\begin{aligned}\Pi_{ft}(\mathbf{P}_{ft}) &= (P_{f1t}-c_{f1t})Q^D_{f1t}(\mathbf{P}_{ft}) + (P_{f2t}-c_{f2t})Q^D_{f2t}(\mathbf{P}_{ft}) + (P_{fbt}-c_{f1t}-c_{f2t})Q^D_{fbt}(\mathbf{P}_{ft})\\ &=(P_{f1t}-c_{f1t})\int_{P_{f1t}}^{\infty}\int_{0}^{P_{fbt}-P_{f1t}}\psi_f(x,y)dydx + (P_{f2t}-c_{f2t})\int_{0}^{P_{fbt}-P_{f2t}}\int_{P_{f2t}}^{\infty}\psi_f(x,y)dydx\\ &+(P_{fbt}-c_{f1t}-c_{f2t})\left[\int_{P_{f1t}}^{\infty}\int_{P_{fbt}-P_{f1t}}^{\infty}\psi_f(x,y)dydx + \int_{P_{fbt}-P_{f2t}}^{P_{f1t}}\int_{P_{fbt}-x}^{\infty}\psi_f(x,y)dydx\right]\end{aligned}$$

Taking derivative with respect to p_{f1t} results in the following equation.

$$\begin{aligned}&\int_{P_{f1t}}^{\infty}\int_{0}^{P_{fbt}-P_{f1t}}\psi_f(x,y)dydx + (P_{f1t}-c_{f1t})\frac{\partial}{\partial P_{f1t}}\Big[\int_{P_{f1t}}^{\infty}\int_{0}^{P_{fbt}-P_{f1t}}\psi_f(x,y)dydx\Big]\\ &+(P_{fbt}-c_{f1t}-c_{f2t})\frac{\partial}{\partial P_{f1t}}\Big[\int_{P_{f1t}}^{\infty}\int_{P_{fbt}-P_{f1t}}^{\infty}\psi_f(x,y)dydx + \int_{P_{fbt}-P_{f2t}}^{P_{f1t}}\int_{P_{fbt}-x}^{\infty}\psi_f(x,y)dydx\Big] = 0\end{aligned}$$

Note that the first term corresponds to $Q^D_{f1t}(\mathbf{P}_{ft})$. For the second term, denote $G_{f1t}(P_{f1t}, P_{fbt}, x) = \int_0^{P_{fbt}-P_{f1t}}\psi(x,y)dx$ and $H_{f1t}(P_{f1t}, P_{fbt}) = \int_{P_{f1t}}^{\infty} G_{f1t}(P_{f1t}, P_{fbt}, y)dy$. Then applying the Leibniz rule gives the following for the second term.

$$\begin{aligned}\frac{\partial}{\partial P_{f1t}}H_{f1t}(P_{f1t},P_{fbt}) &= -G_{f1t}(P_{f1t},P_{fbt},P_{f1t}) + \int_{P_{f1t}}^{\infty}\frac{\partial}{\partial P_{f1t}}G_{f1t}(P_{f1t},P_{fbt},x)dx\\ &= -\int_0^{P_{fbt}-P_{f1t}}\psi_f(P_{f1t},y)dy - \int_{P_{f1t}}^{\infty}\psi_f(x,P_{fbt}-P_{f1t})dx\end{aligned}$$

Similarly, for the third in the first order condition, let $G_{f2t}(P_{fbt}, x) = \int_{P_{fbt}-P_{f1t}}^{\infty}\psi_f(x,y)dy$ with $H_{f2t}(P_{f1t}, P_{fbt}) = \int_{P_{f1t}}^{\infty} G_{f2t}(P_{f1t}, P_{fbt}, y)dy$, and for the fourth term let $G_{f3t}(P_{fbt}, y) = \int_{P_{fbt}-y}^{\infty}\psi_f(x,y)dy$ with $H_{f3t}(P_{f1t}, P_{f2t}, P_{fbt}) = \int_{P_{fbt}-P_{f2t}}^{P_{f1t}} G_{f3t}(P_{fbt}, y)dy$.

Then taking derivative following the Leibniz rule gives the following expressions.

$$\begin{aligned}
\frac{\partial}{\partial P_{f1t}} H_{f2t}(P_{f1t}, P_{fbt}) &= -G_{f2t}(P_{f1t}, P_{fbt}, P_{f1t}) + \int_{P_{f1t}}^{\infty} \frac{\partial}{\partial P_{f1t}} G_{f2t}(P_{f1t}, P_{fbt}, y) dy \\
&= -\int_{P_{fbt}-P_{f1t}}^{\infty} \psi_f(P_{f1t}, y) dy + \int_{P_{f1t}}^{\infty} \psi_f(x, P_{fbt} - P_{f1t}) dx \\
\frac{\partial}{\partial P_{f1t}} H_{f3t}(P_{f1t}, P_{f2t}, P_{fbt}) &= G_{f3t}(P_{fbt}, P_{f1t}) \\
&= \int_{P_{fbt}-P_{f1t}}^{\infty} \psi_f(P_{f1t}, y) dy
\end{aligned}$$

Plugging these terms into the original f.o.c and using the definition of $Q_{f1t}(\mathbf{P}_{ft})$ gives,

$$\begin{aligned}
&Q^D_{f1t}(\mathbf{P}_{ft}) - (P_{f1t} - c_{f1t})\Big[\int_0^{P_{fbt}-P_{f1t}} \psi(P_{f1t}, y) dy + \int_{P_{f1t}}^{\infty} \psi_f(x, P_{fbt} - P_{f1t}) dx\Big] \\
&+ (P_{fbt} - c_{f1t} - c_{f2t})\Big[-\int_{P_{fbt}-P_{f1t}}^{\infty} \psi_f(P_{f1t}, y) dy + \int_{P_{f1t}}^{\infty} \psi_f(x, P_{fbt} - P_{f1t}) dx + \int_{P_{fbt}-P_{f1t}}^{\infty} \psi_f(P_{f1t}, y) dy\Big] \\
&= Q^D_{f1t}(\mathbf{P}_{ft}) - (1 - \mu^{-1}_{f1t}) P_{f1t} \int_0^{P_{fbt}-P_{f1t}} \psi_f(P_{f1t}, y) dy + [(1 - \mu^{-1}_{f2t}) P_{f2t} - d_{fbt}] \int_{P_{f1t}}^{\infty} \psi_f(x, P_{fbt} - P_{f1t}) dx
\end{aligned}$$

where the equality comes from $(P_{f1t} - c_{f1t}) = (P_{f1t} - c_{f1t})\frac{P_{f1t}}{P_{f1t}} = (1 - \mu^{-1}_{f1t})P_{f1t}$ and $(P_{f1t} + P_{f2t} - c_{f1t} - c_{f2t} - d_{fbt}) = (P_{f1t} - c_{f1t})\frac{P_{f1t}}{P_{f1t}} + (P_{f2t} - c_{f2t})\frac{P_{f2t}}{P_{f2t}} - d_{fbt} = (1 - \mu^{-1}_{f1t})P_{f1t} + (1 - \mu^{-1}_{f2t})P_{f2t} - d_{fbt}$. Derivatives with respect to P_{f2t} and P_{fbt} are similar thus omitted.

B. Regression Analysis for Production Side Markups

B.1 Framework from the Production Side

In this section, I describe how markups using production data (section 4) were estimated. These production side markups are recovered following method directly from De Loecker et al. (2016) and the Chinese Manufacturing data. Consider following production function for firm f producing product j at time t :

$$Q^s_{fjt} = F_{jt}(\mathbf{V}_{fjt}, \mathbf{K}_{fjt})\Omega_{ft}$$

where Q^s denotes the physical output (the quantity) of product j produced by firm f at time t. $\mathbf{V}$ denotes a vector of variable inputs that the firm can freely adjust, such as materials and $\mathbf{K}$ is a vector of fixed inputs with adjustment costs such as labor and capital. Combine inputs into a vector $\mathbf{X} = \{\mathbf{V}, \mathbf{K}\}$, and denote the price of input v is denoted as W^v_{fjt}. The productivity of firm f at time t is denoted as Ω_{ft}. Lower case variables indicate the log terms of their capital counterparts. Then, firm's cost minimization problem results in the following expression for markups at the firm-product-year level.

$$\mu_{fjt} = \theta^v_{fjt}\left(\frac{P_{fjt}Q^s_{fjt}}{W^v_{fjt}V^v_{fjt}}\right) = \theta^v_{fjt}(\alpha^v_{fjt})^{-1} \tag{17}$$

where θ^v_{fjt} refers to product j's output elasticity of flexible input v.

I use single-product firms to estimate the production function and the output elasticity θ^v_{fjt} as suggested by De Loecker et al. (2016). To account for the endogeneity issue caused by unobserved productivity terms, the control function approach suggested by Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015) is used to estimate the production function. To account for bias caused by using only single-product firms, I apply a sample selection correction procedure following Olley and Pakes (1996) and De Loecker et al. (2016).

The basic idea of the control function approach is to come up with an equation for the unobserved term ω_{ft} that can be used to eliminate endogeneity bias. Following De Loecker et al. (2016), I employ the Ackerberg et al. (2015) method on single-product firms to estimate production functions. Specifically, I use the material demand function to come up with an equation for ω_{ft}. Assume the material demand function for single-product firm producing good j is

$$m_{ft} = m_t(\omega_{ft}, k_{ft}, l_{ft}, \mathbf{z}_{ft}) \tag{18}$$

where $\mathbf{z}_{ft} = \{\mathbf{L}_f, P_{ft}, EXP_{ft}, \tau^{output}_{it}, \tau^{input}_{it}\}$ with $\mathbf{L}_f$ is firm-specific exogenous factors such as age, location, ownership status, and affiliation status, $\mathbf{EXP}_{ft}$ is export dummy, and $\tau^{output}_{it}, \tau^{input}_{it}$ are output and import tariffs for industry i.

Inverting equation (18) gives the control function for the unobserved productivity ω_{ft} as the following.

$$\omega_{ft} = h_t(\mathbf{x}_{ft}, \mathbf{z}_{ft})$$

To construct the moment conditions, consider the following law of motion for productivity.

$$\omega_{ft} = \eta(\omega_{ft-1}, EXP_{ft-1}, \tau_{it-1}^{output}, \tau_{it-1}^{input}, SP_{ft}) + \xi_{ft} \tag{19}$$

where ξ_{ft} denotes the unexpected innovation to productivity and SP_{ft} is included in the law of motion to correct for selection bias for using only single-product firms.

In the first step to estimate production function, I separate the unanticipated shocks and/or measurement error term ϵ_{fjt} from the rest of the terms that are known to the firm.

$$q_{fjt} = \phi_{jt}(\mathbf{x}_{ft}, \mathbf{z}_{ft}) + \epsilon_{fjt} \tag{20}$$

where $\phi_{jt}(\cdot)$ is equal to $f_j(\mathbf{x}_{ft}; \boldsymbol{\beta}) + \omega_{ft}$. This allows us to express productivity ω_{ft} as a function of data and predicted output $\hat{\phi}_{fjt}$ from the first step.

$$\omega_{ft}(\boldsymbol{\beta}) = \hat{\phi}_{fjt} - f_j(\mathbf{x}; \boldsymbol{\beta})$$

Combining this with the law of motion for productivity in equation (19), we can recover the innovation term ξ_{ft} by

$$\xi_{ft}(\boldsymbol{\beta}) = \omega_{ft}(\beta) - E[\omega_{ft}(\boldsymbol{\beta}) | \omega_{ft-1}(\boldsymbol{\beta}), EXP_{ft-1}, \tau_{it-1}^{output}, \tau_{it-1}^{input}, SP_{ft}] \tag{21}$$

Then, the moment conditions in the second step that identify the parameters of the production function are

$$E[(\xi_{ft}(\boldsymbol{\beta}) + \epsilon_{fjt})\boldsymbol{Y}_{ft}] = 0 \tag{22}$$

where $\boldsymbol{Y}_{ft}$ contains all the variables that are in the firm's information set at time t such as lagged materials, current predetermined capital, labor, and their higher order interaction terms, as well as lagged output prices, lagged tariffs, and their appropriate interactions with the inputs.

B.2 Regression results using markup from the Production Side

Table 7: Regressions with Markups from the Production Side

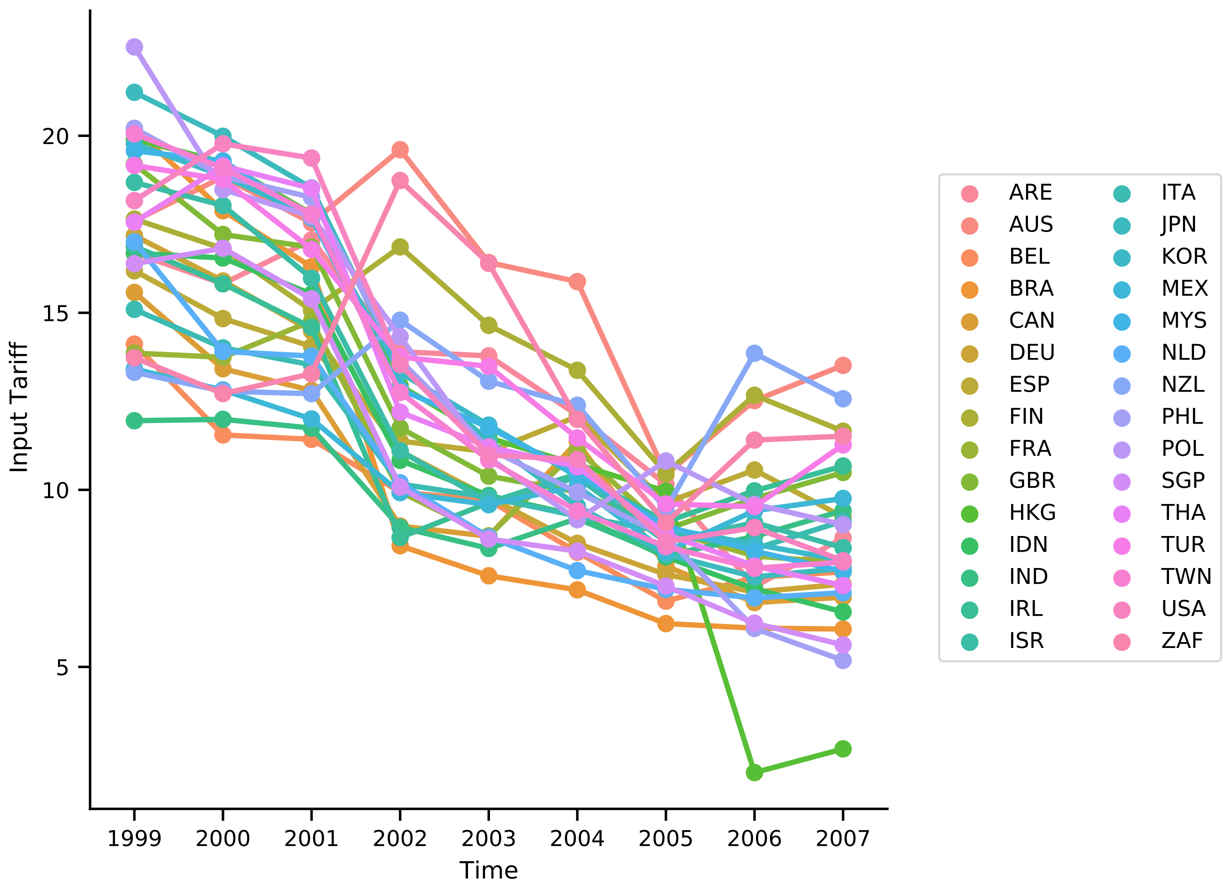
	$log\mu_{ft}$				
	(1)		(2)		(3)
$D_{MF_{ft}}$	0.0801	D_{TL}	0.0871**	τ_{dt}^{to}	-0.0193
	(0.0479)		(0.0386)		(0.1073)
		$D_{TL}D_{MF_{ft}}$	-0.0198	$\tau_{dt}^{to}D_{MF_{fdt}}$	0.2288**
			(0.0473)		(0.1158)
				τ_{dt}^{ti}	-0.2558***
					(0.0713)
				$\tau_{dt}^{ti}D_{MF_{fdt}}$	0.0887***
					(0.0310)
Firm FE	Yes	Firm FE	Yes	Firm FE	Yes
Year FE	Yes	Year FE	No	Year FE	No
F-statistic	1.13e-05	F-statistic	46.81	F-statistic	34.50
Observation	1,137	Observation	1,137	Observation	1,137

Note: This table reports the regression coefficients. The dependent variable is (log) markup. Each column is an OLS regression result of log markup on firm heterogeneity for observations for CPU and GPU with various levels. Column (1) shows the results for baseline transactions, and column (2) shows the results at the firm-destination market-yearly level. Finally, column (3) shows the results at the firm-yearly level. The standard errors are in parenthesis and are bootstrapped. Significance : * 10 percent, ** 5 percent, *** 1 percent.

C. Supplementary Documentation

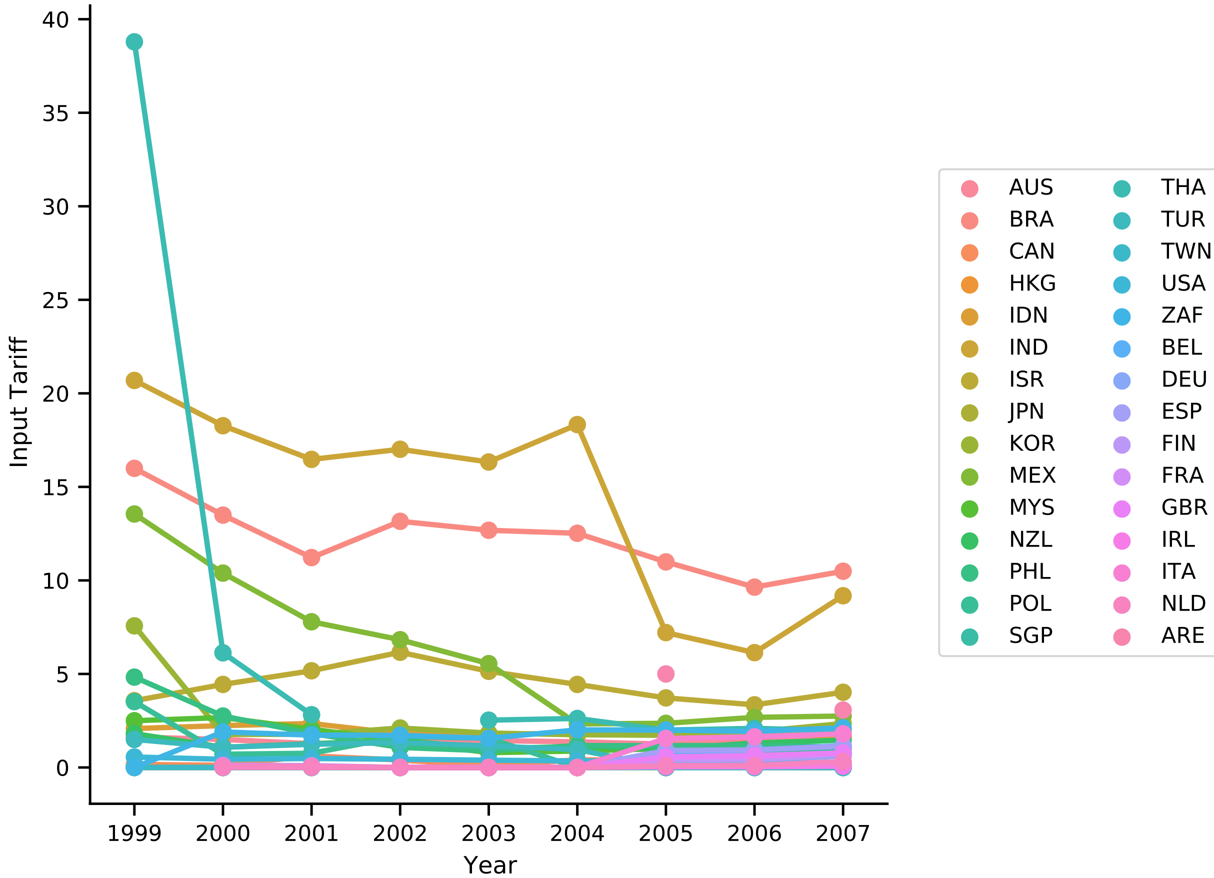
This section provides additional figures that supplement the main materials. Specifically, Figures 6 and 7 display input and output tariffs for each year at the destination market level. Figure 8 shows firm-year markups when the lognormal marginal distribution is used instead of the uniform marginal distribution.

Figure 5: Input Tariffs for CPU and GPU from 1999 to 2007



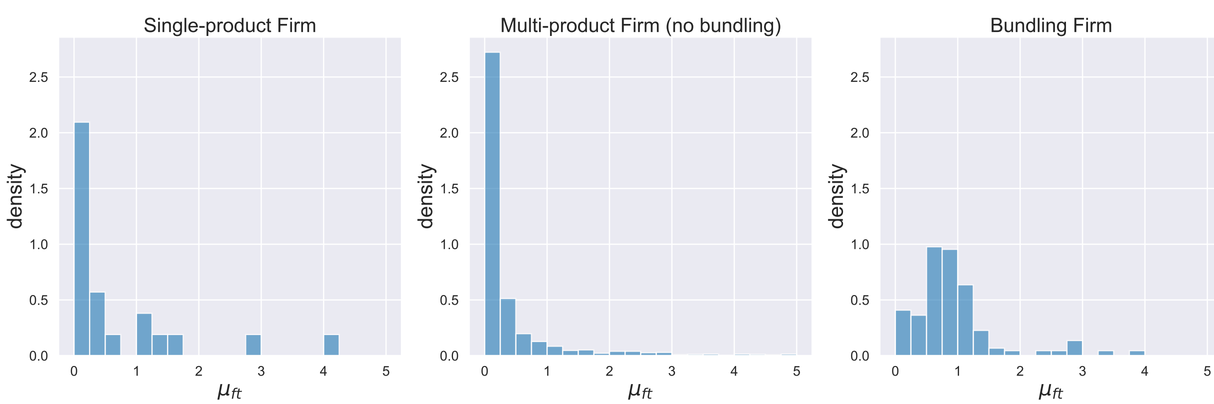
Note: This figure plots market-level input tariffs for China from 1998 to 2007 with respect to HS2 level 84 which contains CPU and GPU.

Figure 6: Output Tariffs for CPU and GPU from 1999 to 2007



Note: This figure plots market-level output tariffs for China from 1998 to 2007 with respect to HS2 level 84 which contains CPU and GPU.

Figure 7: Markups (μ_{ft}) of CPU and GPU by firm-types



Note: This figure is the same figure as figure 4, but it is recovered under the assumption that consumer valuation for each product follows log normal distribution rather than uniform distribution.