Product Bundling, Joint Markups and Trade Liberalization

Ji Hye Heo*

January 20, 2023

Abstract

Product bundling is a frequent practice that multi-product firms use 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 product bundling pose a challenge for previous methods in estimating markups. Utilizing the linkages across prices in the firm profit maximization problem, I propose a method to estimate transaction-level markups incorporating multi-product firms' decisions to bundle products. Focusing on Chinese exporters, multi-product firms that bundle products achieve markups that are approximately 30% higher than firms without product bundling. The markup premiums that bundling firms enjoy have been partially driven out since competition has increased due to China's WTO accession.
Keywords: Joint markups, multiproduct firms, product bundling, consumer valuation estimation, trade liberalization, pro-competitive effect
JEL Codes: D22, D43, F12, F13, F14, L11, L13,

*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: their ability to engage in product bundling, a practice whereby firms sell multiple goods in a single package.[1] In this paper, I study (i) how product bundling impacts a firm's markups and (ii) how the impacts of trade liberalization may differ for bundling and non-bundling firms.

Multi-product firms that engage in product bundling make pricing decisions jointly across the products they decide to bundle. Since previous methods of recovering markups rely on an implicit assumption that firms price goods independently, markups for bundling firms are not correctly captured by methods from the literature. This is especially true when the firm leverages its market power from one product market to another by bundling its goods. 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. The 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, the markups are recovered using the information from the first-order conditions.

[1] As a provider of multiple goods, 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*. If buyers can only purchase products as a bundle, it is called *pure bundling*. Depending on the market structure and the firm's market power, a firm engaging in bundling practices can price a bundle at a higher (*bundling premium*) or lower (*bundling discount*) price. This paper focuses on mixed bundling practices with price discounts, the most prominent case in product bundling.

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 approach from Berry et al. (1995). This significantly departs from the widely used method of recovering 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 methodology also departs from that of Berry et al. (1995), where the dimensionality problem was solved by switching to the product characteristic space, and thus, data on sales, product characteristics, and market share are required. Instead, the consumer tastes across products are estimated herein using transaction data, which eases the burden on the data compared Berry et al. (1995). By incorporating product bundling into the framework, this paper characterizes the difference in markup strategies 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) was proposed, the bundling literature focused on either theoretically extending the basic framework[3] or analyzing bundling practices in the 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

[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 do not contain price information but only revenue, only revenue elasticities can be obtained, not output elasticities. With 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 practices in a competitive setting where there are an arbitrary number of firms and Chen and Riordan (2013) for general conditions for the profitability of product bundling where a copula is used to model the stochastic dependence of consumer values.

[4]Regarding software, see the *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 PCs with Internet Explorer. Additionally, Crawford and Yurukoglu (2012) studied short-run welfare in the television channel market when á la carte policies that require distributors to offer individual channels for sale to consumers are introduced. The simulation results showed that increased input costs offset consumer benefits from purchasing individual channels.

the main analysis. Specifically, 84% firms in the data are multi-product firms that sell more than one product during the sample period. Out of firms that sell both ADPMs (*automatic data procession machines*) and ADPM accessories[5], 20% engage in product bundling. 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 among firms with different bundling decisions. The empirical analysis for Chinese exporters' electrical machines (ADPMs and ADPM accessories) from 2000 to 2006 shows that multi-product firms with product bundling enjoyed approximately 32.9% higher markups at the firm-market-year level than their counterparts without bundling practices. These differences in markups across bundling and non-bundling firms can 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.

Last, this paper adds to the literature studying the relationship between markups and competition in response to trade reform. Changes in market competitiveness force 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 find that (i) the incomplete pass-through of input costs declines to prices and (ii) there is a pro-competitive effect on markups. However, in their setting, each product's markup is assumed to be *independent* of the other products' markups even though most production occurs 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 the increased competition induced by trade liberalization results in a decrease in markup dispersion across firms for computer parts. This pro-competitive effect may partly come from increased competition driving out product bundling.

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

[5] ADPMs are machines that use logically interrelated operations performed in accordance with preestablished programs to furnish data. Computer parts such as CPUs (*central processing units*), GPUs (*graphics processing units*), and SSDs (solid state drives) fall into this category. Examples of ADPM accessories are coolers, server racks, and mounts.

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, the empirical results are presented, and section 5 concludes this paper.

2 Data and Trade Policy Background

I first describe the Chinese Customs data (CCD) in section 2.1 because these data determine 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 markup behavior across firms, time, and international markets. The CCD record Chinese firm-level exports and imports between 2000 and 2006 at the destination market–month level with corresponding HS6 codes, quantities, values, and firm characteristics such as names, ownership, addresses, and cities.

There are a few things to note about this data set. First, because these are customs data, the entire empirical analysis is focused on exporter firms and their export transactions.[6] Second, the framework for recovering markups requires transaction data such that, ideally, transactions are recorded between each seller and buyer firm in a short period of time. While the frequency of CCD is at the monthly level, which is a good measure for international trade, there is aggregation on the buyer side. This buyer-side aggregation may lead to misclassifying multiple single–good 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 the

[6] De Loecker and Warzynski (2012) show that exporter firms, on average, have higher markups than domestic firms. In contrast, Yang (2021) document that Chinese exporter firms have lower markups than nonexporters because China has a comparative advantage in low-markup products. Regardless, if there is not a systematic difference across firms with different pricing strategies, then focusing on exporters will not lead to significantly different results from those of domestic firms.

USA for the years 2016 and 2018.[7] Last, unlike production data where domestic and foreign quantities are aggregated, destination market-level transaction records allow me to incorporate demand-side characteristics into the framework and carry out the analysis by destination market. Hence, in the main empirical analysis, the markups are recovered at the firm-destination-product-month level and aggregated to various levels, such as the firm-year level.

2.1.1 Getting Comparison Prices for Multi-Product Firms

The framework for recovering 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 the products sold during a given transaction, the prices of unsold products (including the bundle) for multi-product firms need to be recovered. These unobserved prices are recovered based on the firm's actual behaviors using the monthly feature of the customs data. The key intuition is to get the unobserved prices from the most recent month.

Consider a benchmark case with two products, product 1 and product 2. For firm f in destination 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, dis) be the final *comparison* prices for product 1, product 2, both products combined, and the bundling discount for the transaction used to estimate consumer valuations and recover markups. The bundling discount is calculated as $dis = (x_1 + x_2) - x_b$. If a transaction is a multi-product transaction $(y_1, y_2) = (1, 1)$, with $dis > 0$, then it is classified as a transaction with product bundling. Transactions with either $(y_1, y_2) \neq (1, 1)$ or $dis = 0$ are not classified as bundling transactions.[8]

First, for a given transaction, if the price is observed, the comparison price is simply the 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

[7] Firm-to-firm-level transaction data show that most transactions (88%) remain the same when seller firm-to-buyer firm transactions are aggregated to the seller firm-to-buyer market, indicating that the chances of misclassifying non-bundling transactions as bundling transactions are slim.

[8] For example, a transaction where $(y_1, y_2) = (1, 0)$ with $dis > 0$ is a transaction where the buyer buys only 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 $dis = 0$ is simply a transaction with multiple products and is not classified as a bundled transaction.

for $(y_1, y_2) = (1, 1)$, $x_b = p_1 + p_2$. If a price is not observed for product j, then $y_j = 0$ for $j = 1, 2$. Then, for a given firm-destination-year, I find the most recent 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 most recent transaction with $(y_1, y_2) = (1, 1)$. I use the price from the most recent transaction as the comparison price, i.e., $x_j = p_j^c$, where p_j^c denotes the price from the most recent transaction. For $j = b$, I find the most recent transaction where $(y_1, y_2) = (1, 1)$ and apply it as $x_b = p_1^b + p_2^b$, where p_j^b are prices from the most recent multi-product transaction. If there are no transactions with $(y_1, y_2) = (1, 1)$, I simply set this as $x_b = x_1 + x_2$.

Table 1 presents a basic example of the price recovery procedure for a given firm-destination pair. For the January transaction, $(y_1, y_2) = (1, 0)$ means only product 1 is sold. The observed price is simply transferred to the comparison price depicted with superscripts *. Since January is the most recent transaction for the February and June transaction, the January price is also used as a comparison price depicted with superscripts **. Similarly, the February transaction reveals a comparison price for product 2, and the June transaction shows the price for multi-product transactions. Then discounts can be calculated using prices of single goods case – (x_1, x_2) and both goods case – (x_b). Since the June transaction has both items with $(y_1, y_2) = (1, 1)$ and positive discounts, $dis > 0$, this transaction is classified as a bundled transaction.

Table 1: Example of Recovering Comparison Prices

	Observed in the Data				*Final Comparison Prices*				
transaction	y_1	y_2	p_1	p_2	x_1	x_2	x_b	dis	Bundle?
2000, Jan	1	0	80	-	80*	120	170	30	N
2000, Feb	0	1	-	120	80**	120	170	30	N
2000, June	1	1	70	100	80**	120	170	30	Y

Note: This table shows how prices are recovered for unobserved prices.

2.1.2 Firm Type Definition and Data Description

After unobserved prices are recovered, firms that sold both products can be classified into bundling and non-bundling firms. First, firms are categorized into single- or multi-product firms depending on whether they sold both product 1 and product 2 to each destination in a given year. For example, if firm f produced multiple products but sold only ADPMs to destination d in year t, then the firm is classified as a single-product firm in destination d in year t. Once unobserved prices are recovered for multi-product firms that have sold products of interest, multi-product firms are further divided into bundling firms and non-bundling firms based on whether there was a bundling transaction in market d in year t.

In this paper, ADPMs and ADPM accessories are selected for analysis.[9] The products are selected based on the following criteria. First, there must be enough observations. Electrical machines were one of the most exported goods from China during the sample period. Additionally, the relationship between goods must be considered. Goods are chosen based on whether firms are likely to produce both goods and sell them as a bundle to importing firms in the destination. ADPMs and ADPM accessories are both parts of electrical machines that are frequently produced by the same manufacturers. A basic description of the data for ADPMs and ADPM accessories is summarized in Table 2.

Table 2: Summary Statistics

	Observation	(%)
Firm-Destination-Year Pair Characteristics(fdt)	35,800	100.0%
Sold either ADPMs or ADPM accessories	28,010	78.24%
Sold both ADPMs and ADPM accessories	7,790	21.76%
without product bundling	6,379	81.89%
with product bundling	1,411	18.11%
Transaction Characteristics(fdm)	158,368	100.0%
Have sold either ADPMs or ADPM accessories	131,503	83.03%
Have sold both ADPMs and ADPM accessories	26,865	16.96%
Transaction has one item	22,843	85.02%
Transaction has both items	4,022	14.97%
Bundling transaction	1,647	40.95%

Note: The subscripts indicate the following: f is for firms, d is for destination, j is for products(ADPMs or ADPM accessories), and m and t are time subscripts that stand for month and year, respectively.

In the upper panel, Table 2 shows that based on the transaction-side classification, the majority of firms sold either ADPM or ADPM accessories given the destination-year pair. Specifically, out of 35,800 firm-destination-year pairs, firms that sold only one product out

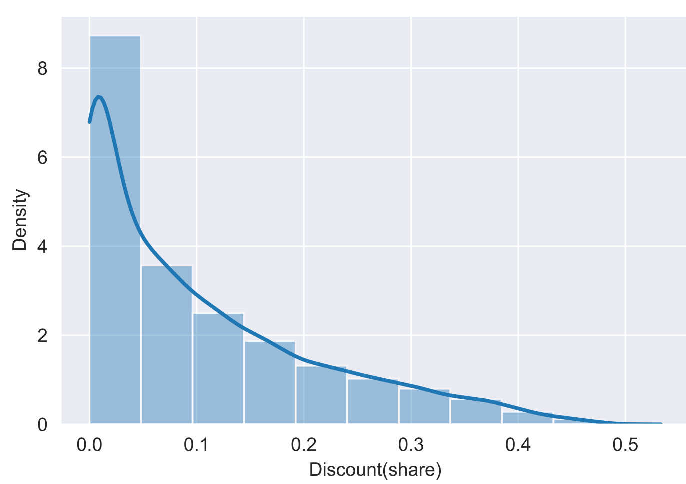
[9]The products are classified at the HS6 code level. Specifically, the ADPMs are {847130, 847141, 847149, 847150, 847160, 847170, 847180}, and the ADPM accessories are {847330}.

of two products account for approximately 78% of firm-destination-year pairs. This is due to transaction-side classification and focusing on only two products. Note that most of these firms are multi-product firms that only happen to sell only one of the products of interest.

In the main empirical analysis, 7,790 firms that sold both ADPMs and ADPM accessories to market d in year t are used. This is to compare the effect of bundling on markups between firms that can offer the bundling option with ADPMs and ADPM accessories. Out of firms that sold both ADPMs and ADPM accessories to destination d in year t, approximately 35% of firms engaged in product bundling with ADPMs or ADPM accessories. If the size of a bundled product is increased or multiple bundles are considered, the number of firm-destination-year pairs that sell all bundled products will increase.

The bottom panel describes the baseline transactions, defined at the firm-destination-month level. There are a total of 158,368 transactions, and of those, approximately 17% sold ADPMs and ADPM accessories. Among the transactions involving both ADPMs and ADPM accessories, 40.95% were bundled transactions. Figure 1 plots the share of bundling discounts for those bundling transactions. A power–law feature is shown for the bundling discount, where most of the discounts are 10% or less of the original price.

Figure 1: Bundling discount share for bundled transactions



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

2.2 *WTO Accession and Tariff Reductions*

China's WTO accession, which took place in 2001, has induced substantial tariff reductions (see Lu et al. (2015)). In this section, I document the impact of China's trade liberalization on electric machines using tariff data from the 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 destinations where China had the most transactions for ADPMs and ADPM accessories. They account for 94.77% of the quantity exported and 97.13% of trade value.[10]

Figure 2 displays the evolution of China's aggregated market access, output and input tariffs.[11] Market access tariffs are tariffs that partner country firms face when exporting to China, whereas output tariffs are those that Chinese exporters face. Output tariffs are aggregated using each destination's trade value as weights. For the input tariffs, I follow De Loecker et al. (2016) and construct them for each destination by passing the tariff data at the ISIC Rev3 level to China's input–output matrix table for 1995–2010 and then using the values as weights to create the aggregated input tariffs.[12] Figure 2 shows that trade liberalization brought a sharp decline in both output and input tariffs for ADPMs and ADPM accessory products and a modest decline in market access tariffs. Specifically, the output tariff declined from approximately 12% to 4%, and the input tariff declined significantly from approximately 19% to 8%.

3 A Framework for Estimating Markups

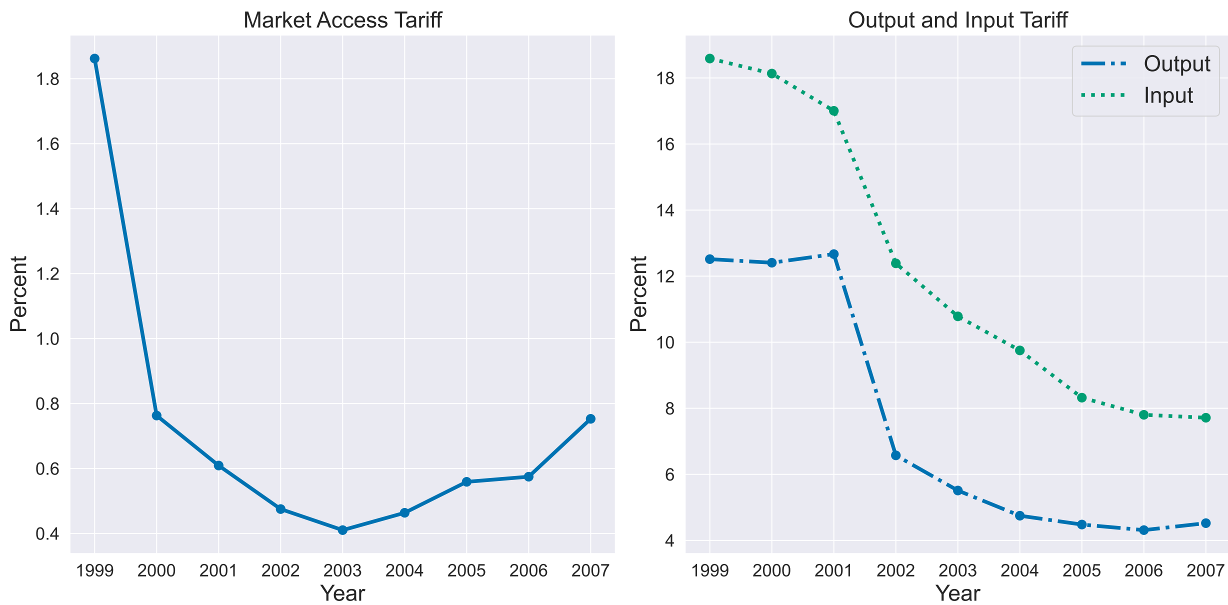
To incorporate joint markups and product bundling, I introduce an empirical model from the bundling literature. While non-bundling firms price goods independently, a bundling

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

[11]Market access, output and input tariffs for each market are displayed in the appendix.

[12]The formal definition of the input tariff is $\tau_{idt}^{input} = \sum_k a_{ki}\tau_{kdt}^{output}$, where τ_{kdt}^{output} is the export tariff for destination d to China in 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 ADPMs and ADPM accessories from 1998 to 2007



Note: This figure plots aggregated market access and, output and input tariffs for China from 1999 to 2007 for HS2 level 84, which contains both ADPMs and ADPM accessories.

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 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 chooses.

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, the model recovers marginal costs and consumer valuations across goods from the information revealed 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, the set of individual goods (bundles) for firm f in year t are denoted as $\mathcal{G}_{ft}(\mathcal{B}_{ft})$, and the number of components in the set are $G_{ft}(B_{ft})$. Let J_{ft} be the total number of products that the multi-product firm sells, either as individual products or as a product bundle. For example, if firm f produces two discrete products and sells three products—each individual product 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 firm f's constant marginal cost for a single product $j \in \mathcal{G}_{ft}$ in destination d and year t.[13] 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.[14] For example, in the case of $G_{ft} = 2$, $c_{fdbt} = c_{fd1t} + c_{fd2t}$ and $P_{fdbt} = P_{fd1t} + P_{fd2t} - dis_{fdbt}$ with $dis_{fdbt} > 0$, where the subscript b refers to a bundled product comprising both product 1 and product 2. Multi-product firms that do not engage in bundling practices could be interpreted as having $dis_{fdbt} = 0$, that is, as effectively selling both goods simultaneously. Thus, while the subsequent discussion assumes that multi-product firms bundle individual products, it could easily be applied to multi-product firms without bundling by setting $dis_{fdbt} = 0$.

The model assumes that consumers for each firm desire at most one unit of each good and demand each good independently of their consumption of the other goods.[15] For these consumers, consider the consumer valuations for G_{ft} goods at transaction month m, $\mathbf{v}_{fdm} = (v_{fd1m}, ..., v_{fdG_{ft}m})$, which are distributed according to the unknown distribution function $\Psi_{fd}(\mathbf{v}_{fdm})$.[16] Let $\psi_{fd}(\mathbf{v}_{fdm})$ and $\psi_{fdk}(v_{fdkm})$ be the probability density function and marginal density function, respectively, for product k for $\Psi_{fd}(\mathbf{v}_{fdm})$. To avoid trivial cases, a positive measure of consumers exists such that $v_{fdjm} \geq c_{fdjt}$ for all j, and resale by consumers is not

[13]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. However, as long as marginal costs are locally constant at the optimal price, additional order or sale of the products will not affect the scale of the marginal costs. Figure 11 in the appendix reveals that there is no clear sign of either increasing or decreasing returns to scale for the main analysis.

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

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

[16]The 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, and market to capture demand characteristics at the firm, and destination levels. This is because there are significant price variations across destinations for the same product by the same seller.

possible.

This paper will focus on a benchmark case where $G_{ft} = 2$ to build on the 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 substantially more derivations. In practice, bundled products do not typically contain many individual products, which eases the burden of derivation and any data restrictions.[17].

The next section describes the framework for recovering markups with the transaction unit based on Chinese customs data; thus, the transactions are at the firm-destination-month level. However, the transactions can be defined based on the available data.[18]

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.[19] Thus, the first-order conditions for each product-level profit entail 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 in month m. Given the consumer valuations for product j in destination d and year t, consumers whose valuations are higher than the price will purchase the good. Thus, $Q^D_{fdjm}(P_{fdjm}) = \int_{P_{fdjt}}^{\infty} \psi_{fdj}(x)dx$.

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

[18]The assumptions on the unit of the marginal costs, consumer valuations, and prices, and hence the markups, can be chosen appropriately based on how detailed the "transaction" is. For example, this paper defines a transaction as between an exporter firm and 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 destination-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 differ by destination market, i.e., Ψ_{fd}, to reflect destination-specific demand characteristics. The unit price for each product follows the unit of transactions, i.e., P_{fdjm}, where m is the month.

[19]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 $dis_{fbt} = 0$, setting the discount equal to zero.

Note that the quantity demanded for good j is only a function of the good j characteristics such as price P_{fdjm} and its marginal distribution ψ_{fdj} and does not depend on other products' characteristics. Then, the first-order condition (2) yields the following equation in terms of the marginal density of valuations for product j, ψ_{fdj}, monthly prices in year t, $P_{fjdm\in t}$, and the marginal cost c_{fdjt}.[20] After the distribution of 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.

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

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) remains the same.

3.2 *Recovering Joint Markups with Product Bundling*

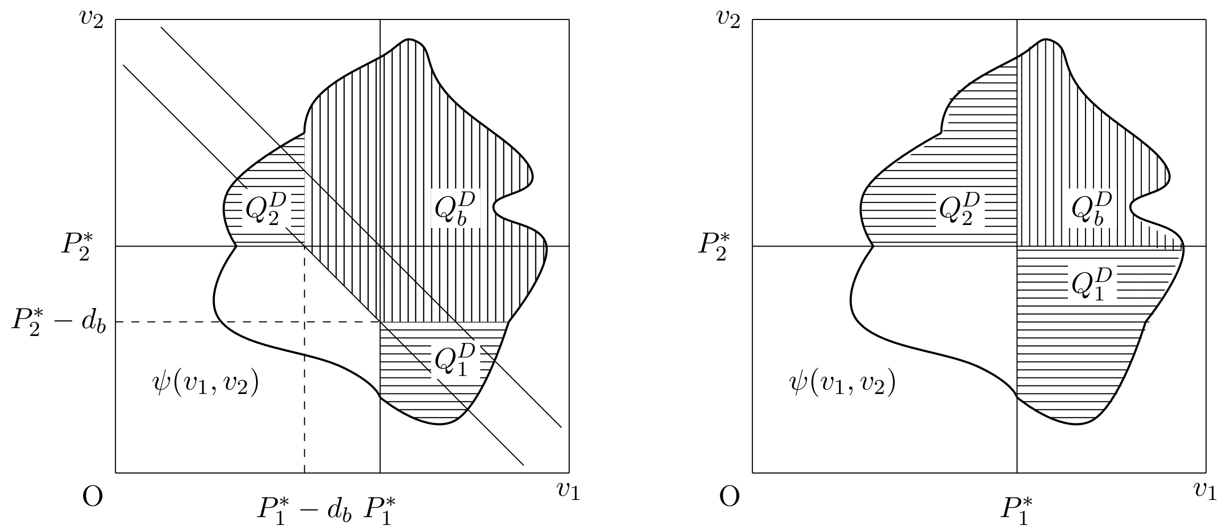
To identify joint markups, I introduce a framework from the bundling literature. The approach employs a model setting similar to those of McAfee et al. (1989) and Chen and Riordan (2013) in that consumer valuations are introduced to capture demand-side information. While their work focuses on finding 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 for estimating 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 firm f only if it gives her the highest utility among all other options. This enables me to write

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

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 a discount. For both figures, subscripts f, d, and m or t are dropped for parsimony.

the quantity demanded for each good j (Q^D_{fdjm}) in terms of prices and the distribution of consumer valuations. For example, when $G_{ft} = 2$,

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

Denoting the vector of prices as $\mathbf{P}_{fdm} = (P_{fd1m}, P_{fd2m}, P_{fdbm})$, 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_1^D, Q_2^D, or Q_b^D will buy good 1, good 2, or the bundled goods, respectively. Note that the quantity demanded for a single good $j = 1, 2$ is a function of not only its price but also the price of the bundled good b and hence the price of the other good, explicitly showing the linkage across goods for firms with joint pricing.

$$
\begin{aligned}
Q^D_{fd1m}(\mathbf{P}_{fdm}) &= \int_{P_{fd1m}}^{\infty}\int_0^{P_{fdbm}-P_{fd1m}} \psi_{fd}(x,y)dydx,\\
Q^D_{fd2m}(\mathbf{P}_{fdm}) &= \int_0^{P_{fdbm}-P_{fd2m}}\int_{P_{fd2m}}^{\infty} \psi_{fd}(x,y)dydx,\\
Q^D_{fdbm}(\mathbf{P}_{fdm}) &= \int_{P_{fd1m}}^{\infty}\int_{P_{fdbm}-P_{fd1m}}^{\infty} \psi_{fd}(x,y)dydx + \int_{P_{fdbm}-P_{fd2m}}^{P_{fd1m}}\int_{P_{fdbm}-x}^{\infty} \psi_{fd}(x,y)dydx,
\end{aligned} \tag{4}
$$

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:

$$
\begin{aligned}
\underset{\mathbf{P}_{fdm}}{\operatorname{argmax}}\, \Pi_{fdt} = \underset{\mathbf{P}_{fdm}}{\operatorname{argmax}} \sum_{m\in t} \Pi_{fdm} &= \underset{\mathbf{P}_{fdm}}{\operatorname{argmax}} \sum_{m\in t}\Big(\Pi_{fd1m} + \Pi_{fd2m} + \Pi_{fdbm}\Big),\\
\text{where} \quad \Pi_{fdkm} &= (P_{fdkm} - c_{fdkt})Q^D_{fdkm}, \quad \text{for all } k \in \{1,2,b\},
\end{aligned}
$$

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

$$
\begin{aligned}
\Pi_{fdt} = &\sum_{m\in t}\Big(\Pi_{fd1m} + \Pi_{fd2m} + \Pi_{fdbm}\Big),\\
=&(P_{fd1t} - c_{fd1t})\int_{P_{fd1t}}^{\infty}\int_0^{P_{fdbt}-P_{fd1t}} \psi_{fd}(x,y)dydx, + (P_{fd2t} - c_{fd2t})\int_0^{P_{fdbt}-P_{fd2t}}\int_{P_{fd2t}}^{\infty} \psi_{fd}(x,y)dydx\\
&+ (P_{fdbt} - c_{fd1t} - c_{fd2t})\Bigg[\int_{P_{fd1t}}^{\infty}\int_{P_{fdbt}-P_{fd1t}}^{\infty} \psi_{fd}(x,y)dydx + \int_{P_{fdbt}-P_{fd2t}}^{P_{fd1t}}\int_{P_{fdbt}-s}^{\infty} \psi_{fd}(x,y)dydx\Bigg].
\end{aligned}
$$

The first-order conditions for the price variables give the following three equations that express the relationship among marginal costs (hence markups) across products in terms of consumer valuation $\psi_{fd}(v_{fd1m}, v_{fd2m})$ and price variables.[21]

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

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

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

where $\mathcal{A}_{fdm} = \int_0^{P_{fdbm}-P_{fd1m}} \psi_{fd}(P_{fd1m}, y)dy$, $\mathcal{B}_{fdm} = \int_{P_{fd1m}}^{\infty} \psi_{fd}(x, P_{fdbm} - P_{fd1m})dx$, $\mathcal{C}_{fdm} = \int_0^{P_{fdbm}-P_{fd2m}} \psi_{fd}(x, P_{fd2m})dx$, $\mathcal{D}_{fdm} = \int_{P_{fd2m}}^{\infty} \psi_{fd}(P_{fdbm} - P_{fd2m}, y)dy$, and $\mathcal{E}_{fdm} = \int_{P_{fdbm}-P_{fd2m}}^{P_{fd1m}} \psi_{fd}(x, P_{fdbm} - x)dx$.[22] Note that after the consumer's valuation distribution $\psi_{fd}(x, y)$ is estimated, the first-order conditions provide the expression needed to identify joint markups.[23]

The equations from system (5) are denoted as $\mathbf{\Gamma}(\mathbf{P}_{fdm}, \mathbf{Q}^D_{fdm}, \psi_{fd}(\mathbf{v}_{fdm}); \mu_{fdj\in\{1,2\}m}) = 0$. Note that $\mathbf{\Gamma}(\mu_{fdj\in\mathcal{G}_{ft}m}) = 0$, 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 overdetermined. I propose recovering the 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 that 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 the case as if all firms offer all possible combinations of a bundle. That is,

[22]For multi-product firms without bundling, taking the first-order conditions with respect to only (P_{fd1t}, P_{fd2t}) or plugging $d_{fdbt} = 0$ into equation (5) yields 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 are robust to numerically estimating the markups of multi-product firms without bundling by plugging in $d_{fdbt} = 0$.

[23]The intuition for the first-order conditions still holds, just as 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, the firm will lose profit from consumers who were on the margin between Q_{fd1m} and not buying.

$B_{ft} = \sum_{b=2}^{G_{ft}-1} \binom{G_{ft}}{b}$ 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, I 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, I 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 prices of bundled goods of which k is a component. For $k \in \mathcal{B}_{ft}$, $Q^D_{fdkm}(\mathbf{P}_{fdm})$ should be a function of its price and the prices of all of the individual products that compose bundle k. After deriving expressions for the quantity demanded, I plug them into the profit function to derive J_{ft} first-order conditions with joint markups. This process is denoted as $\mathbf{\Gamma}(\mathbf{P}_{fdm}, \mathbf{Q}_{fdm}, \psi_{fd}(\mathbf{v}_{fdt}); \mu_{fdj \in \mathcal{G}_{ft} m}) = 0$, and the joint markups are recovered numerically as in the $G_{ft} = 2$ case.

3.3 *Consumer Valuation Estimation*

This section describes how the consumer valuation distribution ψ_{fd} is estimated using the approach proposed by Letham et al. (2014). The joint probability density function, $\psi_{fd}(\mathbf{v}_{fdm})$, 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-destination 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 regarding the correlation structure across goods for $\psi_{fd}(\mathbf{v}_{fdm})$ is possible,[24] if the relationship between goods affects markups in a meaningful way, specific assumptions will likely distort the estimation of markups.[25] Using transaction data, Letham et al. (2014) propose a

[24] See Letham et al. (2014) for a survey of studies that made assumptions of either independence or perfect correlation regarding the correlation structure across goods.

[25] 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 significantly different predictions regarding possible

statistically consistent 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 the 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.

A transaction is defined as a deal between a seller and buyer during a certain period of time. In a retail setting where consumers often buy goods in small amounts, days are a good choice for the time unit. In trade, where buyer firms purchase goods in large amounts from specific sellers, months or years may be an appropriate time unit 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.[26] The other component is the price data for individual products in 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 $\mathbb{C}_{fd}(\cdot)$ for $\Psi_{fd}(\cdot)$ is a distribution function over $[0,1]^{G_{ft}}$ with uniform margins such that $\Psi_{fd}(v_{fd1m}, ..., v_{fdG_{ft}m}) = \mathbb{C}_{fd}(\Psi_{fd1}(v_{fd1m}), ..., \Psi_{fdG_{ft}}(v_{fdG_{ft}m}))$. The copula $\mathbb{C}_{fd}$ contains all information on the dependence structure between the components of $(v_{fd1m}, ..., v_{fdG_{ft}m})$ and combines each marginal distribution Ψ_{fdk} to return the joint distribution Ψ_{fd}. Suppose each marginal distribution is a function of parameters $\boldsymbol{\theta}_{fdj}$, i.e., $\Psi_{fdj}(v_{fdjm}; \boldsymbol{\theta}_{fdj})$, and the copula distribution belongs to a family with parameters $\boldsymbol{\phi}_{fd}$, i.e., $\Psi_{fd}(\boldsymbol{v}_{fdm}; \boldsymbol{\theta}_{fd}, \boldsymbol{\phi}_{fd}) = \mathbb{C}_{fd}(\Psi_{fd1}(v_{fd1m}; \boldsymbol{\theta}_{fd1}), ...\Psi_{fdG_{ft}}(v_{fdG_{ft}}; \boldsymbol{\theta}_{fdG_{ft}}); \boldsymbol{\phi}_{fd})$. 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 marginal

profit when a bundled product is introduced.

[26] 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.

distributions $\Psi_{fdj}(\cdot)$ and copula function $\mathbb{C}_{fd}$, then find the parameters for which $\mathbb{C}_{fd}(\Psi_{fd1}(v_{fd1m}),...,\Psi_{fdG_{ft}}(v_{fdG_{ft}m}))$ is the closest to $\Psi_{fd}(v_{fd1m},...v_{fdG_{ft}m})$ in terms of likelihood.

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

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

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

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.

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

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

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

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) stated, if the demand model is linear, the corresponding valuation distribution is a uniform distribution. If the demand model follows the normal distribution function, the corresponding marginal valuation distribution also follows a normal distribution. For empirical analysis, I follow Letham et al. (2014) in using uniform distributions for the marginal distributions and a Gaussian function for the copula function.

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

$$
\begin{aligned}
l_{ft}(\hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd}) &= \sum_{s=1}^{S} \log \mathfrak{p}_f(\mathbf{y}^s | \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd}) && (10)\\
&= \sum_{s=1}^{S} \log \int \mathfrak{p}_f(\mathbf{y}^s | \mathbf{v}^s, \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd}) \mathfrak{p}_f(\mathbf{v}^s | \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd}) d\mathbf{v}^s && (11)\\
&= \sum_{s=1}^{S} \log \int_{v^{s,l}_{G_{ft}}}^{v^{s,u}_{G_{ft}}} \cdots \int_{v^{s,l}_{1}}^{v^{s,u}_{1}} \psi_{fd}(v^s_1, ..., v^s_{G_{ft}}; \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd}) dv^s_1 ... dv^s_{G_{ft}} && (12)\\
&= \sum_{s=1}^{S} \log \sum_{k=0}^{G_{ft}} (-1)^k \sum_{I \subseteq \{1,...,G_{ft}\}, |I|=k} \Psi_{fd}(\mathbf{v}^s; \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) && (13)
\end{aligned}
$$

where the equality in equations (11) to (12) uses $\mathfrak{p}_f(\mathbf{v}^s | \mathbf{P}^s_{G_{ft}}, \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd}) = \mathfrak{p}_f(\mathbf{v}^s | \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd}) = \psi_{fd}(\cdot; \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd})$ and makes 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) employed 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 for estimating the consumer valuation distribution $\psi_{fd}(\mathbf{v})$ is

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_{fdj} &\in \underset{\boldsymbol{\theta}_{fdj}}{\operatorname{argmax}} \sum_{s=1}^{S} (y^s_j \log(1 - \Psi_{fdj}(P^s_j; \boldsymbol{\theta}_{fd}))) + (1 - y^s_j) \log(\Psi_{fdj}(P^s_j; \boldsymbol{\theta}_{fd}))\\
\hat{\boldsymbol{\phi}}_{fd} &\in \underset{\boldsymbol{\phi}_{fd}}{\operatorname{argmax}} \sum_{s=1}^{S} \log \sum_{k=0}^{G_{ft}} (-1)^k \sum_{I \subseteq \{1,...,G_{ft}\}, |I|=k} \Psi_{fd}(\tilde{\mathbf{v}}^s; \hat{\boldsymbol{\theta}}_{fd}, \boldsymbol{\phi}_{fd})
\end{aligned}
$$

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 compared to non-bundling firms. Additionally, I use tariff reductions from China's WTO accession to see whether trade liberalization differentially affects markups between bundling and non-bundling firms. In studying 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 rests on the dependence of consumer values summarized by the copula and not on marginal costs.[27]

After the consumer valuations that capture demand-side characteristics are estimated, markups for firms with and without bundling can be computed from the first-order conditions as described in the previous section. The recovered markup estimates and empirical analysis reveal several major findings. First, by comparing markups to those recovered using the De Loecker et al. (2016) method, I find that while markups for non-bundling firms are similar across methods, the traditional method cannot fully capture markups for bundling firms.[28] Second, I investigate the relationship between markups and firm types across destinations and time. These analyses cannot be done using previous methods, where product bundling and joint pricing decisions were not incorporated into the estimation process. I find that bundling firms have higher markups and that these markup differences reduce with a decrease in input tariffs since it brings new entrants into the market.

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

[27] 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 consider the marginal cost, the choice to bundle or not purely depends on whether the profit from additional purchases by offering a discount, dis_{fdbm}, overrides profit loss by offering a discount. 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 the proof.

[28] There is a discussion of how markups from the production side are estimated using the De Loecker et al. (2016) method, and empirical results are given in the Appendix.

conditions.

Table 3: Markup (μ_{fdjt}) Results

	Total			Non-Bundling			Bundling		
	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std
ADPMs	2.22	1.40	1.90	2.15	1.26	2.01	2.46	2.27	1.41
Accessories	2.31	1.49	1.94	2.28	1.33	2.07	2.40	2.26	1.31

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

Table 3 presents recovered product markups for bundling and non-bundling firms at the firm-destination-year level. On average, ADPMs and ADPM accessory products are priced approximately 120% higher than their original costs.[29] Within the sample, the mean and the median values of the recovered markups are higher for bundling firms than non-bundling firms. Standard deviations are relatively high across firm types due to markups in the upper tail. For both bundling and non-bundling firms, the maximum value of markups is around 8 to 10.

Figure 4 shows the histogram of markups recovered from the transaction–side approach suggested by this paper and the production–side approach from the previous literature. For this figure, single-product firms export only one good, either ADPMs or ADPM accessories, including outside options. Multi-product firms export more than one good, with one or both goods being ADPMs or ADPM accessories. Bundling firms export both ADPMs or ADPM accessories with product bundling.[30] This definition is to compare markup distribution from that of the production side.

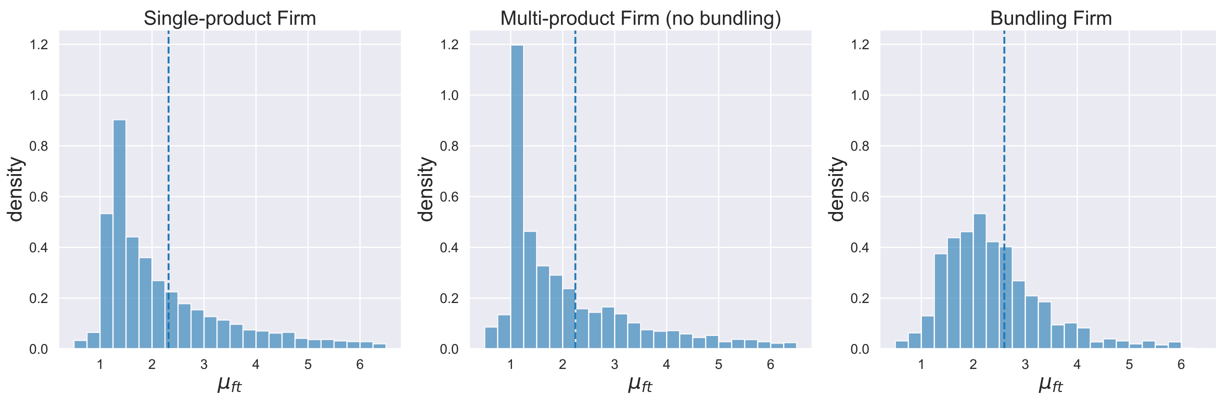
Firstly, Figure 4 (a) shows that differences in the markups across firm types are not fully captured on the production side, where bundling firms are absent. While single- and multi-product firms without bundling practice have peaked markup distribution with strong power law features, the distribution of bundling firms is relatively more even compared to non-bundling firms and is towards higher values. This indicates that bundling firms may have different markup structures compared to non-bundling firms. These stark differences

[29]Note that a markup value of 1.5 means that the firms obtain 50% of the marginal cost as profit for each unit. For example, with an ADPM with a marginal cost of $100, the price is set at $150.

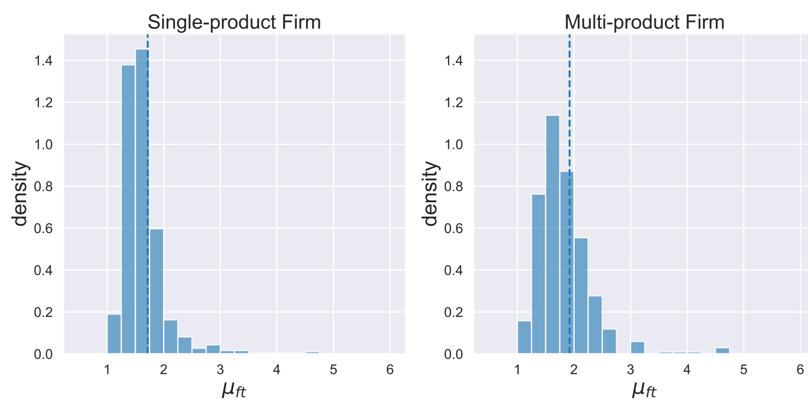
[30]Markups recovered using the lognormal distribution with the Gaussian copula instead of the uniform distribution with the Gaussian copula show similar results and are presented in the Appendix.

Figure 4: Markups (μ_{ft}) of ADPMs or ADPM accessories by firm type

(a) Transaction Side



(b) Production Side



Note: Figure (a) plots the histogram of firm-year level markups for ADPMs or ADPM accessories that is recovered using the transaction–side approach suggested by this paper. Figure (b) plots the histogram of firm-year level markups from the production side approach using the Chinese manufacturing data. Since the manufacturing data follows different product codes, this histogram plots markups for electrical machines. Note that the y-axis is the normalized density and the x-axis is μ_{ft}. Here, the top and bottom 3% of values are trimmed for both cases.

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.

4.2 Markups, Firm Heterogeneity and Trade Liberalization

The relationship between product bundling and markups depicted in Figure 4 may explain one additional channel relevant to why multi-product firms dominate international trade, in addition to 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 and then across time, using China's WTO accession as a trade liberalization event. For the empirical 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.[31]

4.2.1 Markups and Firm Heterogeneity

I first study how markups differ between bundling and non-bundling firms using the following regression equation:

$$\log \mu_{fdjm} = \delta_{FE} + \delta_B D_{BF} + \mathbf{x}'\beta + \varepsilon_{fdjm} \tag{14}$$

where D_{BF} is a dummy variable for multi-product firms that engage in bundling practices. To capture any destination, time, or ownership[32] trends, the destination, year, and firm ownership fixed effects are included in δ_{FE}. To capture firm size, the quantities are included in the covariate X as well as input, output, and market access tariffs. In this regression, δ_{BF} measures the percentage premium that the multi-product firm with bundling has over multi-product firms that do not engage in bundling (i.e., the "bundling premium").

Table 4 shows the results of equation (14) at various levels of markups, and the results align with economic intuition. The bundling dummy coefficients are all significant and positive across all specifications. For the baseline transactions in column (1), bundling firms have, on average, 19.1% higher markups than non-bundling that sold both products. When

[31] The baseline markups are recovered at the firm-destination-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.

[32] 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_{BF}	0.1746***	0.1908***	0.2845***
(*Product Bundling Premium*)	(0.0316)	(0.0351)	(0.0229)
Destination FE	Yes	Yes	Yes
Product FE	Yes	Yes	No
Year FE	Yes	Yes	Yes
Ownership FE	Yes	Yes	Yes
F-statistic	73.35	31.99	41.48
Observation	30,887	16,040	7,362

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 bundling firm dummy for observations for ADPMs or ADPM accessories with various levels. Column (1) shows the results for baseline transactions, which is the firm-destination-product-monthly level, and columns (2) and (3) show the results at the firm-destination-product-yearly level and firm-destination-yearly level. The standard errors are in parentheses and are clustered at the firm and yearly level. Significance : * 10 percent, ** 5 percent, *** 1 percent.

monthly markups are aggregated to the yearly level, the bundling dummy coefficient increases slightly and shows that bundling firms, on average, have 21.0% higher markups than non-bundling multi-product firms. Similarly, multi-product firms with bundling enjoyed 32.9% higher markups compared to non-bundling firms at the firm-destination-yearly level.

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 the true market power. In columns (1) and (2), the *product bundling premium* associates product bundling with firms with market power – higher markups. However, once we move from column (2) to (3), where markups are aggregated at the firm-level and capture firm-level decisions, the large difference in the *product bundling premium* from column (2) to (3) shows that for multi-product firms, firm-level joint decisions such as product bundling may require analysis at both the product and firm levels to fully characterize their market power. In short, these results show that firms could potentially utilize product bundling to exercise market power and retain higher markups compared to other firms.

4.2.2 Bundling and Additional Sales

Findings from the bundling literature indicate that firms engage in bundling practices to increase their profits by increasing the probability of selling additional products via a small discount on bundled products. To investigate whether this is how bundling firms obtain higher firm level markups than non-bundling, I follow the literature and run the following probit regression for only firms that sell both ADPMs and ADPM accessories.

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

where MPT_{fdm} is a dummy that is equal to one when a transaction between firm f and destination d in month m is a multi-product transaction with ADPMs or ADPM accessories. A positive value for the coefficient δ_{BF} indicates that bundling firms have a higher probability of selling goods together than other firms. I expect a positive value for $\delta_{\phi_{fd}}$ since customers with high $\delta_{\phi_{fd}}$ value ADPMs are more likely to value ADPM accessories more and thus have a higher chance of buying both goods. Vector X includes parameters related to consumer valuations such as maximum and minimum ($\theta_{\mathbf{fd}}$), and prices ($\mathbf{p_{fdm}}$).

Table 5: Probit Regression for Multi-Product Purchases

	MPT_{fdm}		
δ_{BF}	0.3601***	0.3947***	0.3814***
	(0.008)	(0.008)	(0.009)
$\delta_{\phi_{fd}}$		0.2712***	0.2510***
		(0.007)	(0.007)
c_{fd1t}			−0.0005***
			(0.0001)
c_{fd2t}			−0.003***
			(0.001)
Other Covariates	Yes	Yes	Yes
LLR-pvalue	***	***	***
Observation	29,148	29,148	29,148

Note: The dependent variable is a dummy MPT_{fdm} that equals one when a given transaction involves both ADPMs and ADPM accessories. Other covariates such as marginal parameters ($\theta_{\mathbf{fd}}$) and prices ($\mathbf{p_{fdm}}$) are included to control for demand side information. The coefficients show marginal effect of a unit increase from zero for each regressor. The standard errors are in parentheses. Significance : * 10 percent, ** 5 percent, *** 1 percent.

Table 5 displays the marginal effect of a unit increase from zero for each regressor. First, the correlation between the consumer's values for ADPMs and APDM accessories captured by ϕ_{fd} is positive and significant across specifications at the 1% level, as expected. If the consumers change their perception of ADPMs and their accessories from independent goods to perfect complements, multi-product sales increase by 25%. Additionally, the bundling firm coefficient is positive and significant at the 1% level across regressions. As indicated by the literature, bundling firms have a higher probability of selling more goods (both ADPMs and ADPM accessories) than others. Specifically, controlling for both the demand- (consumer taste) and supply- (marginal cost) side characteristics, being a bundling firm increases the probability of multi-product sales by 38%. This result shows that bundling firms are more likely to sell multiple goods as a bundle than other multi-product firms, increasing profit. In this case, if multi-product firms bundle products with low markups with products with high markups, then bundling firms can increase overall firm-level markups. This can be seen in Figure 5 where markups for both ADPMs and accessories are plotted. Figure 5 shows negative correlations (-0.15) between markups for ADPMs and accessories, indicating that bundling firms bundle products with low and high markups.

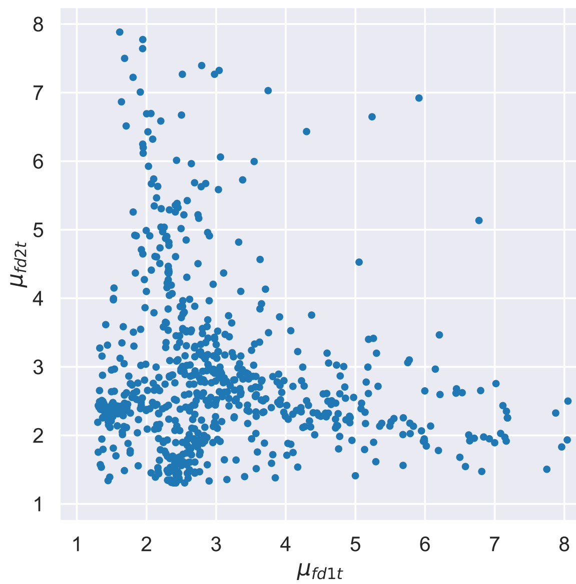
4.2.3 Markups and Trade Liberalization

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

$$\log \mu_{fdjm} = \delta_{FE} + \delta_{ti}\tau_{dt}^{ti} + \delta_{ti*BF}\tau_{dt}^{ti}D_{BF} + \mathbf{x}'\beta + \varepsilon_{fdjm} \tag{16}$$

where τ_{dt}^{ti} is input tariffs for each destination at the yearly level and δ_{FE} includes appropriate fixed effects for each level of analysis. I focus on the effects of input tariffs, given that they reflect exporters' input costs. The covariate $\mathbf{x}$ contains other tariffs such as market access

Figure 5: Correlation across products for Bundling Firms



Note: This figure scatterplots the transaction-level markups for ADPMs and accessories for bundling firms. The correlation value is -0.15, indicating firms bundle low and high markup goods.

and output tariffs along with their cross terms with bundling firm dummy. δ_{ti} captures the effect of a one unit change in input tariffs on markups for non-bundling firms, $\delta_{ti} + \delta_{ti*BF}$ captures the effect on the markups on multi-product firms with bundling.

Table 6 presents the results at various levels of markups. Because input tariffs decreased, the positive sign on the coefficient corresponds to a decrease in markups. The effect of a reduction in input tariffs for firms that sold both ADPMs and ADPM accessories and their cross terms with bundling firm dummy are positive and significant across specifications. Overall, a decrease in input tariffs brings all firms a downward trend in markups. Also, a positive δ_{ti*BF} indicates that markup differences between bundling and the non-bundling firm have decreased. Specifically, when input tariffs declined by 1% points at the baseline transaction level, markups decreased by 0.62% and 1.66% for multi-product firms without bundling and bundling firms, respectively. A similar trend holds at the firm-destination-product-yearly level where a 1% point decrease in input tariff causes a reduction of markups for multi-product firms without bundling by approximately 0.79%, but markups for bundling firms decrease by 1.94%. At the firm-destination-yearly level, non-bundling firms face a

Table 6: Markups and Trade Liberalization: Tariff Changes

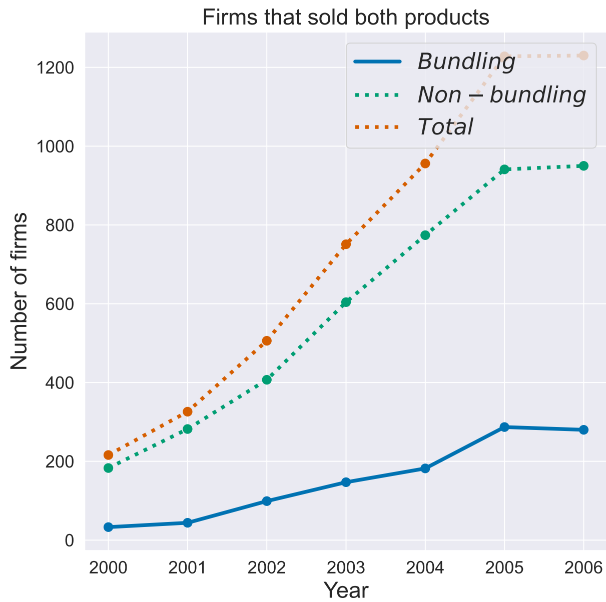
	(1)	(2)	(3)
	$log\mu_{fdjm}$	$log\mu_{fdjt}$	$log\mu_{fdt}$
τ_{dt}^{input}	0.0062***	0.0079***	0.0036**
	(0.0028)	(0.0015)	(0.0017)
$\tau_{dt}^{input} D_{BF}$	0.0106***	0.0116***	0.0243***
	(0.0052)	(0.0028)	(0.0044)
Year FE	Yes	Yes	Yes
Product FE	Yes	Yes	No
Ownership FE	Yes	Yes	Yes
F-statistic	46.57	21.26	24.51
Observation	30,887	11,582	7,362

Note: This table reports the coefficients from regression (16). The dependent variable is (log) markup. Each column is an OLS regression result of log markup on tariff changes for observations for ADPMs or ADPM accessories with various levels. Column (1) shows the results for baseline transactions, which are at the firm-destination-product-monthly level, and columns (2) and (3) show the results at the firm-destination-product-yearly level and firm-destination-yearly level. The standard errors are in parentheses and are clustered at the firm and yearly level. Significance : * 10 percent, ** 5 percent, *** 1 percent.

0.36% decrease in markups, and bundling firms face a 2.75% reduction in markups. This means that if the input tariff were to decline by 10% as it did from 2000 to 2006, multi-product firms with bundling would lose approximately 27.5% of their markups.

The pro-competitive effects of trade liberalization can explain this downward trend in markups and markup differences. Specifically, a decrease in input tariff is associated with a decrease in input costs for firms. This has two different effects. Firstly, if there is a significant incomplete pass-through of costs to prices, as found in De Loecker et al. (2016), then markups can increase with the decline of input costs. On the other hand, a decrease in input cost lowers entry barriers for firms, bringing entrants into the market, and making competition more fierce. This effect will lower the markups. I found that both bundling and non-bundling firms have nearly complete pass-through rates while the number of firms increased significantly, as shown in Figure 6.

Figure 6: Changes in composition and number of firms



Note: This figure plots the changes in the number of firms across time for both bundling and non-bundling firms in the case of ADPMs and ADPM accessories. The number of firms increased dramatically over the years.

5 Conclusion

Recently, firm-level analysis has been a central focus in research attempting to understand international trade, e.g., research on multi-product firms, productivity, networks, and markups. In this paper, I examine 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 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 previous literature. Comparing the estimated markups for Chinese exporters to markups recovered using the De Loecker and Warzynski (2012) method shows that the production-side approach may miss one important channel (*bundling*) that captures joint pricing decisions. By offering a discount, bundling firms incentivize customers to purchase more products and can leverage market power from one product market to another. Thus, bundling enables firms to increase overall firm-level markups.

My study also contributes to the literature on the relationship between markups and trade characteristics. While previous studies focused on gains from trade at the aggregate level, I study how these effects may differ across firms depending on their decision to bundle or not. Tariff reductions from trade liberalization bring pro-competitive effects by reducing the markups and markup differences across firms.

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.

Yang, Xuebing. (2021) “Markups of Chinese exporters: Export discount and product choice.” *The World Economy*, 44, 2382–2409.

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 of Profit Maximization

A.1 Derivation of the expressions

Here, I demonstrate the steps for calculating the first-order conditions in Section 2. Recall that 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 the 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 term 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 the 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 $Q_{f1t}(\mathbf{P}_{ft})$ yields

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

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 $(P_{f1t} + P_{f2t} - c_{f1t} - c_{f2t} - d_{fbt}) = (P_{f1t} - c_{f1t})\frac{P_{f1t}}{P_{f1t}} + (P_{f2t} - c_{f2t})\frac{P_{f2t}}{P_{f2t}} - d_{fbt} = (1 - \mu^{-1}_{f1t})P_{f1t} + (1 - \mu^{-1}_{f2t})P_{f2t} - d_{fbt}$. The derivatives with respect to P_{f2t} and P_{fbt} are similar and thus omitted.

B. Regression Analysis for Production–Side Markups

B.1 Framework of the Production Side

In this section, I describe how the markups using production data (section 4) were estimated. These production–side markups were recovered by directly following the method of De Loecker et al. (2016) and the Chinese Manufacturing data. Consider the 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 the inputs into a vector $\mathbf{X} = \{\mathbf{V}, \mathbf{K}\}$, and denote the price of input v 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 capitalized counterparts. Then, the 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 for 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 the 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 develop 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 to single-product firms to estimate the production functions. Specifically, I use the material demand function to develop an equation for ω_{ft}. Assume the material demand function for a 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 $\mathbf{L}_f$ are firm-specific exogenous factors such as age, location, ownership status, and affiliation status, $\mathbf{EXP}_{ft}$ is the export dummy, and $\tau^{output}_{it}, \tau^{input}_{it}$ are the output and import tariffs for industry i.

Inverting equation (18) gives the control function for the unobserved productivity ω_{ft} as $\omega_{ft} = h_t(\mathbf{x}_{ft}, \mathbf{z}_{ft})$. To construct the moment conditions, consider the following law of motion for productivity.

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

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

In the first step of estimating the production function, I separate the unanticipated shocks and/or the 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 the 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

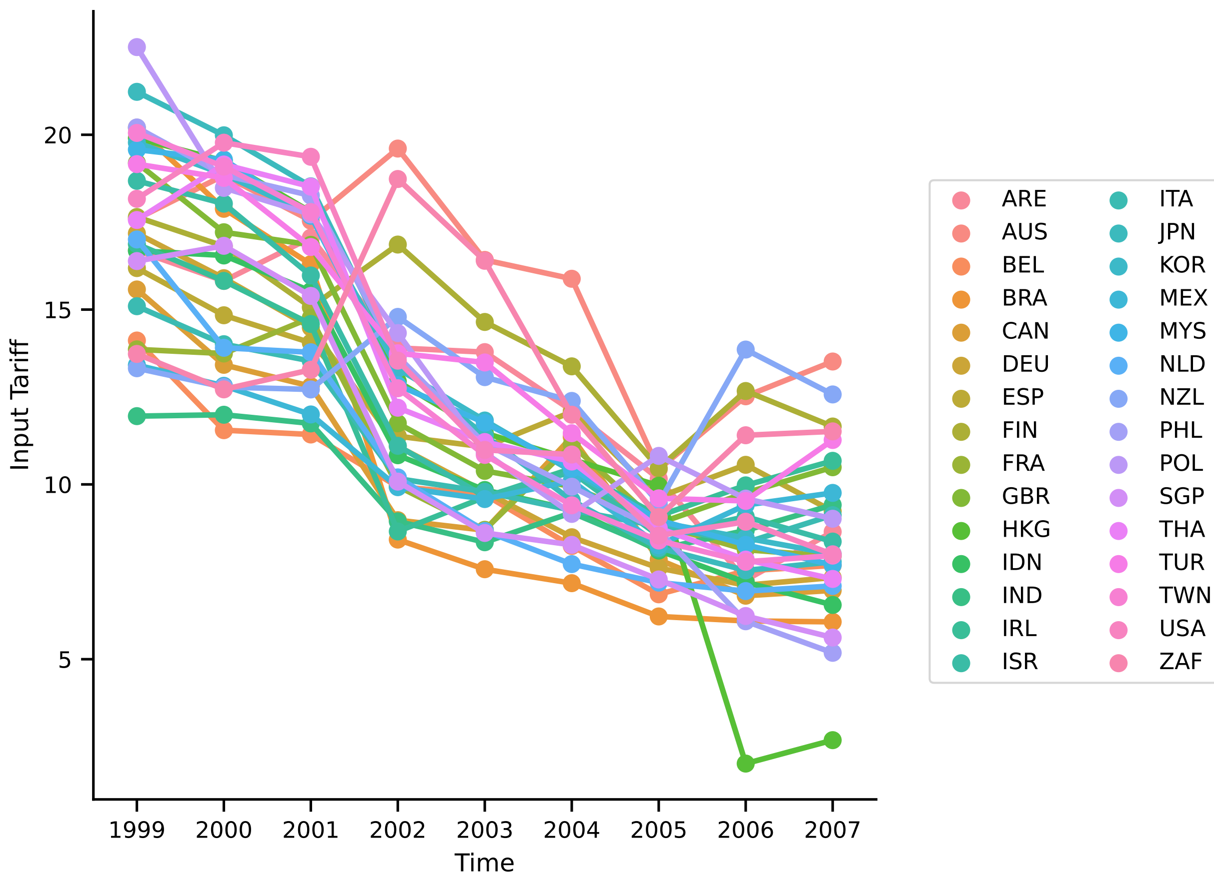
$$E[(\xi_{ft}(\boldsymbol{\beta}) + \epsilon_{fjt})\boldsymbol{Y}_{ft}] = 0 \tag{22}$$

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 and, labor, and their higher–order interaction terms, as well as the lagged output prices, lagged tariffs, and their appropriate interactions with the inputs.

C. Supplementary Documentation

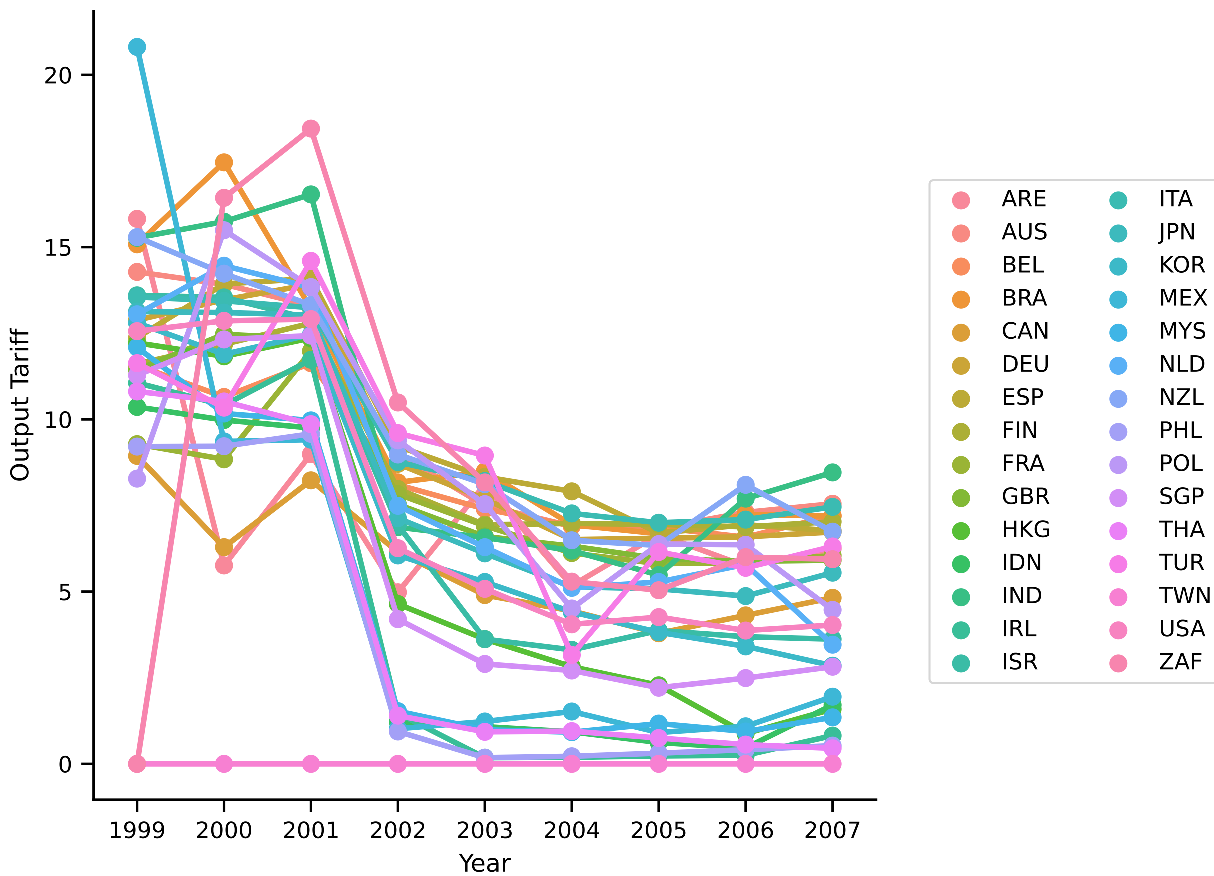
This section provides additional figures that supplement the main materials. Specifically, Figures 6 and 7 display input and output tariffs for each year at the destination market level. Figure 8 shows firm-year markups when the lognormal marginal distribution is used instead of the uniform marginal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

Figure 7: Input Tariffs for ADPMs and ADPM Accessories from 1999 to 2007



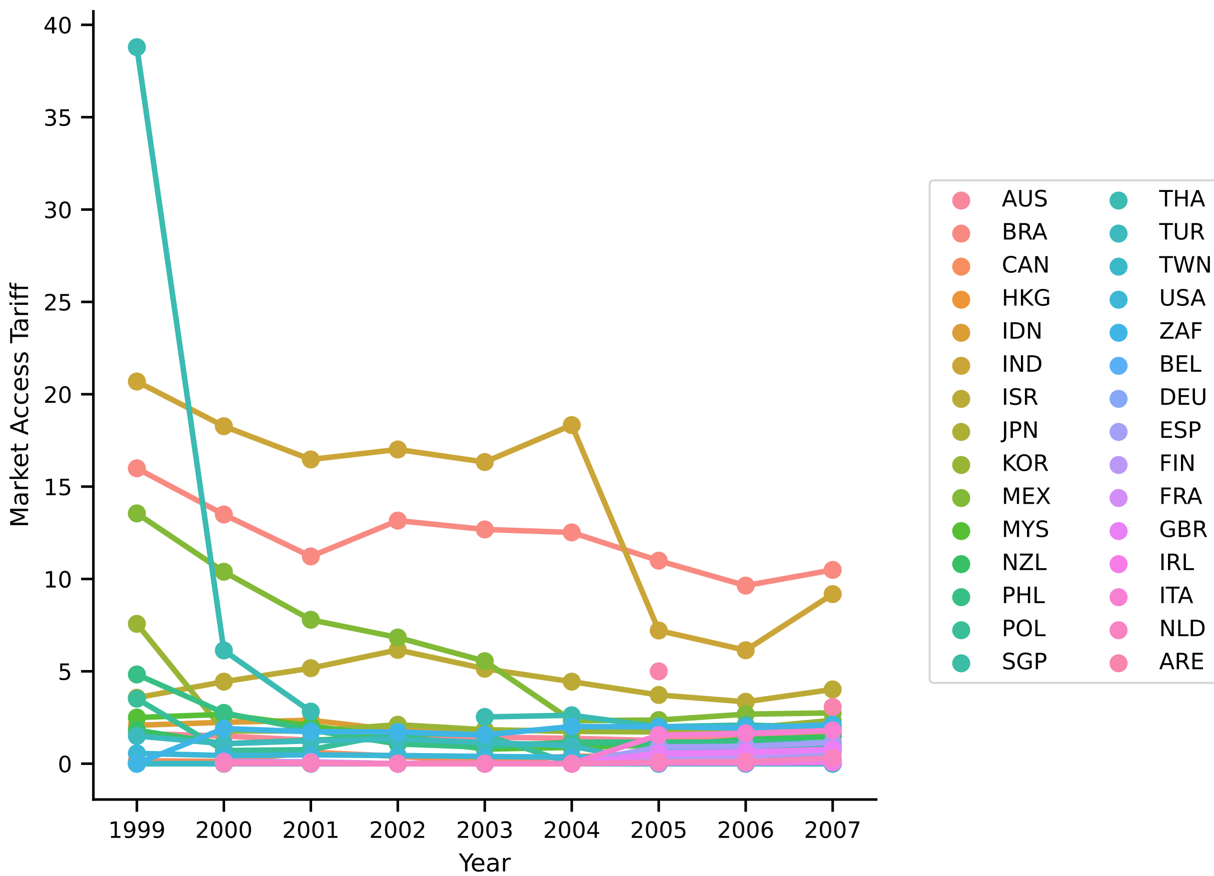
Note: This figure plots market-level input tariffs for China from 1998 to 2007 with respect to HS2 level 84, which contains ADPMs and ADPM accessories.

Figure 8: Output Tariffs for ADPMs and ADPM Accessories from 1999 to 2007



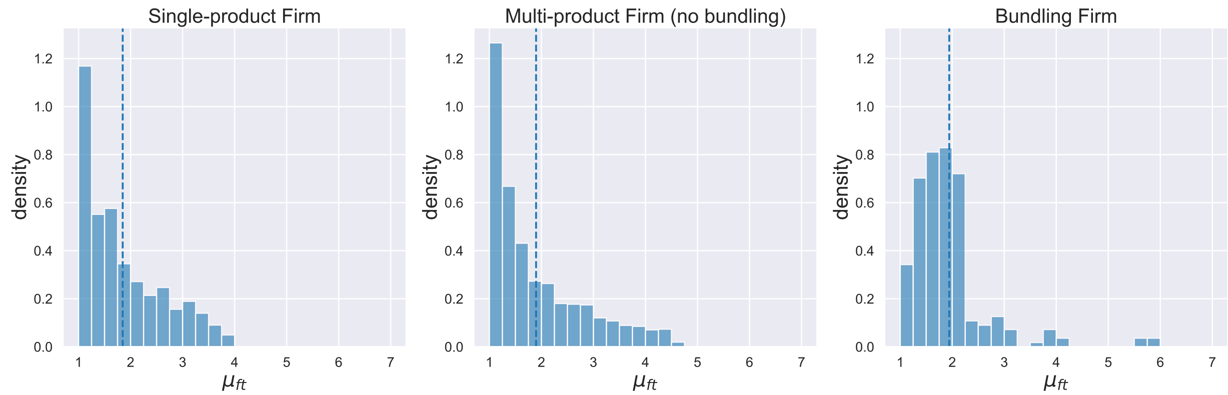
Note: This figure plots market-level output tariffs for China from 1998 to 2007 with respect to HS2 level 84, which contains ADPMs and ADPM accessories.

Figure 9: Market Access Tariffs for ADPMs and ADPM Accessories from 1999 to 2007



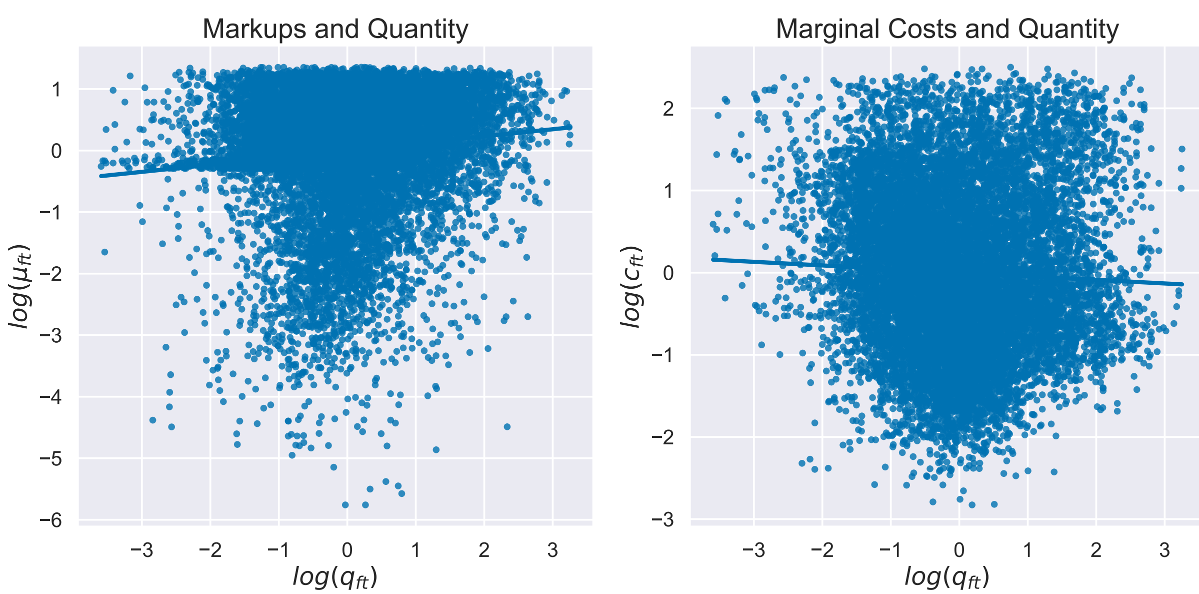
Note: This figure plots market-level market access tariffs for China from 1998 to 2007 with respect to HS2 level 84, which contains ADPMs and ADPM accessories.

Figure 10: Markups (μ_{ft}) of ADPMs and ADPM Accessories by Firm Type



Note: This figure is the same as figure 4. However, it is recovered under the assumption that consumer valuation for each product follows a lognormal distribution rather than a uniform distribution. The overall markup distribution across firm types remains the same as in the uniform case.

Figure 11: Relationship between markups/costs and quantities



Note: This figure plots the relationship between markups and quantities on the left panel and cost and quantities on the right panel. There is no clear sign for increasing or decreasing returns to scale.