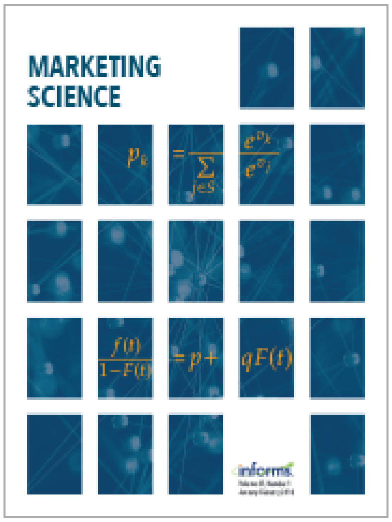


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

When Franchisee Service Affects Demand: An Application to the Car Radiator Market and Resale Price Maintenance

Tongil "TI" Kim

MARKETING SCIENCE
Vol. 40, No. 1, January–February 2021, pp. 101–121
ISSN 0732-2399 (print), ISSN 1526-548X (online)
http://pubsonline.informs.org/journal/mksc

When Franchisee Service Affects Demand: An Application to the Car Radiator Market and Resale Price Maintenance

Tongil "TI" Kim[a,b]

[a] Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080

Contact: ti.kim@utdallas.edu, https://orcid.org/0000-0003-3277-4337 (TK)

Received: February 11, 2017
Revised: June 4, 2019; September 14, 2019; April 28, 2020
Accepted: May 5, 2020
Published Online in Articles in Advance: September 25, 2020

https://doi.org/10.1287/mksc.2020.1243

Abstract. It is well understood that a downstream firm's service can impact the performance of vertical channels. Although many academic works address the service provision of the downstream firm, empirically quantifying the impact has been challenging because the downstream firm's service is often unobservable to the researcher. I propose a new empirical framework that incorporates the downstream firm's unobserved endogenous service provision by modifying the standard demand model. I apply this empirical framework to proprietary data from a franchise network in the car radiator market to quantify the downstream firms' (e.g., franchisees') endogenous service. Counterfactuals under maximum resale price maintenance (RPM) policies show that the standard demand model ignoring the franchisees' endogenous service reduction (i.e., service externality) results in more optimistic counterfactual predictions than the developed framework does. Such service externality can be mitigated if the service provision cost is lower for franchisees. Last, I examine boundary conditions: under the extreme regime of maximum RPM aiming to fully extract franchisees' profit, I find that information asymmetry is a greater concern for the upstream firm within the focal industry. Additionally, when service externality is combined with channel information asymmetry, maximum RPM at such extremes may no longer increase the franchisor's profit.

History: K. Sudhir served as the editor-in-chief and Carl Mela served as associate editor for this article.
Funding: This research was supported by a Graduate Fellowship, Dean's Dissertation Fellowship, Lam Research Fellowship, and Institute for Business Innovation Fellowship at the University of California, Berkeley. There was no funding from data provider.
Supplemental Material: Data replication files and the online appendix are available at https://doi.org/10.1287/mksc.2020.1243.

Keywords: services marketing • franchising • business-to-business (B2B) • vertical channel • empirical industrial organization • structural model

1. Introduction

In a vertical channel, a downstream firm influences customer demand by choosing the optimal mix of price and nonprice factors such as service. When ownership of the downstream firm is separate from that of the upstream firm, however, it is difficult for the upstream firm to induce a desired level of downstream firm service. If an upstream firm can perfectly monitor price and service decisions of the downstream firm at zero cost, it can write a complete contract to correct any market inefficiency. As Winter (1993) points out, however, service is often not only unobservable but also not easily contractible. This leaves the upstream firm to rely on contracting and enforcing the observed and monitored portion of the downstream firm's decisions.

Although it has been standard practice to include a downstream firm's service in analytical models, modeling it empirically has been difficult because it is often unobservable to researchers. Unlike price, data on the downstream firm's service are not readily available. Their limited availability has led empirical researchers often to assume exogeneity by not micromodeling service despite its theoretical importance. Standard random coefficient demand models (e.g., Berry et al. 1995, Nevo 2000) assume that a firm's only endogenous decision is price and lump the firm's unobserved endogenous service provision into an error term. Thus, it may not be sufficient to study a vertical channel where the downstream firm's service is often a critical decision variable. By extending the standard random coefficient demand model, I present a new empirical framework that micromodels service when it is not readily observed. This paper demonstrates that we can extend the model if we are willing to make mild assumptions on the firm's profit: the firm is assumed to incur cost to provide service as it simultaneously determines the optimal level of price and service. This supply-side assumption allows us to write first-order conditions with respect to retail price and a downstream firm's service that show the link between retail margin and service. The combination of these first-order conditions allows us to quantify the unobserved service. Although there have been studies that empirically address one type of firm's observed service (e.g., Goeree 2008, Conlon and Mortimer 2015, Murry 2017), a firm's service is multidimensional and often difficult for the researcher to observe in its entirety. Ignoring such service may lead to biases in model predictions, and this is the first paper that attempts to quantity the effect of a firm's unobserved endogenous service on demand, in conjunction with endogenous price.

To quantify unobserved endogenous service, I develop a two-step estimator where demand parameters are estimated in the first step using the standard random coefficient demand model, and the supply-side parameter (i.e., service parameter) is estimated in the second step. The traditional random coefficient demand model is a subset of the developed framework: when the service parameter is zero, the framework is statistically equivalent to the standard random coefficient demand model. I estimate the developed model using proprietary data from a national franchise network that sells automobile radiators to repair shops and dealerships in the United States. The estimation method addresses endogeneity issues regarding price and service that are potentially correlated with the demand shock and can bias parameter estimates. I address these concerns by using fixed effects and instrumental variables (IVs). The estimated service parameter is positive and statistically significant, and it suggests that the franchisee's service accounts for 9.11% of the average price of $107.22, highlighting the importance of service effort in this business-to-business (B2B) industry of a relatively commoditized product.

Unlike the standard demand model, the model developed here acknowledges that both the observed price and the unobserved service of a firm are determined simultaneously. This has substantial implications on counterfactual results. If endogenous service is important in the industry but captured as part of demand shock in the standard demand model, any counterfactuals regarding changes to the price would assume that service stays constant regardless of the profit margin, which counters the literature that connects profit margin and service level (e.g., Telser 1960, Klein and Murphy 1988, Winter 1993, Mathewson and Winter 1998). Without accounting for the unobserved endogenous service, policy simulations on price would result in biased counterfactual predictions.

I demonstrate such bias in counterfactuals with resale price maintenance (RPM). RPM involves the upstream firm enforcing pricing constraints to which a downstream firm must comply. The upstream firm can then improve the channel performance (for a literature review and further details, see Sudhir and Datta 2008). In particular, maximum RPM (or maximum resale price ceiling) has been known as a tool for an upstream firm to extract the downstream firm's rent to improve channel efficiency. However, with the presence of downstream firm service externality (i.e., the downstream firm provides endogenous service), maximum RPM's effectiveness in extracting the rent through price alone may be diminished. The first counterfactual highlights the importance of accounting for the unobserved endogenous franchisee service in predicting market outcomes of simple maximum RPM by comparing results under the full model against predictions under the standard demand model. I find that maximum RPM with a price ceiling $10 below the optimal franchisee price (or a 9.77% lower price) would result in a 10.26% reduction in the downstream firm's service level under the full model, whereas the standard demand model would predict a constant level of franchisee service despite franchisees facing a lower profit margin. Consequently, use of the standard demand model would result in rosy predictions on franchisee and franchisor profits. Such service externality may be mitigated if the service cost were lower for the franchisees, for example, improvement of software with customer relationship management (CRM) functionalities and marketing tools or financial incentives tied to certain service activity goals. The second counterfactual finds that franchisee service reduction in response to the aforementioned maximum RPM would be negligible if the service cost were 10% lower.

In the third counterfactual, I examine boundary conditions of maximum RPM. It has been noted that maximum RPM may not be sufficient to fully extract the downstream firm's rent if the downstream firm has private information about the retail cost or the local demand that the upstream does not have or the downstream firm's service externality exists (Tirole 1988, Perry and Porter 1990). Under the extreme regime of maximum RPM aiming to fully extract franchisees' profits, I find that information asymmetry is a greater concern for the upstream firm within the focal industry. Additionally, when service externality is combined with channel information asymmetry, maximum RPM at such extremes may no longer increase the franchisor's profit.

This paper proceeds as follows. I review related literature in Section 2 and discuss data and the focal industry (the car radiator market) in Section 3. I provide an illustrative example to show the importance of accounting for service in Section 4. I then specify the full econometric model with customer utility, franchisee service, and the profit function in Section 5. In Section 6, I discuss the estimation strategy, identification of service, and endogeneity issues. Estimation results (including estimated unobserved service) and validity checks are presented in Section 7.

In Section 8, I conduct counterfactuals of maximum RPM policies before the conclusion and directions for future research in Section 9.

2. Related Literature

The term *service* can be defined in many different ways depending on the business context. In studies focusing on vertical channels, a downstream firm's service is generally defined as the presale treatment or demonstration of the product (Telser 1960) or promotional activities that "make goods more attractive to customers" (Tirole 1988, p. 177). In a similar vein, Mathewson and Winter (1998) describe it as information provided to customers at the point of sale and the effort and talent of the seller. For more specific examples, Winter (1993, p. 62) characterizes service as "a short cashier line, convenient fitting rooms, well-organized inventory, prominent shelf space, and informed staff." Following the literature, I define service as multidimensional presale marketing promotional activities that increase demand.

In addition to price, the important role of nonprice competition in distribution channels has been emphasized in many studies. Mathewson and Winter (1985) note that service provision by a downstream firm may be contracted but that monitoring and enforcing it are costly. Lal (1990) finds that when the service of both the franchisor and the franchisee affects demand, royalty payments, along with franchisor monitoring franchisees, are needed. Desai and Srinivasan (1995) investigate the relationship between the lack of observability of franchisee service and price contracts in the context of a principal's signaling strategy toward a risk-neutral agent. Romano (1994), Iyer (1998), and Raju and Zhang (2005) also investigate service competition in channels.

A few empirical studies examine the effect of observed service on demand. Conlon and Mortimer (2015) estimate a model of a retailer's inventory decision while holding price fixed. Goeree (2008) models the observed advertising decision along with price in the personal computer market. Murry (2017) also studies observed advertising with endogenous price in the vertical channel of the automobile industry. Although these studies focus on quantifying the effect of a single type of observed service, a firm's service is often multidimensional and cannot be fully observed by the researcher, which can lead to biased model predictions. In contrast to the previous literature, I develop a model of unobserved endogenous service that subsumes advertising as well as other market expansion efforts, in conjunction with endogenous price, in a vertical channel. Yang et al. (2019) use this framework and apply it to multiple product settings to determine salespeople's unobserved effort using automobile dealership data.

Studies have pointed out that the service provision of the downstream firm often depends on the retail margin. Telser (1960), Klein and Murphy (1988), Winter (1993), and Mathewson and Winter (1998) discuss the link between the retail margin and the level of service. Mathewson and Winter (1998) explain that a higher margin increases the level of a downstream firm's service because of the higher marginal benefit of providing service. Such an interrelationship between the downstream price and service was also highlighted in the U.S. Supreme Court's decision to initially ban maximum RPM, citing that "[m]aximum prices may be fixed too low for the dealer to furnish services essential to the value which goods have for the consumer or to furnish services and conveniences which consumers desire" (*v. Herald Co.*, 390 U.S. 145, 1968).[1] The validity of this view may be difficult to empirically evaluate because such services are often not observed. To address the issue of unobservability of service, I combine the first-order conditions with respect to price and service. As a result, the model developed here retains the link between retail margin and service level. This link plays an important role in quantifying the downstream firm's unobserved service in policy simulations.

This paper also contributes to the literature on RPM by micromodeling the downstream firm's service. As Telser (1960) points out, the need for RPM arises if a downstream firm is a separate entity from an upstream firm. Furthermore, literature has acknowledged the importance of accounting for the downstream firm's service in studying RPM (Perry and Porter 1990, Mathewson and Winter 1998, Lafontaine and Slade 2008). In particular, maximum RPM has long been known to be effective in extracting the downstream firm's rent and mitigate double marginalization problem. However, maximum RPM is not sufficient in rent extraction once uncertainty is introduced or the downstream firm provides services to consumers that make the upstream firm's product more attractive (Tirole 1988, Perry and Porter 1990). For instance, if the downstream firm has private information about the retail cost or the local demand that the upstream firm does not have, finding the right level of price restraint will be difficult for the upstream firm. The downstream firm's service externality also exists when service is an endogenous decision in addition to price. The simplest solution to the service externality may seem to be to specify the level of promotional service in the contract. However, as Winter (1993) and Tirole (1988) note, service is difficult to observe and enforce because the courts (or involved parties) cannot measure it precisely, forcing the upstream firm to rely on restraints on easily observable and enforceable decisions such as price. The developed framework allows us to compare the two sources of difficulty in implementing maximum RPM. Results demonstrate that the downstream firm's information asymmetry is a greater concern at the boundary conditions, but service externality also plays an important role in counterfactual predictions.

3. Industry Description and Data

The focal industry in this study is the car radiator industry, from which I obtained proprietary data. In this section, I describe the specifics of the study before I discuss the data.

3.1. Car Radiator Industry Description

In the auto repair industry, service centers—repair shops, body shops, dealerships, and so on—rarely carry much inventory because the number of different parts needed to repair the immense variety of auto makes, models, years, and editions is in the millions. Instead, these service centers rely on special distributors for parts delivery when needed (e.g., when a vehicle owner brings in a car for repair). The distribution of car radiators is no exception. The radiator, which dissipates heat from the engine, comes in various sizes, materials, fits, and other specifications, making it impractical for service centers to carry extensive inventories.

I obtained sales data from a franchise network that distributes car radiators in North America. A typical business process is as follows: a radiator fails, and the driver takes the car to a service center. The service center then calls a franchisee of this car radiator distributor (and possibly its competitors) in the region; obtains information on the price, product, and service; and makes a purchasing decision. Therefore, the data record the B2B transactions between the distributor (hereafter, the *franchisee*) and the service center (hereafter, the *customer*). Note that customer here refers to the service center, not the driver of the car with a failed radiator. Franchisees set their own prices on radiators and are assumed to be risk neutral residual-claimant-profit maximizers. In addition to setting the prices, franchisees endogenously provide service to customers to expand the market and promote the business. As part of the franchising agreement, franchisees transfer a certain percentage of revenue (royalty) to the franchisor. This vertical relationship involves franchisees paying a royalty of 6%–10% of gross revenue on a typical 20-year contract.

The focal company carries a large number of noncompatible, noncompeting product lines. Each product line fits particular specifications for a car make, model, year, and edition: a radiator for a 1998 Honda CR-V, for example, will not fit into a 2011 Volkswagen GTI. In other words, each product line is considered a separate market because products across product lines are not compatible. Auto radiators have been around for a long time, and the product is relatively commoditized; model-specific radiators are manufactured by multiple different companies. Therefore, a franchisee's level of service is quite important to be able to differentiate itself from a competitor.

Each franchisee operates in an exclusive territory, defined by a discrete cluster of ZIP Codes that typically comprises 750,000 to 1.5 million people. The exclusive territory eliminates encroachment or free-riding issues in this study, which allows me to quantify the franchisee's own service without worrying about any spillover effect from other franchisees.

3.2. Service in the Industry

The service in this industry can be promotional service, sales service, managerial activities, local advertising, or any other market expansion service. Here I broadly define *service* as the business's presale input or effort that increases customer utility and effectively shifts the demand curve upward at the time of purchasing decisions. Franchisees' service is a combination of many observable and unobservable market expansion services and efforts in the industry. For instance, the count of franchisees' calls and visits may be observable if there is a central software system that keeps track of downstream firms' activities. However, the quality of these calls and visits or how knowledgeable and friendly a franchisee is toward customers in a given period is generally not observed by the franchisor or the researcher. Other examples include whether a franchisee offers extended hours of operation for customers' convenience or whether a franchisee follows the franchisor's other customer service guidelines in a given time period. Although the franchisor provides service guidelines to franchisees regarding hours of operation, product warranties, and part return policies, interviews with the franchisor and franchisees have revealed that the level of adherence to these guidelines is not consistent across franchisees or across time because monitoring and enforcement are difficult for the franchisor. In addition, individual franchisees may run local advertising (e.g., local newspaper ads), but it is not observed by the franchisor or the researcher.[2] This study explicitly models and quantifies these unobserved service activity goals, acknowledging that service provision is the franchisee's own endogenous decision.

3.3. Data

I obtain a unique data set from the franchise network that includes information about various aspects of its business.[3] First, I use two sets of proprietary data—sales transaction data and customer lookup (i.e., product inquiry) history—to obtain two product characteristics: price and the probability of same-day delivery. These data sets include information on the top 71 product lines (approximately 50% of quantity sales) in the repair business at the U.S. franchisees between January 2009 and August 2011, aggregated to monthly levels (32 time periods).[4] Among them, I focus on high-demand product lines with substantial market potential across multiple markets and franchisee markets with a substantial number of product lines in a given time to obtain reliable estimates. This results in 9,274 product line–franchisee–month unbalanced observations based on 25 unique franchisee markets and 68 product lines. Further data-cleaning details can be found in Online Appendix A.1. Sales transaction data provide us with information on the cost of goods and the final sales prices of the sold radiators. Customer lookup data contain information about whether same-day delivery was possible at the time of inquiry. If the inquired radiator model is available at the local franchise business warehouse, it can be delivered within hours, which customers highly value. If not, it may take a day or several days to ship the radiator from the manufacturer, another supplier, or another franchisee. The probability of same-day delivery of a radiator is calculated by averaging the availability at the local warehouse of each product line in each market. For the focal firm, inventory management is automated using third-party inventory management software that the franchisor deploys for the entire franchise network. Although the exact algorithm of the third-party software is proprietary, interviews suggest that the inventory replenishment algorithm is likely based on the commonly used economic order quantity, which minimizes inventory cost by assuming that demand is constant and that the inventory is depleted at a fixed rate (for a history and review of the economic order quantity literature, see Erlenkotter (1990)). Evidence also shows that the probabilities of same-day delivery of two different ownership structures (company-owned locations and franchisee locations) are very similar, suggesting that individual locations do not actively change the inventory level in response to demand shocks: the mean of same-day delivery for company-owned (franchisee) locations is 0.947 (0.948), and a two-sample t-test cannot reject the null hypothesis that they are equal ($p = 0.70$). Also, interviews revealed that unlike price, franchisees do not actively manage the inventory level (or the threshold for restocking) of hundreds of different types of radiators. Hence, the probability of same-day delivery is considered an exogenous variable and separate from franchisees' endogenous decisions.

The data have the following descriptive statistics. The average price of radiators is \$107.22 (standard deviation of \$23.55). The average probability of same-day delivery is 95% (standard deviation of 11%). The cost of goods is averaged at \$57.34 (standard deviation of \$14.30). The company also provided information about the customers' observed characteristics, including business type (repair shop, dealership, auto parts shop, or other) and ZIP Code in the customer lookup data, which I match to the 2010 U.S. Census median household income. Note that income level is an attribute of the final consumer (service center customer). The two most common customer types are radiator shops and repair shops, comprising about 74% of inquiries. The rest are junkyards, fleets, and other businesses. The mean of median household income is approximately \$52,150. Table 1 summarizes these statistics.

Note that franchisees must pay the franchisor a royalty of 6%–10% of their revenue. Lafontaine (1995, 1999), Graddy (1997), Thomadsen (2005), and Kalnins (2003) have shown that the royalty rate effectively creates a double marginalization problem by raising the transfer price to franchisees. Therefore, higher royalty rates are generally associated with higher prices, which I confirm by running regressions of price on royalty rates with product line and time dummies. The regression results in Table 2 show that a 10% increase in the royalty rate is associated with a \$6.47 increase in the average price.

I also gather data on service proxies: the average number of sales visits and calls that franchisees make in each month. Franchisees in the estimation sample average 528.42 sales visits and 154.20 calls per month (standard deviations of 569.50 visits and 345.60 calls). These data are later used to validate the estimated service in Section 7.2.2.

Market share is calculated by dividing demand by market potential. *Market potential* is defined as the

Table 1. Data Summary Statistics

Variable	Mean	Standard deviation	p10	Median	p90
Price (\$)	107.22	23.55	79.50	104.49	137.41
Same-day delivery probability	0.95	0.11	0.84	1.00	1.00
Cost of goods (\$)	57.34	14.30	40.05	56.50	76.10
Median income (\$)	52,150.53	11,500.59	40,757.80	51,358.06	66,800.00
Radiator/repair	0.74	0.16	0.52	0.75	0.93

Note. The observation is at the product line franchisee–time level.

Table 2. Price vs. Royalty Regression Results

Model:	(1)	(2)	(3)
Royalty rate (%)	0.875***	0.766***	0.647***
	(0.188)	(0.105)	(0.0984)
Constant	99.33***	—	—
	(1.689)		
Product line fixed effects	No	Yes	Yes
Year fixed effects	No	No	Yes
N	9,274	9,274	9,274
R^2	0.0020	0.7696	0.7980

Notes. The dependent variable is price. Robust standard errors are in parentheses.
$^{***}p < 0.01$.

total demand for car radiators for repair. The total annual car radiator market in the United States was estimated at $900 million in 2011 by the focal company. I first determine the total demand for radiator repairs by dividing $900 million by the average price of all radiators in my data. I then use the distributions of inquiries for each product line, franchise outlet, and time in the customer lookup data to determine market potential.[5] This results in an average market share of 22.72% (standard deviation of 7.50%) for the focal company.[6] Further details on data cleaning and market potential/share calculations are included in Section A.1 of Online Appendix A.

4. Illustrative Example

Before presenting the full econometric model, I illustrate a simple theoretical model that lays out the link between price and endogenous effort, which is the working foundation of the main model to be presented later.

Suppose that a franchisee is selling a product to customers in a single-period market with linear demand. Assume that customers dislike higher price p and prefer higher franchisee service e with diminishing marginal utility: $D(\mathrm{P}, e) = 1 - \mathrm{P} + e$. Under a typical franchise contract, a franchisee is required to pay a certain percentage of the revenue r (royalty rate) to its franchisor in order to operate as part of the franchise network. The franchisee then maximizes the following profit function over price and service: $\pi(\mathrm{P}, e) = [\mathrm{P}(1-r) - c]D(\mathrm{P}, e) - (1/2\lambda)e^2$. The first term is the profit from selling the product with a constant marginal cost c after the royalty payment $\mathrm{P} \cdot r$ to the franchisor, and the last term is the cost of providing service with a parameter λ. From the first-order condition with respect to service, the optimal level of service is expressed as $e^* = \lambda[\mathrm{P}^*(1-r) - c]$. The relationship between retail margin $\mathrm{P}^*(1-r) - c$ and demand D can be easily established from the first-order condition with respect to price. Then we can quantify service based on observed demand once λ can be empirically estimated. Combining these two first-order conditions is the key to quantifying service in the framework developed in this study.

5. Model

Now I develop a full econometric model that captures the franchisee's simultaneous decisions on price and service. I incorporate franchisee's service in the standard random coefficient demand model by extending it with assumptions on the supply side. The resulting model carries similarities to the simple theoretical model in Section 4 in that first-order conditions with respect to price and service show the link between retail margin and service. This link allows us to quantify the unobserved service.

5.1. The Demand Side

Consider an exclusive territory of the focal firm's franchise f (i.e., a geographic market f) that offers H_f noncompatible, noncompeting product lines that are effectively separate markets. For each product line h, customer i in f's territory either buys one from f or chooses the outside option. The outside option for the customer is typically either buying from a competing firm or not buying at all. Note that a car cannot be driven without a radiator; hence, not buying means disposing of the car in this industry. It is reasonable to think that the probability of disposing of a car, which is typically worth (tens of) thousands of dollars, does not significantly change in response to a change in radiator prices of around $100. Hence, I assume that a radiator is bought if the need arises. In other words, the outside option is buying from competitors.

Customer i gets the indirect utility in Equation (1) if it buys a radiator of product line h from the focal radiator distributor franchise business f at time t. Otherwise, the customer gets the normalized mean zero utility. I also allow the outside option to vary across product lines, franchisee, and time by specifying fixed effects:

$$u_{ihft} = \sum_k x_{hkft}\tilde{\beta}_{ik} + \rho_h + f(e_{ft}) + \xi_{hft} + \epsilon_{ihft}, \qquad (1)$$

$$\tilde{\beta} = \bar{\beta}_k + \sum_q a_{iq}\beta^o_{kq} + v_{ik}\beta^u_k. \qquad (2)$$

The term x_{hkft} is product line h's kth observed product characteristic. In the data, the probability of same-day delivery and price are observed product characteristics. Interviews have revealed that the focal franchise network faces different levels of competition across product lines. For instance, one of their main competitors primarily carries radiators for import cars. Therefore, different customer preferences for the focal firm across product lines are captured through product line fixed effects ρ_h.

The term $f(e_{ft})$ captures a function of the franchisee's endogenous service that each potential customer observes (and incorporates into the purchase decision) but the researcher does not observe. I assume that service enters utility linearly; that is, $f(e_{ft}) = e_{ft}$.[7]

The common demand shock ξ_{hft} is observed by customers but not by the researcher. I further decompose it to $\xi_{hft} \equiv \xi_{ft} + \Delta\xi_{hft}$ so that franchisee–time factors (e.g., different levels of competition across franchisees and time, as well as macroeconomic seasonality) can be separately captured by fixed effects ξ_{ft}. The term $\Delta\xi_{hft}$ is then the residual demand shock that is assumed to be orthogonal to IVs. The term ϵ_{ihft} is an idiosyncratic random term and is assumed to be an independently and identically distributed type I extreme value.

The random coefficient vector $\tilde{\beta}$ with respect to observed product characteristics accommodates customer heterogeneity: $\bar{\beta}$ represents the common, mean taste parameter, a_{iq} denotes customer i's qth observed attribute with parameter β^o, and v_{ik} represents an unobserved attribute with parameter β^u. Customer utility then can be rewritten as follows:

$$u_{ihft} = \delta_{hft} + \mu_{ihft} + \epsilon_{ihft}, \tag{3}$$

where

$$\delta_{hft} \equiv \sum_k x_{hkft}\bar{\beta}_k + \rho_h + \psi_{ft} + \Delta\xi_{hft}, \tag{4}$$

$$\psi_{ft} \equiv f(e_{ft}) + \xi_{ft}, \tag{5}$$

$$\mu_{ihft} \equiv \sum_k \sum_q x_{hkft} a_{iq} \beta^o_{kq} + \sum_k x_{hkft} v_{ik} \beta^u_k, \tag{6}$$

and δ_{hft} is the mean utility that is common across customers and subsumes ρ_h and ψ_{ft}. Note that ψ_{ft} consists of a function of endogenous service effort e_{ft} and exogenous demand shock ξ_{ft}. Equation (5) essentially constitutes the main equation where the relationship between demand residuals and service is estimated in Section 6.2. The term μ_{ihft} (Equation 6) accounts for customer heterogeneity in preference for different product characteristics, where the first term captures the effect of observed customer attributes a, and the last term captures the effect of unobserved attribute v. Then the probability of customer i purchasing product line h from franchise business f at time t is given by the following expression:

$$f_{ihft}(\delta_{hft}, \mathbf{X}, a_i, v_i, \boldsymbol{\beta^o}, \beta^u) \equiv \frac{\exp(\delta_{hft} + \mu_{ihft})}{1 + \exp(\delta_{hft} + \mu_{ihft})}. \tag{7}$$

The predicted market share is then calculated by integrating it over all customers:

$$s_{hft}(\delta_{hft}, \mathbf{X}, \boldsymbol{a}, \boldsymbol{v}, \boldsymbol{\beta^o}, \beta^u) = \int \frac{\exp(\delta_{hft} + \mu_{ihft}(a_i, v_i))}{1 + \exp(\delta_{hft} + \mu_{ihft}(a_i, v_i))} P(da, dv). \tag{8}$$

The distribution of the unobserved customer attribute v is assumed to be standard normal. Random draws $(\boldsymbol{a}, \boldsymbol{v})$ are taken for each product line h at each franchise business f to simulate the market share expression in Equation (8).

5.2. The Supply Side

Each franchise business f is assumed to decide on price and service level to maximize the following profit at time t:

$$\max_{p_{ft}, e_{ft}} \pi_{ft}(\boldsymbol{p}, \boldsymbol{e}; \boldsymbol{c}, \boldsymbol{s}, \boldsymbol{M}, \boldsymbol{r}, \boldsymbol{\theta}) = \sum_{h=1}^{H_f} [p_{hft}(1 - r_f) - c_{hft}] M_{hft} s_{hft} \times (\delta_{hft}, \mathbf{X}, \boldsymbol{a}, \boldsymbol{v}, \boldsymbol{\beta^o}, \beta^u) - \frac{1}{2\lambda}\sqrt{M_{ft}} e^2_{ft}, \tag{9}$$

where p_{hft}, c_{hft}, $s_{hft}(\cdot)$, and M_{hft} are the price, marginal cost, market share, and total market potential of product line h at franchisee f at time t, respectively. The variable r_f denotes the franchise royalty rate at franchise business f, and $M_{ft} \equiv \sum_{h=1}^{H_f} M_{hft}$ represents the total market potential at franchise business f at time t. The terms $\mathbf{p}$, $\mathbf{c}$, $\mathbf{s}$, $\mathbf{M}$, and $\mathbf{r}$ are the vector representations of the aforementioned terms. The vector $\boldsymbol{\theta}$ is a set of the parameters to be estimated. Note that the first term is the sum of monetary profits from selling $M_{hft}s_{hft}(\cdot)$ units of product h at a retail margin of $p_{hft}(1 - r_f) - c_{hft}$ at business f at time t. The last term represents the cost of providing service with the inverse of a service parameter $\lambda > 0$.

The total cost of service is defined by multiplying $\sqrt{M_{ft}}$ by e^2_{ft} for two reasons. First, I restrict the cost of service per customer to be quadratic (e^2_{ft}), consistent with the assumptions used in both analytical and empirical marketing studies examining various topics such as channels (Iyer 1998), advertising (Murry 2017), sales forces (Misra and Nair 2011, Chung et al. 2014, Yang et al. 2019), and teams (Villas-Boas 2020).[8] Second, given the nature of services in the industry, I assume economies of scale with respect to the total number of potential customers in the market; thus, $\sqrt{M_{ft}}$ in the cost function. Examples of services include sales visits, extended hours of operation, and advertising. Regions with large market potentials tend to be highly populated areas in the data (correlation coefficient of 0.35, $p < 0.01$); therefore, I expect economies of scale in making sales visits with large market potential. Extended hours of operation and prevalent quantity discounts in the advertising industry also suggest economies of scale with larger markets. However, if the interpretation of unobserved service differs in other studies, different functional forms may be needed. The role of different cost functions is further examined with empirical analyses in Section 7.2.3.

The first-order conditions with respect to price and service at f at time t are given by

$$(1-r_f)s_{hft} + [p_{hft}(1-r_f) - c_{hft}]\frac{\partial s_{hft}}{\partial p_{hft}} = 0, \quad \forall h = 1, \ldots, H_f, \tag{10}$$

$$\sum_{h=1}^{H_f}[p_{hft}(1-r_f) - c_{hft}]M_{hft}\frac{\partial s_{hft}}{\partial e_{ft}} - \frac{1}{\lambda}\sqrt{M_{ft}}e_{ft} = 0, \tag{11}$$

where

$$\frac{\partial s_{hft}}{\partial p_{hft}} = \int f_{ihft}(a_i, v_i)[1 - f_{ihft}(a_i, v_i)] \times \frac{\partial(\delta_{hft} + \mu_{ihft}(a_i, v_i))}{\partial p_{hft}} P(da, dv), \tag{12}$$

$$\frac{\partial s_{hft}}{\partial e_{ft}} = \int f_{ihft}(a_i, v_i)[1 - f_{ihft}(a_i, v_i)]P(da, dv). \tag{13}$$

From Equation (10), the expression for the optimal price is given by $p^*_{hft} = c_{hft}/(1-r_f) - s_{hft}/(\partial s_{hft}/\partial p_{hft})$. The intuition is straightforward: the optimal price is higher when the marginal cost is higher and/or the magnitude of price sensitivity ($\partial s_{hft}/\partial p_{hft}$) is smaller (i.e., demand is less elastic). The royalty rate, $0 < r_f < 1$, effectively creates a double marginalization problem by raising the transfer price to franchisees by $1/(1-r_f)$. Hence, the optimal price increases in the royalty rate, as shown in Table 2.

From Equation (11), the optimal service can be expressed as

$$e^*_{ft}(\cdot) \equiv \lambda \sum_{h=1}^{H}\left\{[p_{hft}(1-r_f) - c_{hft}]\frac{\partial s_{hft}}{\partial e_{hft}}\right\}\frac{M_{hft}}{\sqrt{M_{ft}}}. \tag{14}$$

More service is exerted when the service parameter λ is larger. This expression can be considered to be a form of a weighted sum of retail margins. The intuition is clear: the franchisee exerts more service when the retail margin, $p_{hft}(1-r_f) - c_{hft}$, is larger, which is consistent with the intuition from the simple linear demand model in Section 4. Note that this expression contains unobserved marginal cost c, so we cannot directly quantify service from this equation. Fortunately, this marginal cost (or profit margin) is estimated in the first-order condition with respect to price.

Note that the first-order conditions with respect to price establish the relationship between observed market shares and unobserved profit margin, which are used to estimate the marginal cost in standard demand models. In addition, the first-order conditions with respect to service define the relationship between the profit margin and unobserved, endogenous service. Combining the two first-order conditions in Equations (10) and (11), the optimal service can be expressed as a weighted sum of observed market shares, with weights being functions of royalty rates, market share derivatives (i.e., sensitivities with respect to price and service), and market potential:

$$e^*_{ft}(p, c, r, M; \psi, \beta^o, \beta^u, \lambda) \equiv \lambda\left[-(1-r_f)\sum_{h=1}^{H}s_{hft}\frac{\partial s_{hft}/\partial e_{ft}}{\partial s_{hft}/\partial p_{hft}}\frac{M_{hft}}{\sqrt{M_{ft}}}\right]. \tag{15}$$

This equation allows us to quantify the level of service in terms of observed and simulated variables: royalty rate, market potential, and market shares are observed, and the derivative terms can be obtained through simulation. Once the service parameter λ is estimated, unobserved endogenous service can be quantified.

Note that λ essentially provides a statistical test of service: if the null hypothesis of $\lambda = 0$ cannot be rejected, then service is not significant in the industry. In other words, the traditional random coefficient demand model is a subset of the developed framework: when $\lambda = 0$, the framework is statistically equivalent to the standard random coefficient demand model.

I now illustrate how the service parameter, along with demand parameters, can be estimated.

6. Estimation

To quantify unobserved endogenous service, I develop a two-step estimator where demand parameters are estimated in the first step using the standard random coefficient demand model (e.g., Berry et al. 1995, 2004; Petrin 2002), and the supply-side parameter λ is estimated in the second step using Equations (5) and (15).[9] To preview the estimation method, I first recover franchisee–time fixed effects ψ_{ft} in Equation (4). Then, in the second step, I estimate Equation (5) by regressing the estimated fixed effects ψ_{ft} on the service expression in Equation (15), where the λ can be estimated in a linear regression. This approach has similarities to the use of brand-specific dummy variables in Nevo (2000, 2001) in that the dummy variables are first estimated in a random coefficient demand model, and the estimated brand effects are regressed on product characteristics to recover taste coefficients. Whereas Nevo (2000, 2001) regresses the estimated fixed effects on fully observed product characteristics, my approach regresses the fixed effects on the unobserved service, expressed as a function of observed and simulated quantities from first-order conditions.

This approach has a few advantages. First, including franchise–time fixed effects in the first step improves the fit of the demand-side estimation. Also, this reduces the correlation between price and the demand shock $\Delta\xi_{hft}$ by explicitly accounting for franchisee–time variation using fixed effects. These are the reasons echoed in Nevo (2000, 2001), advocating the inclusion of brand-specific dummy variables in a demand model estimation. Note that to introduce franchisee–time fixed effects, we require observations on more than one product line, which we have for the focal firm. This two-step approach also allows us to conduct various robustness checks on identification assumptions in recovering service.

6.1. First Step: Demand Estimation

The first step of the estimation allows us to recover demand parameters, including product line fixed effects ξ_h and franchisee–time fixed effects ξ_{ft} using the standard random coefficient demand model estimation.

6.1.1. Berry et al. (1995) Moments. The first set of moments matches the predicted market shares $s_{hft}(\cdot)$ to the observed market shares in the data s^N_{hft}:

$$G_1(\boldsymbol{\psi}) \equiv s_{hft}(\boldsymbol{\delta}(\boldsymbol{\theta}), \boldsymbol{\theta}) - s^N_{hft} = 0, \ \forall h, f, t, \qquad (16)$$

where θ is the set of parameters to be estimated, including ρ_h and ψ_{ft}. Berry (1994) shows that there is a unique value of δ that matches these two market shares, which can be found using a contraction mapping as in Berry et al. (1995). Additional moment conditions can be constructed by making assumptions on the demand shock $\Delta\xi$ that it is uncorrelated with a set of IVs including exogenous variables Z:

$$G_2(\boldsymbol{\theta}) \equiv E\left[\Delta\xi_{hft}(\boldsymbol{\theta})\,|\,Z_{hft}\right] = 0, \ \ \forall h, f, t. \qquad (17)$$

Price in Equation (4) is potentially correlated with the residual demand shock $\Delta\xi_{hft}$; this can create an endogeneity problem, as Villas-Boas and Winer (1999) note. The residual demand shock is not observed by the researcher, but the franchisee observes it and takes it into account when deciding on price and service, which can bias the parameter estimates. To address this endogeneity concern, I collect the total number of manufacturers of a given radiator product line h at time t at other noncompeting franchisees to instrument radiator prices. The product inquiry data from the focal franchise network allow one to identify manufacturers that supply a particular radiator to each franchisee at a given time.[10] From an exclusion standpoint, the residual demand shock $\Delta\xi_{hft}$ at franchisee f is assumed to not be correlated with the number of radiator manufacturers at other noncompeting franchisees. Although this assumption is likely to hold under exclusive territories of the focal franchise, it may be violated when franchisees have substantial overlaps of territories. The correlation between price and this IV arises because more manufacturers of a given product line means more bargaining power for the franchisor to lower the price for all franchisees while negotiating with manufacturers. Other IVs include the squared value of the IV and its interaction with the same-day availability, along with other exogenous variables. Consistent with literature, results show that the price sensitivity is significantly attenuated under ordinary least squares (OLS); further analyses examining the IVs are included in Section B.1 of Online Appendix B.

6.1.2. The Micromoments. The third moment matches the predicted covariance between product characteristics x_{hft} and observed customer attributes $\boldsymbol{A}$ (a dummy for radiator/repair shop and median household income) to the observed covariance. As Berry et al. (2004) show, these moments are particularly useful to pin down microparameters $\boldsymbol{\beta^o}$. For each kth product characteristic in each market, the following micromoments can be defined:

$$\begin{aligned} G_3(\theta) &\equiv x_k\left\{\frac{1}{n_b}\sum_{i_b=1}^{n_b}\boldsymbol{A}_{i_b} - E\big(\boldsymbol{A}\,|\,y_i=1,\boldsymbol{\beta}\big)\right\} \\ &\approx x_k\left\{\frac{1}{n_b}\sum_{i_b=1}^{n_b}\boldsymbol{A}_{i_b} - \frac{\frac{1}{N_s}\sum_r \boldsymbol{A}_r P\big(y_i=1\,|\,A_r, v_r, \boldsymbol{\beta}, \boldsymbol{\delta}(\boldsymbol{\beta})\big)}{\boldsymbol{s}^N}\right\} \\ &= 0, \end{aligned} \qquad (18)$$

where n_b is the total number of customers that bought the product, and i_b is the index for customer i that bought the product. The distribution of customer attributes may vary across locations and time, so I take draws of customer attributes at the franchisee and month levels.[11]

6.1.3. The Objective Function. Denoting the sample analogue of stacked moment conditions $(G_2(\theta)G_3(\theta))'i$ as $\hat{m}(\boldsymbol{\theta})$, the optimal estimators are obtained using a two-stage generalized method-of-moments approach (Hansen 1982) with the following expression:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min}\ \ \hat{m}(\boldsymbol{\theta})'\,\hat{W}(\boldsymbol{\theta})\,\hat{m}(\boldsymbol{\theta}), \qquad (19)$$

where the weighting matrix $\hat{W}$ is an identity matrix in the first stage and $E[Z'\xi(\hat{\boldsymbol{\theta}}_{\mathbf{1}})\xi(\hat{\boldsymbol{\theta}}_{\mathbf{1}})'Z]^{-1}$ in the second stage using the estimates from the first stage $\hat{\boldsymbol{\theta}}_{\mathbf{1}}$. Details on the asymptotic variance and standard error calculations can be found in Section B.3 of Online Appendix B.

6.2. Second Step: Service Estimation

Once demand parameters are estimated, we can then proceed to estimate supply-side parameter or service parameter λ in the second step. Here I estimate $\psi_{ft} \equiv f(e_{ft}) + \xi_{ft}$ (Equation (5)) to quantify unobserved endogenous service e_{ft}. Note that Equation (15) can be rewritten as $e_{ft} = \lambda X_e$, where X_e is expressed as follows:

$$X_e \equiv -(1 - r_f) \sum_{h=1}^{H} s_{hft} \frac{\partial s_{hft}/\partial e_{ft}}{\partial s_{hft}/\partial p_{hft}} \frac{M_{hft}}{\sqrt{M_{ft}}}. \quad (20)$$

The term X_e can be considered a synthetic regressor generated by theory after combining first-order conditions with respect to price and service. Fortunately, variables in X_e are either observed or can be simulated. Because ψ_{ft} is estimated in the first step, λ can be estimated in a linear regression assuming that ξ_{ft} is a structural error at the franchisee–time level that is uncorrelated with X_e.

Note that e_{ft} is uniquely defined for the following reasons. First, X_e is a set of uniquely defined regressors in the second step because δ_{hft} is uniquely found in the first step via contraction mapping (Equation (16)). With the estimated δ_{hft}, all terms in X_e including $\partial s_{hft}/\partial e_{ft}$ and $\partial s_{hft}/\partial p_{hft}$ can be calculated. In other words, X_e is uniquely calculated given the unique values of δ_{hft}. I then estimate Equation (5), $\psi_{ft} \equiv f(e_{ft}) + \xi_{ft} \equiv \lambda X_e + \xi_{ft}$, in a regression setting to pin down λ. Putting it differently, I first obtain the synthetic regressor X_e from the first step and then estimate the λ parameter in the second step, uniquely defining service $e_{ft} = \lambda X_e$.

6.2.1. Identification Assumptions. Estimating the second step in a linear regression will result in biased estimates if ξ_{ft} is correlated with X_e. This is a valid concern because ξ_{ft} is part of δ_{hft} (Equations (4) and (5)), which affects s_{hft}, $\partial s_{hft}/\partial e_{ft}$, and $\partial s_{hft}/\partial p_{hft}$ terms in X_e (Equation (20)).[12] I address this endogeneity concern in two ways. First, I use the local sales tax as an IV. The local sales tax varies at the county level and is assumed to affect the total number of cars sold and hence the total market potential M (the number of cars whose radiators need to be repaired) in X_e. In contrast, the sales tax on radiator repair is considered to have little correlation with the customer's choice of this radiator network over outside options ξ_{ft}. The exclusion restriction is that the focal firm's customers (repair shops, dealerships, etc.) are not likely to shop across county lines with differing sales taxes in mind to save money on a \$100 radiator. Because the sales tax rate instruments service that is unobserved, it is difficult to conduct a test for validation. However, if it is valid, we would expect a strong relationship between the sales tax rate and partial components of service that are observed (i.e., sales visits and calls).[13] I find that the observed service components are negatively correlated with the sales tax rate. Additionally, the sales tax rate is negatively correlated with market potential, as expected. Further analyses on the IV can be found in Section C.1 of Online Appendix C.

Second, I estimate the second step using a fully saturated fixed-effects model: franchisee fixed effects (to control for market-specific characteristics or qualities as well as time-consistent fixed service effort by different franchisees) and monthly fixed effects (to control for any time-varying macroeconomic shocks and seasonality that may affect service level).

As will be discussed in Section 7.2.1, results show that λ under OLS is estimated to be significantly smaller than the estimates with an IV or saturated fixed effects.

7. Results

7.1. Demand Estimates

Table 3 presents demand taste parameter estimates. Their signs are as expected: Customers prefer same-day delivery (0.6046, $p < 0.01$) and low prices on average (−0.1014, $p < 0.01$). Customer heterogeneity parameters on price are also found to be significant: customers in areas with a higher income are less price sensitive (0.0008, $p < 0.05$), and the two largest customer segments (radiator/repair shops) are less price sensitive (0.0026, $p < 0.01$) than other types of customers. The unobserved taste parameter is also statistically significant (0.0321, $p < 0.01$), capturing other unobserved heterogeneous customer attributes with respect to price. The resulting own price elasticity is −2.82, on average. Heterogeneous taste parameters with respect to same-day delivery are found to be not statistically significant.

Table 4 shows the summary statistics of estimated product line fixed effects (ρ_h) and franchisee–time fixed effects (ψ_{ft}). There is substantial variation in both sets of estimates, resulting in standard deviations

Table 3. Demand Parameter Estimates: Taste Parameters

Taste parameters	Variable	Estimate	Standard error	p-value
Macro parameters	Price (\$)	–0.1014***	0.014	<0.001
	Same-day delivery	0.6046***	0.1927	0.002
Microparameters on price	Income (00K)	0.0008**	0.0004	0.029
	Radiator/repair shops	0.0026***	<0.001	<0.001
	Unobserved	0.0321***	<0.001	<0.001

Note. Robust standard errors are reported.
$^{**}p < 0.05$; $^{***}p < 0.01$.

Table 4. Demand Parameter Estimates: Fixed Effects

Fixed effects	Minimum	Median	Mean	Standard deviation	Maximum	No. of estimates significant at 5%
Product line (ρ_h)	−8.599	−6.203	−5.9425	1.5043	−1.7493	66/67
Franchisee-time (ψ_{ft})	−1.960	0.252	0.259	0.7211	2.0496	197/449

of 1.5043 and 0.7211, respectively. More important, many of the fixed-effects estimates are statistically significant at 5%.

The marginal cost vector c_{hft} estimated from the first-order condition in Equation (10) has a mean of \$62.7, a median of \$61.2, and a standard deviation of \$17.2. Because this company is a parts distributor, not a manufacturer, the largest component of the marginal cost is the cost of goods, or the cost of radiators, and I find that approximately 90% of the estimated marginal cost counts toward the cost of goods (\$57.3, on average).

7.2. Service Estimates

The standard demand model would stop at the demand estimation in the preceding section, assuming that any franchisee–time unobserved and endogenous service is captured by ψ_{ft} and is constant under any policy changes. In contrast, this paper separately recovers the unobserved endogenous service portion of ψ_{ft} as outlined previously.

7.2.1. Service Parameter Estimation. I estimate $\psi_{ft} \equiv f(e_{ft}) + \xi_{ft}$ (Equation (5)) to quantify unobserved endogenous service e_{ft}, which can be rewritten as $e_{ft} = \lambda X_e$, where X_e is shown in Equation (20). Considering X_e as a synthetic service regressor derived from first-order conditions, I can then regress ψ_{ft} on X_e to estimate λ with the linear assumption $f(e_{ft}) = e_{ft}$. In this section, I estimate this model and compare it against IV models and a saturated fixed-effects model discussed in Section 6.2.1.

First, I plot estimated franchisee–time fixed effects ψ_{ft} against X_e in Figure 1 to visually the check the relationship. The plot shows a positive relationship between the recovered franchisee–time fixed effects from the standard demand estimation and the analytical synthetic service regressor, separately derived from the first-order conditions. This implies that franchisees' service increases ψ_{ft}, which subsequently increases customer utility.[14]

Column (1) in Table 5 shows the results of OLS regression, and the service parameter λ is estimated to be positive and significant (0.0029, $p < 0.01$), as Figure 1 demonstrates. In column (2), I use the local sales tax rate as an IV[15] and find a significantly larger estimate (0.0097, $p < 0.01$), highlighting the importance of addressing the endogeneity concern. The first-stage F-statistic is 50.20, implying sufficient explanatory power of the IV. The results are similar with

Figure 1. (Color online) Scatter Plot of the Second-Stage Estimation

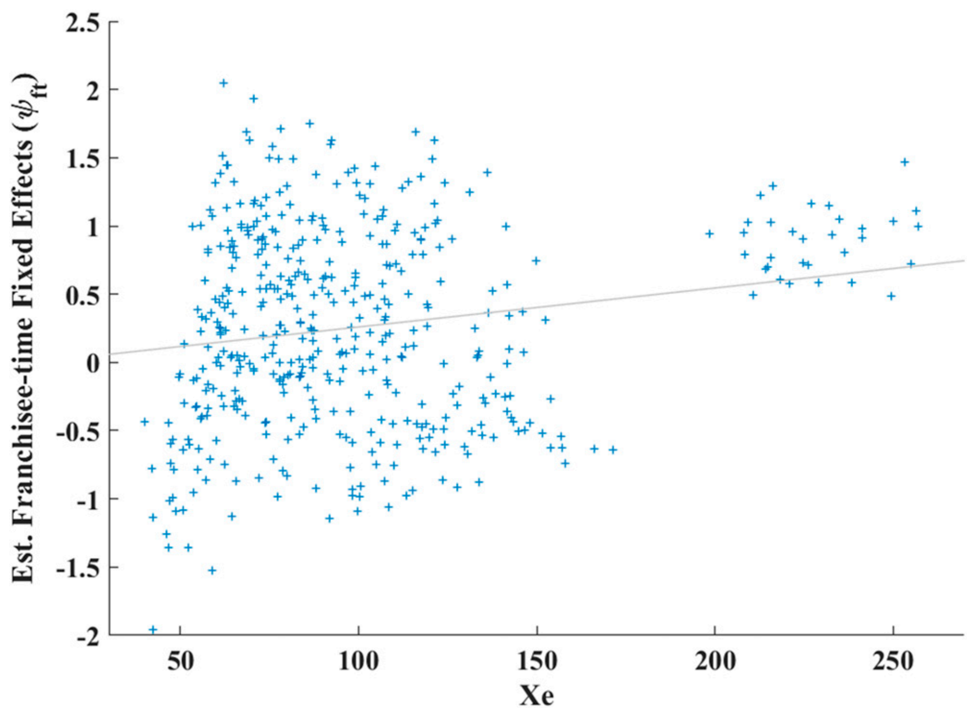


Table 5. Service Parameter Estimates

	OLS	IV 1	IV 2	Saturated fixed effects
	(1)	(2)	(3)	(4)
λ	0.0029***	0.0097***	0.0116 ***	0.0124***
	(0.0009)	(0.0033)	(0.0027)	(0.0015)
Constant	–0.0265	–0.7076**	—	—
	(0.0951)	(0.3152)		
Franchisee fixed effects	No	No	No	Yes
Monthly fixed effects	No	No	Yes	Yes
N	450	450	450	450
R^2	0.199	—	—	0.844
First-stage partial R^2	—	0.101	0.1014	—
First-stage F-statistic	—	50.20	47.04	—

Notes. Standard errors reported in parentheses. Standard errors in columns (1) and (2) are bootstrapped standard errors incorporating the imprecision from the first step. Standard errors for models with fixed effects (columns (3) and (4)) are asymptotically robust standard errors because bootstrapping with fixed effects generates missing observation issues. OLS, ordinary least squares; IV, instrumental variable.
$^{**}p < 0.05$; $^{***}p < 0.01$.

the inclusion of monthly fixed effects in column (3). Column (4) shows the results of the saturated fixed-effects model in which franchisee fixed effects and monthly fixed effects are included to control for franchisee-specific and time-specific factors: λ is slightly larger and significant (0.0124, $p < 0.01$), but it is not statistically different from the IV estimates. To be conservative, subsequent analyses are conducted based on the IV estimate of column (2).[16]

One potential criticism of this approach is that because the dependent measures in this second step are estimated measures from the demand estimation in the first step, errors from the first step may lead to an inefficient estimate of λ. I find that bootstrapped standard errors incorporating the imprecision from the first step are only slightly larger than unadjusted standard errors and do not qualitatively affect overall results. Details on bootstrapping and result comparison can be found in Section C.3 of Online Appendix C.

7.2.2. Estimated Service and External Validation. The franchisee service measure e_{ft} can be quantified from the data by substituting in the estimated parameters into Equation (15). Because resulting monthly service is measured in utility, I convert it to dollars by dividing it by the estimated price coefficient. Figure 2 presents the estimated unobserved endogenous service level in dollars for a sample of franchisees. The graph shows a substantial variation of service across time as well as across franchisees. Note that franchisee 47 provides much higher levels of service than other franchisees in the graph. Interestingly, the focal company's chief marketing officer corroborated that franchisee 47 was internally perceived as the best-run franchisee.

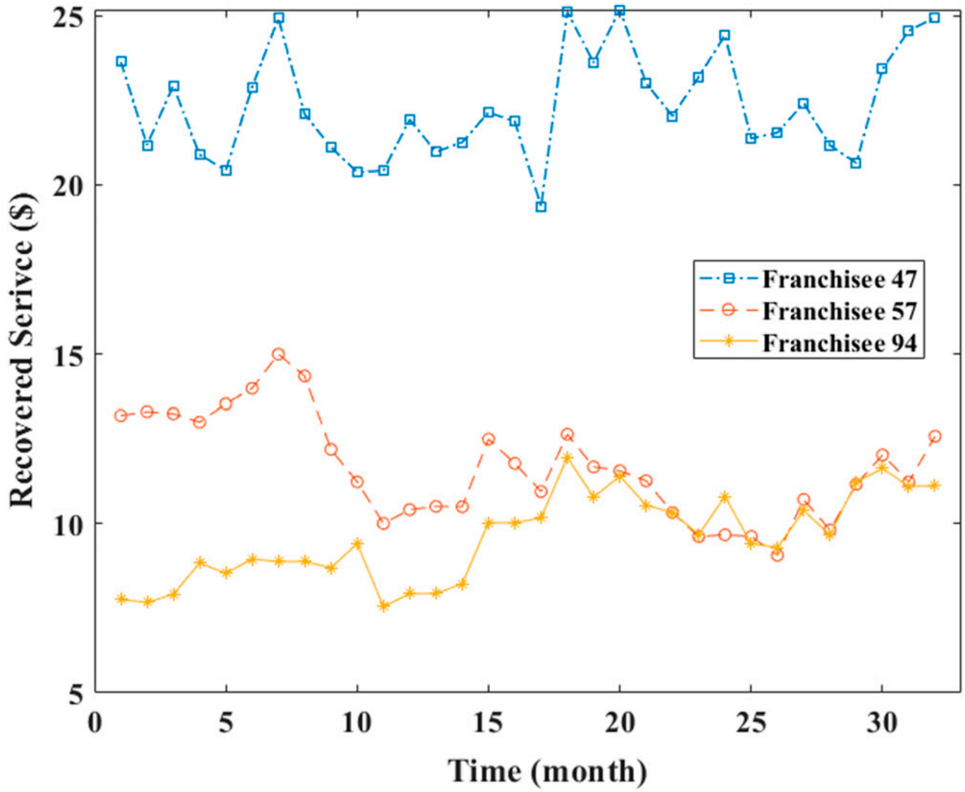
Overall, the resulting service ranges from \$3.89 to \$25.18, with a mean of \$9.77 and a standard deviation of \$4.38.[17] This suggests that service accounts for 9.11% of the average price of \$107.22, highlighting the importance of service effort in this B2B industry of a relatively commoditized product.

Because this paper aims to quantify unobserved service, inherently, there are limited ways to validate the estimated service. In order to conduct an external validation, I gather data on observed promotional activities. The focal company collects data on the frequency of visits and calls at franchisees over time. Franchisees regularly visit and call customers. Making sales visits entails discovering customers' needs and providing them with promotional materials (e.g., magnets, candies, coupons, and gift cards). Sales calls can also be made over the phone. If the sales visits and calls are indeed a subset of the franchisee's service that is captured by the model, the recovered service should be positively correlated with them, which is checked using linear regressions.[18]

Results in Table 6 show positive relations between estimated service and the numbers of calls and visits. It is important to note that the service considered in this study comprises multidimensional activities, which include promotional services, sales calls and visits, managerial activities, local advertising, or any other market expansion effort. Here I was able to collect only data on proxies that capture the frequencies of visits and calls; other metrics of service, such as the quality of those visits and calls, adherence to service guarantees mandated by the franchiser, local advertising, or employment of other sales tactics, are not available. Despite the limited scope of available proxies, I observe an R^2 of 0.563 with visits and calls in the regression.

7.2.3. Functional Form Assumptions. Recall that given the nature of services in this industry, economies of scale are assumed in the cost function (i.e., $(1/2\lambda)\sqrt{M_{ft}}e_{ft}^2$). In this section, I examine how cost functions affect

Figure 2. (Color online) Recovered Unobserved Service



service estimation by presenting results of the second-step service estimates under different assumptions (diseconomies of scale and constant returns to scale) using homogeneous logit models.[19] Column (1) of Table 7 shows estimates of the second step with service economies of scale—the most appropriate assumption based on the economic interpretation of services in the data. Column (2) presents results under service diseconomies of scale (e.g., the service cost in profit is $(1/2\lambda)M_{ft}^2e_{ft}^2$), and column (3) shows estimates under the assumption of constant returns to scale in service provision (e.g., the service cost is $(1/2\lambda)M_{ft}e_{ft}^2$).

All specifications show that the service parameter λ is positive and statistically significant, implying unobserved endogenous service in the data. To better understand economic implications, services ($e_{ft} = \lambda X_e$) are again converted to dollar terms in the last two rows. Columns (1) and (2) exhibit similar average services of \$5.75 and \$4.69, respectively. In contrast, column (3) shows a substantially larger service measure, at \$29.16. This is mainly driven by the negative and statistically significant constant term: when X_e is zero (e.g., retail margins of all product lines are zero in Equation (14)), franchisee–time fixed effects ψ_{ft} have a negative mean. One may choose to include the constant term in the conversion, which then results in a comparable service level of \$5.35. However, I prefer to treat the constant term as part of exogenous service because it does not vary with franchisee incentives, for example, retail margin.

The constant-returns-to-scale assumption in column (3) also produces the least conservative measure relative to the standard demand model estimation that assumes away unobserved endogenous service. Counterfactuals based on this assumption will show larger divergence between this framework and the standard demand estimation, further demonstrating the importance of accounting for unobserved endogenous services. This example also demonstrates that choosing

Table 6. Estimated Unobserved Service vs. Proxies

	Calls only	Visits only	Calls, visits
Model:	(1)	(2)	(3)
Calls	0.011***		0.009***
	(0.001)		(0.001)
Visits		0.010***	0.008***
		(0.001)	(0.001)
Constant	11.730***	8.114***	7.896***
	(0.248)	(0.428)	(0.342)
N	450	450	450
R^2	0.362	0.349	0.563

Notes. The dependent variable is the estimated unobserved service. Asymptotically robust standard errors are in parentheses.
$^{***}p < 0.01$.

Table 7. Functional Forms

	(1)	(2)	(3)
Service cost in profit	Economies of scale	Diseconomies of scale	Constant returns to scale
λ	0.0006***	22.34***	0.119***
	(0.00008)	(5.075)	(0.00737)
Constant	−0.0114	0.0186	−0.677***
	(0.0283)	(0.0323)	(0.0554)
R^2	0.0703	0.046	0.293
N	450	450	450
Average of service ($)	5.75	4.69	29.16 (5.35 including constant term)
Standard deviation of service ($)	2.70	2.20	5.52

Note. Asymptotically robust standard errors are reported in parentheses.
***$p < 0.01$.

the functional form purely based on the data fit ($R^2 = 0.293$) may not be ideal in this framework.[20]

To summarize, the magnitude of estimated service depends on the functional form assumptions, and careful justification based on the domain knowledge of the nature of services in the data is needed. This limitation of functional form assumptions affecting estimated measures is not unique in dealing with unobserved endogenous service. Nevo (2000, pp. 10–11) notes that the patterns of price elasticity are determined by the functional form of observed price (e.g., linear price, logarithmic price). A different approach may be needed to relax these functional form assumptions and eventually estimate this nonparametrically, and I leave this as an area for future research.

8. Counterfactuals: Maximum Resale Price Maintenance

The main distinction between this framework and the standard demand estimation arises when some components of alternative-specific dummies ξ in the standard demand estimation are endogenous as opposed to assumed exogenous demand shock. For instance, when a firm whose unobserved endogenous service is captured in ξ faces some price restraint from its upstream firm or a regulatory agency, counterfactuals using the standard demand model would assume that ξ stays constant, which counters the literature that connects profit margin and service level (e.g., Telser 1960, Klein and Murphy 1988, Winter 1993, Mathewson and Winter 1998). Without accounting for the unobserved endogenous service, policy simulations would result in biased predictions, and the framework developed in this study can address this issue by econometrically measuring it.

Accounting for unobserved service under a price restraint is particularly relevant in the case of maximum RPM. Maximum RPM allows the upstream firm to set a price ceiling for the downstream firm's price as a way for the upstream firm to control for the retail price and extract the downstream firm's rent, which may benefit final customers. Although maximum RPM has been theoretically known to solve double marginalization problems in channel management,[21] it may not be an effective vertical channel control tool when the downstream firm's service externality exists (Tirole 1988, Perry and Porter 1990). The standard demand model does not explicitly model such service externality, assuming that the franchisee does not change its service level even when its profit margin is squeezed by a maximum price restraint. On the contrary, the framework developed in this study models the franchisee's endogenous service, allowing it to change in response to the price restraint. The first counterfactual highlights the importance of accounting for the unobserved endogenous franchisee service in predicting market outcomes by comparing results under the full model against predictions under the standard demand model (Section 8.1).

In Section 8.2, I examine the role of service cost. One natural way to mitigate franchisees pulling back service to final customers in response to maximum RPM is to lower the cost of service provision for franchisees. In this counterfactual, I show that a lower service cost can counter franchisees' service reduction in response to maximum RPM.

In Section 8.3, I consider boundary conditions of maximum RPM by conducting a thought experiment of an extreme case where the franchisor aims to fully extract franchisees' rent. Maximum RPM may be ineffective in fully extracting the downstream firm's rent if the downstream firm has private information about the retail cost or the local demand that the upstream does not have or the downstream firm's service externality exists (Tirole 1988, Perry and Porter 1990). Therefore, I compare the two sources of hindrance in implementing maximum RPM at the extreme.

Note that counterfactual results in this section are partial equilibrium results: impacts on franchisor, franchisees, and customers are studied under the assumption that competitors do not react to the counterfactual changes in the focal franchise network. In all counterfactuals, I assume that franchisees decide on the sales price (as long as it satisfies the maximum resale price restraint) and provide service. The focal franchise network uses a universal point-of-sales software system that tracks sales of all radiators, so the final sales price is fully observed.

8.1. Accounting for Unobserved Endogenous Service

In this section, I compare predictions of maximum RPM under the standard demand model (i.e., invariant unobserved franchisee service) and under the full model developed in this study (i.e., unobserved endogenous franchisee service). As a simple example, suppose that the franchisor aims to enforce maximum RPM by setting a price ceiling that is \$10 below the franchisee's optimal price in the data.[22] For instance, if a franchisee's optimal price for a radiator is \$107.22 in the current data, the counterfactual tests to see how the market outcomes would have been different had the franchisor enforced a maximum RPM at \$97.22. Column (1) of Table 8 lists the average values of various market outcomes (price, franchisee service, market shares, etc.) observed in the data, where there is no price restraint. Column (2) presents counterfactual results of the proposed maximum RPM under the standard demand model that assumes away downstream service externality, whereas column (3) shows the results under the full model that accounts for service externality. Both columns (2) and (3) report the mean percentage change relative to the observed data in column (1), along with confidence intervals.

First, as discussed previously, the standard demand model implicitly assumes a constant level of franchisee service in the face of maximum RPM. In contrast, the full model predicts a 10.26% reduction in the franchisee service level (column (3)), which amounts to a \$1 decrease from the estimated \$9.77 service in Section 7.2.2. Consequently, the standard demand model overestimates the impact on market shares: column (2) shows the market share increasing from 0.227 to 0.299 (31.96% increase), whereas column (3) predicts the market share increasing from 0.227 to 0.295 (30.26% increase). It is then not surprising that the standard demand model overestimates the increase in franchisor royalty profit (18.29% increase in column (2) versus 17.15% in column (3)). Results also show that the price restraint's effect on franchisee profit is smaller when ignoring unobserved endogenous franchisee service (4.06% decrease in column (2) versus 4.21% decrease in column (3)). In other words, the standard demand model makes "rosy" predictions of maximum RPM because it ignores the downstream firm's service externality.

Note that results in Table 8 are based on a mild maximum RPM (i.e., \$10 below the observed price), which explains the small differences regarding whether service externality is considered (i.e., column (2) versus column (3)). The importance of accounting for service externality increases as we consider more restrictive maximum RPM. For instance, under maximum RPM at \$15 below the observed price, counterfactuals with and without service externality predict franchisor profit increases of 25.97% and 28.25%, respectively. Full counterfactual results under this maximum RPM can be found in Section D.1 of Online Appendix D. At the boundary condition of full rent extraction from franchisees, counterfactuals with and without service externality predict franchisor profit increases of 63.88% and 74.22%, respectively, resulting in a substantially more optimistic prediction when service externality is ignored (see columns (1) and (3) of Table 10).

Table 8. Prediction Comparisons Under Maximum RPM at \$10 Below Franchisee Optimal Price

Model:	(1)	(2)		(3)	
Scenario:	No restraint (average price = \$107.22)	Counterfactual: Maximum RPM at \$10 below (average price = \$97.22)			
		Under the standard demand model (no service externality)		Under the full model with unobserved endogenous service (with service externality)	
Model used:	Current data Average	Average	Mean Δ% [95% CI]	Average.	Mean Δ% [95% CI]
Franchisee service (utility)	0.285	0.285	0.00% [0.00%, 0.00%]	0.255	−10.26% [−10.47%, −10.04%]
Market share	0.227	0.299	31.96% [31.82%, 32.11%]	0.295	30.27% [30.13%, 30.41%]
Franchisor profit (\$)	3,696.31	4,361.70	18.29% [18.14%, 18.44%]	4,302.94	17.15% [17.00%, 17.31%]
Franchisee profit (\$)	13,606.84	13,065.78	−4.06% [−4.12%, −3.99%]	12,160.20	−4.21% [−4.28%, −4.14%]

Notes. The maximum RPM here assumes that the price ceiling is set at \$10 below the franchisee optimal price. Franchisor profit refers to franchisor royalty profit. Franchisee profit includes the economic cost of service provision. Results for price and market shares, including standard errors for the percentage difference between the observed data and the counterfactuals, are calculated across 9,274 product line franchisee–month observations. Other results are calculated at and across franchisee–month observations.

8.2. Role of Service Cost

In this section, I explore the role of service cost in maximum RPM. Note that franchisees incur costs in providing service to customers, and the cost of service provision is captured as $1/\lambda$ in Equation (9). How would the counterfactual results differ if the cost were 10% lower?[23] Lowering the cost of franchisee service provision in the focal firm can take many forms. For example, the franchisor may improve the franchise network enterprise software with advanced CRM functionalities such as daily sales visit recommendation engines and marketing email and call management tools to make service provision easier for franchisees. Another possibility is to provide financial incentives tied to certain service activity goals, thus effectively reducing franchisees' cost to provide service. Intuitively, a lower service cost should increase the optimal service level, making maximum RPM a more attractive vertical control tool to the franchisor via mitigating service externality.

Table 9 compares the counterfactual results with the estimated cost of service in the current data (column (2)) with those with a 10% lower service cost (column (3)) under the previously discussed maximum RPM (i.e., a price ceiling $10 below the franchisee optimal price). It also reports the percentage changes of market outcomes relative to column (1). Recall that the full model previously predicted 10.26% less franchisee service (column (2)). If the franchisee's cost of service provision were 10% lower, however, the franchisee's service reduction would be negligible (0.14%; 95% confidence interval (CI) [−0.10%, 0.38%]), as shown in column (3). In other words, a lower service cost can counter the service externality problem when employing maximum RPM. Without substantial franchisee service externality, market shares and franchisor profit are higher in column (3). It is also inferred that franchisee profit increases when the service cost is lower. However, the net franchisor profit may be lower if the cost of lowering service cost (e.g., investment in the software technology) is high and falls on the franchisor.

8.3. Boundary Condition: Full Rent Extraction

It has been noted that maximum RPM may not be sufficient to fully extract the downstream firm's rent if the downstream firm has private information about the retail cost or the local demand that the upstream does not have or the downstream firm's service externality exists (Tirole 1988, Perry and Porter 1990). While the concept of full rent extraction discussed in analytical works may not be realistic in empirical settings, because it ignores issues such as participation constraints (i.e., franchisees may leave the network with zero profit) and out-of-sample predictions, it provides a thought experiment on boundary conditions of maximum RPM rather than a direct prescription of practical implementation. For instance, if a franchisor were to consider an extreme regime of maximum RPM aiming at extracting a substantial portion of franchisees' rent, which factor (private information or service externality) would be more important? In this section, I compare the two sources of hindrance in implementing full rent extraction maximum RPM. In addition to comparing predictions with and without service externality, I compare counterfactuals with varying degrees of downstream private information: a franchisor may not fully extract a franchisee's rent if the downstream firm has private information about marginal cost.[24] Under the full-information condition, I assume that the franchisor knows the marginal costs of the franchisees. Under the private-information condition, I assume that the franchisor knows only the distribution of the marginal cost for each product line but not the actual marginal cost of each franchisee at a given time.

Table 9. Role of Service Cost

Model:	(1)	(2)		(3)	
Scenario:	No restraint (average price = $107.22)	Counterfactual under the full model with maximum RPM at $10 below (average price = $97.22)			
		With estimated service cost		With 10% lower service cost	
Service cost:	Average	Average	Mean Δ% [95% CI]	Average	Mean Δ% [95% CI]
Franchisee service (utility)	0.285	0.255	−10.26% [−10.47%, −10.04%]	0.284	0.14% [−0.10%, 0.38%]
Market share	0.227	0.295	30.27% [30.13%, 30.41%]	0.299	31.94% [31.80%, 32.08%]
Franchisor profit ($)	3,696.31	4,302.94	17.15% [17.00%, 17.31%]	4,361.50	18.28% [18.13%, 18.42%]
Franchisee profit ($)	13,606.84	12,160.20	−4.21% [−4.28%, −4.14%]	12,241.97	−3.75% [−3.82%, −3.69%]

Notes. The maximum RPM here assumes that the price ceiling is set $10 below the franchisee optimal price. Franchisor profit refers to franchisor royalty profit only. Results for price and market shares, including standard errors for the percentage difference between the observed data and the counterfactuals, are calculated across 9,274 product line franchisee–month observations. Other results are calculated at and across franchisee–month observations.

In the first counterfactual, I assume that the franchisor fully observes franchisees' marginal cost (e.g., no private information), and there is no service externality (e.g., franchisee service is fixed with respect to any price restraints). To fully extract rent, the franchisor would set the maximum resale price such that franchisees' profit margin $p(1-r)-mc$ becomes zero, and the resulting demand changes are estimated using the standard demand model, which subsumes service as part of franchisee–time fixed effects ψ_{ft}.

In the second counterfactual, I relax the full-information assumption on the marginal cost by assuming that the franchisor knows only the distribution of marginal costs for each product line. In this counterfactual, the price restraint is set such that $p_{hft}(1-r_f)-E(mc_h)$ is zero, where $E(mc_h)$ is the expected marginal cost for product line h. Note that for some franchisees in a given time, this may result in negative profit margins if $E(mc_h)$ is lower than the true marginal cost mc_{hft}. In those cases, I assume that the franchisees do not offer those product lines.

Next, I relax the assumption that there is no service externality. In the third counterfactual, I assume that the franchisor fully observes franchisees' marginal cost, and franchisees endogenously change the level of unobserved service in response to the price restraint.

In the last counterfactual, I relax both assumptions in that the franchisor knows only the distribution of marginal costs, and franchisees endogenously determine unobserved service in response to the price restraint. The full rent extracting maximum price restraint results in a 32%–36% reduction in price (from \$107.22 to approximately \$69 per radiator) depending on the counterfactual scenarios (see Online Appendix D for detailed results, including raw data).

The resulting market outcome changes relative to the observed data under various scenarios are shown in Table 10. First, only counterfactuals (3) and (4) allow endogenous franchisee service in response to the full rent extraction maximum RPM, predicting 100% and 86% franchisee service reductions. As for demand, assuming full information and no service externality, the first counterfactual predicts a 173% increase in demand (counterfactual (1)). Franchisees' private information lowers this figure to 143% (counterfactual (2)); because of the information asymmetry in the channel, the franchisor's price restraint based on expected marginal cost makes franchisees fail to offer 32% of product lines whose realized profit margins would be negative. In counterfactual (3), assuming full information and endogenous service would result in a 150% demand increase that is smaller than the 173% in counterfactual (1), implying that service externality substantially undermines the upstream firm's vertical channel control on the downstream firm through a price restraint. Information asymmetry in franchisees' marginal cost would further hamper the price restraint's effect on demand (only 125% demand increase in counterfactual (4)).

Counterfactual (1) in Table 10 also shows that under full information without service externality, maximum RPM increases the franchisor's profit by 75%, suggesting that the maximum price restraint would be an attractive vertical channel control tool to the franchisor. However, the second counterfactual shows that information asymmetry substantially reduces the appeal, predicting only a 6.6% increase in its profit. This is mainly driven by the fact that the franchisor can set the price based only on the distribution of marginal costs rather than the true marginal costs, forcing franchisees to stop offering 32% of product lines because of negative profit margins.[25]

The downstream firm's service externality can also undermine the effectiveness of maximum RPM. Counterfactual (3) in Table 10 predicts that under full information, an attempt to fully extract franchisees' rent by imposing maximum RPM increases franchisor profit by 64% (as opposed to 75% without service externality) because franchisees will stop providing service to customers. Furthermore, service externality

Table 10. Full Rent Extraction Counterfactuals Under Various Scenarios

	Mean Δ% relative to observed data [95% CI]			
Model:	(1)	(2)	(3)	(4)
Condition:	Full info + no service externality	Info asymmetry + no service externality	Full info + service externality	Info asymmetry + service externality
Franchisee service	0.00% [0.00%, 0.00%]	0.00% [0.00%, 0.00%]	−100.00% [−100.00%, −100.00%]	−86.23% [−87.46%, −85.00%]
Market share	172.97% [172.29%, 173.64%]	143.30% [142.27%, 144.34%]	150.25% [149.62%, 150.88%]	124.99% [124.04%, 125.94%]
Franchisor profit	74.77% [73.73%, 75.81%]	6.61% [2.47%, 10.75%]	63.88% [62.75%, 65.01%]	0.52% [−3.40%, 4.43%]
Franchisee profit	−100.00% [73.73%, 75.81%]	−81.46% [2.47%, 10.75%]	−100.00% [62.75%, 65.01%]	−81.74% [−3.40%, 4.43%]

Notes. The counterfactual maximum RPM here sets the price ceiling to fully extract franchisees' rent, resulting in 32%–36% lower prices. The table shows the mean percentage change under the counterfactual relative to the current data, which do not have any resale price restraint. I also include 95% confidence intervals (CIs) in brackets. Results for market shares are calculated across 9,274 product line franchisee–month observations. Other results are calculated at and across franchisee–month observations. Raw data are included in Online Appendix D.

combined with information asymmetry (counterfactual (4)) makes maximum RPM an ineffective vertical control tool, predicting a statistically insignificant 0.52% (95% CI [−3.40%, 4.43%]) increase in franchisor profit.

These results show that at least in these data, information asymmetry undermines the effectiveness of maximum RPM more than service externality does. However, when combined (counterfactual (4)), they can erase the entire benefit of maximum RPM to the upper-stream firm. Therefore, it is important to account for the downstream firm's service externality in addition to information asymmetry in the upstream firm's decision on employing price restraint as a way to manage the vertical channel. Ignoring the need to account for service externality simply because of its lack of data would overstate counterfactual results, and this study's framework of econometrically quantifying service addresses such an issue to some extent.

Last, when the upstream firm has full information on the downstream firm, the upstream firm can set the price restraint such that it leaves no rent for franchisees (counterfactuals (1) and (3)). With uncertainty in marginal cost, however, the franchisor cannot fully extract rent, leaving franchisees with approximately 20% of the original profit in counterfactuals (2) and (4).

All counterfactual results in the study are partial equilibrium counterfactuals, meaning that the estimates are based on the assumption that market participants other than the focal franchise network do not strategically respond to the prescribed policy change, mainly because the data in this study contain information from the focal franchise network only. A potential criticism, then, is that all other products are lumped into the outside good, which would likely bias counterfactual results. In other words, such counterfactual results may be overly optimistic because competitive responses are not fully accounted for and modeled. This study lacks data about competitors, and quantifying such competitive response is difficult because of the data limitation. Regardless, I conduct a limited sensitivity analysis against one type of competition scenario, and its details can be found in Section D.3 of online Appendix D.

9. Conclusion and Future Research

Both practitioners and researchers have well understood the importance of a downstream firm's role in a vertical channel because the downstream firm chooses its own optimal mix of price and service provision. Because service is often difficult to observe, monitor, and enforce, economically quantifying it has been a challenge to researchers. Here I propose an empirical framework that micromodels and quantifies the role of the downstream firm's unobserved endogenous service.

Standard random coefficient demand models that are similar to that of Berry et al. (1995) assume that a firm's only endogenous decision is price. This paper argues that we can do more than lump the firm's unobserved endogenous service provision into an error term if we are willing to make a mild assumption on the firm's profit: the firm incurs a cost to provide service because it determines the optimal level of price and service. Using this supply-side assumption, we can write first-order conditions with respect to retail price and a downstream firm's service. By combining these first-order conditions, we can quantify the unobserved service. To quantify unobserved endogenous service, I develop a two-step estimator where demand parameters are estimated in the first step using the standard random coefficient demand model, and the supply-side parameter (i.e., service parameter) is estimated in the second step. The estimation method addresses endogeneity issues with saturated fixed effects and IVs.

The framework here is particularly applicable to industries where service is a salient factor in demand and/or the product is not heavily differentiated (i.e., commoditized). I apply the model to a unique data set—sales transactions, customer information, and franchise arrangement—obtained from a franchise network in the car radiator market, in which an unobserved (to the researcher) service level set by the franchisee is an important factor in the industry. The model estimates that the franchisee's service accounts for 9.11% of the average price of $107.22, highlighting the importance of service effort in this B2B industry of a relatively commoditized product. I validate this measure against field evidence and show that it is positively correlated with observed components of franchisee service.

If endogenous service is important in the industry but captured as part of demand shock in the standard demand model, policy simulations on price can result in biased counterfactual predictions. I demonstrate such biases in counterfactuals using maximum RPM. For example, a maximum RPM with a price ceiling $10 below the optimal franchisee price (or a 9.77% lower price) would result in a 10.26% reduction on the downstream firm's service level under the full model, whereas the standard demand model implicitly assumes a constant level of franchisee service in the face of maximum RPM. Consequently, the standard demand models would make rosy predictions on channel profits. Such a service externality could be countered if the service cost were lower for franchisees; the second counterfactual shows that the service externality would be negligible if the service cost were 10% lower under the aforementioned maximum RPM scenario. In the third counterfactual, I examine boundary conditions of maximum RPM. The literature has pointed out two potential factors that can undermine a upstream firm's ability to extract rent from the downstream firm: information asymmetry in the channel and the downstream firm's service externality. Using the developed framework, I examine how these two factors make it more difficult for the franchisor to fully extract rent. Results show that when both information asymmetry and downstream service externality exist, maximum RPM aiming at fully extracting franchisee rent will not increase franchisor profit.

This paper primarily focuses on presale service because the focal company deems postsale service insignificant relative to presale service in this industry. As a distributor, the company only sells car radiators and does not install them. Because the failure rate of parts is small, there is little postsale service in the focal company. A similar approach may be applied to industries where recovering unobserved postsale service is relevant, and it would be an interesting topic for future research.

As noted previously, the focal franchise network guarantees exclusive territory to franchisees. This provides a unique setting for this paper to abstract away from any spillover effect from nearby franchisees and to focus on quantifying franchisee service. In other settings where free riding, encroachment, or other spillover among franchisees is important, future work may explore ways to extend this model to recover unobserved service of neighboring franchisees and enrich our understanding of franchise management.

Although cost functions were chosen based on an understanding of the economics of the industry, different functional-form assumptions resulted in different magnitudes of service. Getting rid of functional forms altogether and estimating service magnitude nonparametrically would be a natural next step, and I leave this as an area for future research.

Last, the framework presented in this study focuses on a single franchise network because of data limitations. Although I control for some aspects of competition using fixed effects and conduct a sensitivity analysis, this may not fully capture competitive responses. To model unobserved endogenous service in a multifirm setting, one would need to obtain data from multiple firms of multiple product lines and modify the first-order conditions for multiple firms, along with considerations on ownership structure (i.e., royalty r_j for firm j) and asymmetric service parameters (i.e., different λ for each firm). Future research may examine such a multifirm setting and shed light on boundary conditions in which a competitor decreases or increases price and service in response to a firm's lowered price and service because of resale price restraint.

Acknowledgments

The author is indebted to his advisor, Miguel Villas-Boas, for his guidance and teachings for this dissertation work. The author is also grateful to his dissertation committee members—Minjung Park, Denis Nekipelov, and Lucas Davis—and other marketing faculty members and colleagues at Haas School of Business for their insightful and constructive feedback. Thanks also to seminar participants at the University of Washington; Emory University; George Washington University; Washington University, St. Louis; University of Toronto; University of Wisconsin, Madison; Carnegie Mellon University; Cheung Kong Graduate School of Business; University of California-Los Angeles; and Stanford University, as well as participants at the Marketing Science Conference for many helpful comments and suggestions.

Endnotes

[1] In *State Oil Company v. Khan* (522 U.S. 3, 1997), however, the court overruled *Albrecht v. Herald Co.* (390 U.S. 145, 1968). The court found that maximum RPM should be analyzed under "the rule of reason," arguing that maximum RPM lowers prices, which benefits customers. In 2007, minimum RPM became rule of reason in Leegin Creative Leather Products, Inc. v. PSKS, Inc. (551 U.S. 877, 2007) as well.

[2] A search in the Kantar Media database turned up no major advertising for the focal company brand in 2010 other than a very small number of internet ads in January and February.

[3] The company provided the data to the author confidentially. The author received no financial compensation from the company.

[4] The model is estimated at the aggregate level because there is no panel structure to the transaction data. Also, even at the transaction level, there are limited observable sources of heterogeneity that could help improve the analysis through estimation at the individual transaction level. If this framework were applied to rich individual-level transaction data, it might be preferable to estimate the first-step demand model using maximum likelihood to recover alternative-specific fixed effects in each market (see, e.g., Goolsbee and Petrin 2004). The rest of the estimation in quantifying service could then follow in the second step.

[5] Note that a customer (e.g., repair center) may call this franchise network to buy a radiator or purchase one elsewhere without calling this network. Therefore, this calculation of market potential relies on the assumption that the distribution of inquiries at the focal company is not significantly different from those at competitors.

[6] One may also allocate market potential across products and franchisees based on some baseline size of the franchisee market for radiators. Unfortunately, such data were not available for the focal industry.

[7] Alternatively, one may assume a logarithmic function of service in the utility. However, because the logit function is highly nonlinear, the logarithmic function of service does not imply diminishing returns because the shape of the response function depends on the market share. Therefore, I simply allow service to enter the utility linearly without imposing any specific functional forms, similar to price entering the utility in a linear manner.

[8] It also ensures a unique analytical expression of service (Equation (14)) in the first-order condition with respect to service (Equation (11)).

[9] I thank an anonymous reviewer and the associate editor for suggesting this estimation framework.

[10] The manufacturer's IVs vary across product lines, franchisees, and time. Data variation is shown in Section B.1 of Online Appendix B.

[11] The simulation sample size is 150 in the estimation. The derivation of the second equation in Equation (18) involves Bayes' rule and is outlined in Section B.2 of Online Appendix B. More details can be found in Berry et al. (2004).

[12] The sign of $\partial X_e/\partial\xi$ depends on the data, implying that the correlation between ξ and X_e can be either negative or positive. In Table 5, I show that the service parameter increases with an instrumental variable and fixed-effects approaches, suggesting a negative relationship between ξ and X_e in the focal data.

[13] I thank an anonymous reviewer who suggested this test.

[14] The cloud on the left consists of observations whose market potential is less than 4,712, whereas the one on the right consists of observations whose market potential is greater than 5,446. Two markets with 4,712 and 5,446 market potentials are not fundamentally different, and the market potential gap is not substantial in a linear manner. However, use of the square root of the market potential (i.e., $\sqrt{M_{ft}}$ for economies of scale in cost) in calculating X_e makes the gap larger because the two market potentials lie in the flat part of the square root function. In comparison, a plot with linear M_{ft} (i.e., constant return to scale) does not show multiple groups of data points. Further discussions with additional plots can be found in Section C.2 of Online Appendix C.

[15] Note that tax rates vary across franchisees (mean of 0.075, standard deviation of 0.012, minimum of 0, and maximum of 0.0925) only because they were obtained as a snapshot at the time of data extraction.

[16] One can calculate own service elasticity with respect to market shares $(\partial s_{lhft}/\partial e_{ft})(e_{ft}/s_{lhft})$. Service elasticities are found to be 0.099–0.427, depending on the model specifications in Table 5. Yang et al. (2019) find similar but smaller own service elasticities of 0.079–0.199 in automobile dealership data in Japan. This may be because radiators are much cheaper than cars, allowing the marginal service to have a greater impact on purchase probability.

[17] The standard deviation is the sample standard deviation of the mean based on 450 franchisee–time observations.

[18] Alternatively, one may include visits and calls as controls in the second-step estimation, which would result in a statistically significant lambda and visit parameter (0.004 and 0.753, respectively, with $p < 0.01$) but not a significant call parameter (0.222, $p = 0.132$). Estimated visits and calls in this approach, however, do not change in response to price restraints. The proper modeling approach would require additional IVs and accounting for their costs in the profit function (e.g., Goeree (2008) with advertising) so that visits and calls change in response to price restraints, along with unobserved endogenous service. Because the focus of this study is to show that the proposed framework can quantify and recover unobserved endogenous service, the numbers of sales visits and calls were used for postestimation validity checks rather than being directly incorporated into the framework to quantify their effects.

[19] Changing the cost function assumption consequently changes the service regressor X_e in the second-step estimation. A logit model is recommended for this step so that many functional forms can be tested quickly without computational burden.

[20] See discussions on other functional forms in Section C.4 of Online Appendix C.

[21] There are other vertical control tools such as full-line forcing, brand discounts, rebates, and slot allowance (e.g., Shaffer 1991a, b). This study focuses on maximum RPM that links between profit margin and service. Mathewson and Winter (1998, p. 59) call RPM "the most important vertical restraint in terms of both the frequency of use and the number of legal cases generated."

[22] Assume that the franchisor has full downstream information and can calculate the franchisees' optimal price without price restraint.

[23] The cost of service takes the inverse form of λ (i.e., $\frac{1}{\lambda}$) so that the optimal service is linear in λ (Equation (15)). Therefore, a 10% lower service cost equates to replacing $1/\lambda$ with $0.9/\lambda$ in Equation (9).

[24] Similarly, the downstream firm may have private information about the local demand. However, in the policy simulation of full rent extraction where the downstream firm decides on the price, the private information on the marginal cost is most relevant to the upstream firm, which sets the price restraint in relation to the marginal cost.

[25] Franchisees may choose to exit the market, further reducing the franchisor's royalty profit.

References

Berry ST (1994) Estimating discrete-choice models of product differentiation. *RAND J. Econom.* 25(2):242–262.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Berry S, Levinsohn J, Pakes A (2004) Differentiated products demand systems from a combination of micro and macro data: The new car market. *J. Political Econom.* 112(1):68–105.

Chung DJ, Steenburgh T, Sudhir K (2014) Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. *Marketing Sci.* 33(2):165–187.

Conlon CT, Mortimer JH (2015) Efficiency and foreclosure effects of all-units discounts: Empirical evidence. Working paper, New York University, New York.

Desai PS, Srinivasan K (1995) Demand signalling under unobservable effort in franchising: Linear and nonlinear price contracts. *Management Sci.* 41(10):1608–1623.

Erlenkotter D (1990) Ford Whitman Harris and the economic order quantity model. *Oper. Res.* 38(6):937–946.

Goeree MS (2008) Limited information and advertising in the us personal computer industry. *Econometrica* 76(5):1017–1074.

Goolsbee A, Petrin A (2004) The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica* 72(2):351–381.

Graddy K (1997) Do fast-food chains price discriminate on the race and income characteristics of an area? *J. Bus. Econom. Statist.* 15(4):391–401.

Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50(4):1029–1054.

Iyer G (1998) Coordinating channels under price and nonprice competition. *Marketing Sci.* 17(4):338–355.

Kalnins A (2003) Hamburger prices and spatial econometrics. *J. Econom. Management Strategy* 12(4):591–616.

Klein B, Murphy KM (1988) Vertical restraints as contract enforcement mechanisms. *J. Law Econom.* 31(2):265–297.

Lafontaine, F (1995) Pricing decisions in franchised chains: a look at the restaurant and fast-food industry. NBER Working Paper No. 5347, National Bureau of Economic Research, Cambridge, MA.

Lafontaine F (1999) Franchising vs. corporate ownership: The effect on price dispersion. *J. Bus. Venturing* 14(1):17–34.

Lafontaine F, Slade M (2008) Empirical assessment of exclusive contracts. Buccirossi P, ed. *Handbook of Antitrust Economics* (MIT Press, Cambridge, MA), 391–414.

Lal R (1990) Improving channel coordination through franchising. *Marketing Sci.* 9(4):299–318.

Mathewson F, Winter R (1998) The law and economics of resale price maintenance. *Rev. Indust. Organ.* 13(1–2):57–84.

Mathewson GF, Winter RA (1985) The economics of franchise contracts. *J. Law Econom.* 28(3):503–526.

Misra S, Nair HS (2011) A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quant. Marketing Econom.* 9(3):211–257.

Murry C (2017) Advertising in vertical relationships: An equilibrium model of the automobile industry. Working paper, Boston College, Boston.

Nevo A (2000) A practitioner's guide to estimation of random-coefficients logit models of demand. *J. Econom. Management Strategy* 9(4):513–548.

Nevo A (2001) Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2):307–342.

Perry MK, Porter RH (1990) Can resale price maintenance and franchise fees correct sub-optimal levels of retail service? *Internat. J. Indust. Organ.* 8(1):115–141.

Petrin A (2002) Quantifying the benefits of new products: The case of the minivan. *J. Political Econom.* 110(4):705–729.

Raju J, Zhang ZJ (2005) Channel coordination in the presence of a dominant retailer. *Marketing Sci.* 24(2):254–262.

Romano RE (1994) Double moral hazard and resale price maintenance. *RAND J. Econom.* 25(3):455–466.

Shaffer G (1991a) Capturing strategic rent: Full-line forcing, brand discounts, aggregate rebates, and maximum resale price maintenance. *J. Indust. Econom.* 39(5):557–575.

Shaffer G (1991b) Slotting allowances and resale price maintenance: A comparison of facilitating practices. *RAND J. Econom.* 22(1): 120–135.

Sudhir K, Datta S (2008) Pricing in marketing channels. Rao V, ed. *Handbook of Pricing* (Edward Elgar Publishers, Cheltenham, UK).

Telser LG (1960) Why should manufacturers want fair trade? *J. Law Econom.* 3(October):86–105.

Thomadsen R (2005) The effect of ownership structure on prices in geographically differentiated industries. *RAND J. Econom.* 36(4):908–929.

Tirole J (1988) Vertical control. *The Theory of Industrial Organization* (MIT Press, Cambridge, MA), 177–178.

Villas-Boas JM (2020) Repeated interaction in teams: Tenure and performance. *Management Sci.* 66(3):1496–1507.

Villas-Boas JM, Winer RS (1999) Endogeneity in brand choice models. *Management Sci.* 45(10):1324–1338.

Winter RA (1993) Vertical control and price vs. nonprice competition. *Quart. J. Econom.* 108(1):61–76.

Yang B, Chan T, Thomadsen R (2019) A salesforce-driven model of consumer choice. *Marketing Sci.* 38(5):871–887.