Inference of Service with Limited Data
– Apple and China Unicom’s iPhone Exclusivity

Grace Wei Yu
Boston University

December 2, 2014

Abstract

This paper provides a framework for evaluating retailers’ service even when service is not observed. The identification of service comes from the variation in manufacturer’s wholesale prices, in the sense that retailer’s optimization over service enters the manufacturer’s first-order condition and thus alters its price setting. Using new data containing wholesale prices from China Unicom, China’s second largest wireless carrier, I construct and estimate a dynamic structural demand model of consumers with persistent heterogeneous tastes. The demand estimates are then used in the structural supply model to estimate service parameters. I find that service has a significantly positive effect on expanding market demand; however, its impact subsides over time. I argue that this pattern in service provides a potential explanation for Apple’s initial exclusive contract with China Unicom and subsequent contract arrangements.

1 Introduction

This paper provides a framework for learning about the role of retailer service when service is not observed. “Service” in this paper relates to any type of non-contractible brand level promotion choice made by the retailer. For instance, a sales person taking time to explain the features of a complex product is an important example of service for this paper. Given limited access to service data, this paper draws its inference about service through the manufacturer’s first-order condition. Even if the retailer bears the full cost of service provision, nonetheless, the retailer’s optimization of service enters the manufacturer’s first-order condition. As a result, variation in wholesale prices provides a source of identification of the importance of service. Taking advantage of the observed wholesale prices in the data, I am able to numerically estimate the impact of service. I proceed with the estimation in two steps. First, I construct and estimate a dynamic structural demand model featuring heterogeneous consumer preferences. In the second step, given the estimates of the demand model, I compute the manufacturer’s price-cost margins, as a function of parameters governing service provision. Matching the predicted price-cost margin to the observed price-cost margin allows me to estimate how consumers respond to service. The empirical focus is on the service provided by China Unicom, the second-largest wireless carrier in China, which was under an exclusive contract with Apple from 2010 to 2012.
Learning about service is necessary for many reasons. First, a consumer’s purchase decision is subject to service. For example, a well-designed display center can change consumer’s perception of a product; pre-purchase demonstration helps consumers to better infer the quality and performance of the product; and post-purchase service can be viewed as a quality improvement. Consumer’s welfare is improving with service. As service increases, the probability of purchase and thus the firm’s revenue are also higher. Therefore, learning about service is of policy relevance for surplus calculations in the counterfactuals. Second, service affects the demand faced by both the retailer and the manufacturer. Service is often subjected to free-riding, and thus provides useful information for assessments of merger activities between manufacturer and retailer, as well as across retailers. Finally, as the empirical setting happens to be an exclusive contract, service estimations can help us better understand the underlying motivations for such vertical contracts.

Estimation of service has been quite challenging for economists. One reason is the heterogeneous preferences regarding service by infra-marginal consumers (Spence (1975)). Another reason is that service is difficult to measure. Limited data access furthers the difficulty. This paper addresses these issues by specifying a model in which a retailer picks the optimal level of service and manufacturers respond appropriately in picking wholesale price. After knowing the mechanism of service on wholesale prices, I need only focus on the variation of wholesale prices to infer service.

This paper’s contribution to the existing literature is four-fold. First, it brings new empirical evidence to the literature on the inference of service, in particular those that focus on service as a purpose for various vertical constraints. Restained by limited access to service data, most research is theoretical. This paper, to my knowledge, is the first attempt to empirically evaluate the impact of service. The results in turn bring insights and clarifications to policy-makers and inspire better judgments when it comes to cases of vertical constraints in the future. The second contribution is that I present a new method for learning about service from supply chain first-order conditions. This method is not limited to the analysis of service, but can be widely applied to the identification of any variable that is unobservable but nonetheless alters a firm’s first-order condition. In a closely related paper by Villas-Boas (2007), she uses supply chain first-order conditions to infer vertical relationships between manufacturer and retailer. The third contribution is the novelty of data. The data was obtained directly from the industry source – China Unicom – and provides the wholesale prices of smartphone models charged by the manufacturer to China Unicom, a rare opportunity in empirical studies. Besides its use for service estimation, the data also help to expand our understandings of the similarities and differences in wireless industries in China and the United States. The last contribution is that showing that service is important provides a potential explanation for why Apple signs exclusive contracts, which I will further discuss below.

As mentioned, a great challenge and also the main contribution of this paper is to empirically identify the effect of service. The identification strategy is inspired by Bresnahan (1989), who argues that the logical distinction between two theoretical models will yield different comparative statics given a change in an exogenous variable, and the distinction in the comparative statics serves as economics inference that guides us in seeking identification of the model from data. In my case, service enters the consumer’s utility function the same way as other demand shocks, causing a shift of demand. As service is unobserved, we
cannot uniquely pin down the cause when we see a demand shift. However, I can estimate service from the supply chain. After plugging in the retailer’s optimal choice of service from the retailer’s first-order condition, the manufacturer’s first-order condition specifies a relationship between wholesale price, marginal cost, elasticities and market share, where elasticities are a function of service parameters. There would be a different relationship between these variables if service plays no role. The data favor the former relationship, so I can estimate service by finding the parameters that rationalize this relationship.

Showing that service matters is important also because it provides a potential explanation for Apple’s pattern of exclusive contract. Under the concept of service, there are several implications. First, service increases demand. Service provides extra product information for consumers who are less familiar with the product; it can also change consumers’ perceptions of a product’s quality, for example, by having a well-designed product showing center. Either way, it increases the probability that a consumer will make a purchase. Second, because a rival retailer can steal market share from the retailer who provides service by simply lowering prices, it reduces the incentive for retailers to provide service. Finally, service being unobservable gives rise to vertical externality, such that the retailer will provide service that is less than the manufacturer’s ideal level if the retailer cannot share the profit increase from service provision.

Exclusive contract can be an efficient tool for enforcement and motivation to provide service in this setting. Under an exclusive contract, a retailer has the total market share to itself. Based on the first implication, a higher market share provides more incentive for retailers to provide service. Also, an exclusive contract eliminates the service free-riding problem as consumers can only purchase the product from the retailer that provides the service. Further, if service is no longer important after a product becomes well-known, this can potentially explain Apple’s consecutive contract arrangements: an exclusive contract for the initial move followed by agreements with more retailers when the iPhone had gained enough familiarity. A closely related paper by Sinkinson (2014) studies the same type of exclusive contract between Apple and AT&T, however, provides a different motivation for exclusive contracting. I argue that an exclusive contract could be motivated by the manufacturer’s need to induce the retailer to provide optimal service; he argues that an exclusive contract can generate higher prices. Therefore, he focuses more on Apple wanting an exclusive contract and choosing a carrier, while my research assumes Apple already has chosen a carrier and investigates how the service level will vary under the exclusive contract.

The empirical model in this paper is motivated by several theoretical papers that study the enforcement of service. The one that had the most influence is the paper by Perry & Porter (1990) in which they use a representative consumer model and test which combination of vertical constraints will generate the optimal level of service. A different paper by Mathewson & Winter (1984) addresses similar question using a spatial model. They argue that the less optimal service is generated by two types of externality: one is horizontal externality and the other is vertical externality, in which the retailer fails to account for the profit increase for the manufacturer from the service provision. Perry and Porter’s (1990) work is more fitting for my model in the sense that the effect of service is market share dependent such that the marginal contribution of service to sales is higher when market share is higher. My model, however, differs from Perry and Porter’s model; I do not tackle the issue of retailer entry and exit since I have only one retailer, while it plays a heavy
role in their paper. The one-retailer setting automatically captures the implication of exclusive territory, specifically, an increase in equilibrium prices and manufacturers’ profits by altering the perceived demand curve to a less elastic one.\(^1\) In this sense, my model is a combination of Perry & Porter (1990) and Rey & Stiglitz (1995).

The empirical model I use for demand estimation follows Gowrisankaran & Rysman (2012), in which consumers are forward-looking and their choice of purchases is affected by their various traits, including area of residence, educational achievement, age and other demographics. Dynamics play an essential role here, especially for the smartphone market. The durability and rapid evolution of the market creates substantial intertemporal substitutions, requiring consumers to think ahead when they make a purchase. Unlike Gowrisankaran & Rysman (2012), here I allow for only a single purchase, while they allow repurchase behavior. Repurchasing is a reality in that consumers will replace their old smartphones with new ones after several years of use, however, such behavior is not captured by the data because the data ends before or around the time when replacements take place. Another point worth noticing is the price that consumers pay, which is not a one-time payment but a set of consecutive payments throughout the contract. Thus, the price used in the estimation is the discounted sum of the bundle price and consecutive monthly access fees.

In addition to Sinkinson (2014), a number of recent papers (Villas-Boas (2007), Mortimer (2008), Lafontaine & Slade (2008), Zanarone (2009), Asker & Ljungqvist (2010), Bonnet & Dubois (2010) and Bonnet & Dubois (2013), Conlon & Mortimer (2014)) conduct empirical evaluations of vertical constraints. Lafontaine & Slade (2008) conduct a survey on recent developments in empirical assessments of vertical relationships, and divide the literature into groups based on the econometric approach, such as multivariate regression on a cross-section level, natural experiments on legal changes, firm’s value forecast in an event study and estimating a structural model. My paper falls into the last category. Also in the same category are papers by Villas-Boas (2007), Bonnet & Dubois (2010) and Bonnet & Dubois (2013), which derive analytical expressions of a firm’s first-order condition, and use those expressions to conduct comparisons between vertical relationship alternatives and select the one that fits the data the best. While these papers share similarities with mine in the mathematical derivations, they differ from my own in two important respects. First, mine accommodates the service factor, while theirs do not. Second, my focus is on detecting the impact of service on a firm’s price scheme, while theirs is on detecting which pricing scheme best fits the data. Mortimer (2008) and Conlon & Mortimer (2014) study the effect of vertical rebates on downstream retailer’s service, so is more relevant to my paper than others in this regard. However, the service in my paper focuses on types of services that are difficult to verify and are of a broader definition, whereas the service in their paper is specifically about the retailer’s effort in restocking goods.

The rest of the paper proceeds as follows: Section 2 presents and describes data, Section 3 constructs the empirical model for both demand and supply, Section 4 discusses the estimation method and inference, Section 5 presents the empirical results and Section 6 presents the conclusions.

\(^1\)This is the key argument in Rey & Stiglitz (1995) that by this an exclusive territory can be used for reducing interbrand competition
2 Industry Description and Data

2.1 Background of Chinese Wireless Industry

The iPhone entered the Chinese market in November 2009, exclusively sold through China Unicom at the time. China Unicom did not officially introduce the phone+network bundle until August 2010, so the iPhone was the only product that could be either purchased as a non-contracted phone or through a bundle contract during this time. The phone+network bundle was relatively new in China at the time of this data, and China Unicom was a pioneer in initiating bundle sales. The other carriers did not adopt the bundle sales model until March 2010. Although it was widely adopted by carriers, a great portion of the population chose to purchase a non-contracted phone separately from a phone retailer, and then subscribe only to network service through a carrier before and during the data period. At the end of the period covered by this data, 70% of the population was still choosing a non-contracted phone over a bundle from a wireless carrier.

Figure 1 and 2 show the market division among three carriers in both 2G and 3G networks over time. At the end of 2009, there were 726 million mobile phone users total, among which only 1.7% were 3G network users. China Mobile was in a dominant position in both 2G and 3G networks, taking up to 73% of market shares in 2G network and 45% in 3G network; China Telecom was the smallest carrier in 2G network with only 7.3% market shares; and China Unicom started as the smallest 3G carrier with 22% of total 3G market shares. By the end of May 2013, just over three years later, 3G network users had expanded to almost 400 million, representing 33% of mobile users. Market shares have become more balanced among carriers, with China Mobile, China Telecom and China Unicom each having 41%, 28% and 31% shares of the whole market. China Unicom has surpassed China Telecom and become the second largest 3G wireless carrier in China.

Figure 1: Market Shares of Three Carriers in 2009
2.2 Data

The dataset I use in this estimation consists of sales and demographic data. A “product” here is a “phone+network” bundle from China Unicom, where a consumer signs a contract with the carrier for a certain length of time and picks a data plan for network service in exchange for a discounted phone price. Carriers also offer the option of a non-contract phone, which is the same deal a consumer would receive from a phone retailer. In this case, the consumer is not subjected to a contract. Given the huge share of the population with a mobile phone, the market is defined as the consumers that prefer to purchase a phone from a wireless carrier rather than a regular phone retailer, and the outside option includes consumers who purchase a non-contract phone from a carrier (not necessary from China Unicom) and those who purchase a bundle from a rival carrier to China Unicom.

A novelty of this data is that wholesale prices are observable, which is a rare opportunity for empirical economists. Observing wholesale price is essential here. As I mentioned earlier in the introduction, I cannot estimate service without wholesale price. However, the data comes with a limitation: the sales-related information is available for only one carrier, China Unicom. Such a limitation makes it quite difficult to conduct analysis on carriers’ behaviors, nonetheless, the data is reasonably sufficient to study the behaviors of manufacturers, as Unicom carries models of up to 21 various brands.

The sales data contains market-level sales-related information from August 2010 to May 2013, collected directly from China Unicom. Under observation is the “phone+network” bundle, including general information such as quantity sold and bundle list price. Detailed information includes the characteristics and wholesale price of the smartphone that comes with the bundle. Also, for each bundle, the data specifies the contract length and a list of 11 various plan options, which together determine the level of monthly refund. Therefore, the total price a consumer has to pay is an upfront payment of the bundle list price plus the discounted sum of a series of consecutive payments, which is equal to the plan fee net the refund. The plan fee ranges from 66RMB to 866RMB per month.
Table 1: Bundle Summary Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Units</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bundles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List price</td>
<td>RMB</td>
<td>1,971</td>
<td>1,434</td>
<td>499</td>
<td>5,880</td>
</tr>
<tr>
<td>Monthly fee</td>
<td>RMB</td>
<td>33</td>
<td>17</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>Total price</td>
<td>RMB</td>
<td>2,646</td>
<td>1,406</td>
<td>1,181</td>
<td>6,795</td>
</tr>
<tr>
<td>Sales</td>
<td></td>
<td>46,720</td>
<td>64,498</td>
<td>158</td>
<td>609,000</td>
</tr>
<tr>
<td><strong>Phones</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail price</td>
<td>RMB</td>
<td>1,665</td>
<td>1,380.9</td>
<td>399</td>
<td>5,399</td>
</tr>
<tr>
<td>Wholesale price</td>
<td>RMB</td>
<td>1,500.8</td>
<td>1,207.5</td>
<td>330</td>
<td>5,280</td>
</tr>
<tr>
<td>Camera</td>
<td>Mps</td>
<td>397</td>
<td>190.8</td>
<td>130</td>
<td>1,300</td>
</tr>
<tr>
<td>Screen size</td>
<td>Inches</td>
<td>3.55</td>
<td>0.63</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Battery life</td>
<td>mAh</td>
<td>1,446.3</td>
<td>339.1</td>
<td>860</td>
<td>4,000</td>
</tr>
<tr>
<td>CPU</td>
<td>MHz</td>
<td>832</td>
<td>322.6</td>
<td>184</td>
<td>1,741</td>
</tr>
<tr>
<td>Dual-SIMs-dual-standby</td>
<td></td>
<td>0.42</td>
<td>0.49</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Imported</td>
<td></td>
<td>0.46</td>
<td>0.50</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Operating systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Android</td>
<td></td>
<td>0.67</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>iOS</td>
<td></td>
<td>0.09</td>
<td>0.29</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Symbian</td>
<td></td>
<td>0.06</td>
<td>0.24</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Windows</td>
<td></td>
<td>0.03</td>
<td>0.17</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>0.15</td>
<td>0.35</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: An observation is a model-service-month bundle. Market shares are defined as quantity sold divided by market size, with market size defined as the number of mobile phone users at the end of May 2013.

data plan offers better network service with more available usage on voice, text, data and other features. However, a more expensive data plan also amounts to a higher total price, given that the consumer will receive a higher monthly refund out of the bundle list price for that data plan. Unfortunately, other than the aggregate sales of each bundle, the company does not collect data on the sales specific to each data plan. Given the policy that one can only upgrade a data plan (not the other way around) during the contract course, and the fact that consumers do not have an accurate prior assessment of their data usage, I chose the introductory plan for each bundle as the representative data plan to construct the total price, which is either 66RMB or 96RMB depending on the bundle list price.

I further augmented the sales data with additional measures of phone characteristics, which are publicly available. These characteristics were chosen based on a survey report conducted by iiMedia in 2012. In the survey, they asked consumers to name the top ten characteristics that they care about in a phone. I include seven of them here: operating system, manufacturer brand, screen size, camera mega-pixel, battery life, CPU power and dual-SIMs-dual-standby. There were 119 bundles available over the period of 34 months, each with a unique smartphone model that belongs to one of the 21 manufacturer brands.² There are 119 bundles over the period of 34 months, each with a unique smartphone model

²Dual-SIMs-dual-standby is a phone features that allows two sims cards adopted for two different networks working at the same time, as if you had two phones embedded together.
that belongs to one of the 21 manufacturer brands. In terms of well-known operating systems, there were 89 models with Android, 5 models with iOS, 5 model with Symbian and 3 models with Windows. Summary statistics of several key variables are shown in Table 1. For more summary statistics on the manufacturer level, please refer to Table 7 in Appendix.

The first thing to notice is that the list prices vary greatly among bundles, with the most expensive more than ten times higher than the least expensive. Highly correlated with the list prices, the retail prices of smartphones display the same pattern of variation. Because consumers are not limited by data plan choices, the variation of the bundle list prices largely reflects the quality differences of various smartphone models. The monthly fee is the extra charge consumers face each month, which is the plan fee net the monthly refund if the net value is positive, and zero otherwise. It is worth remembering that the absolute value and variation might be underestimated, and as such downward biases are subjected to data limitations, since we do not know consumers’ choices of data plan at the beginning of the contract nor how consumers adjust their data plans later.

![Figure 3: Herfindahl index over time](image)

**Note:** The index is calculated using the bundle sales of manufacturer relative to the total bundle sales of China Unicom in a given period.

Sales also vary greatly within, as well as across, manufacturers. Apple holds the dominant position in bundle sales, followed by Lenovo, Coolpad, ZTE, Huawei and Samsung. The average shares of these six manufacturers together take up to 85% of the total bundle sales per period. This provides preliminary evidence that the market of smartphones is highly concentrated. Figure 3 displays the Herfindahl index over time, which reflects the evolution of market structure. The downward sloping line describes a shift of market structure from highly concentrated to moderately concentrated when further down the road, suggesting the market of smartphones has become more competitive, and smartphones from different manufacturers are more substitutable between each other than before. A closer
look on how such a shift occurs is provided by Figure 5.

If we see market shares from the operating system view, nearly 70% of the smartphones in market are equipped with the Android system. “Others” under “Operating systems” in Table 1 represents operation systems that are essentially Android-based but with some alterations. With these included in the broader definition of Android, Android occupies more than 80% of the smartphone market.

Figure 4: Product characteristics over time

Last, product characteristics are improving over time. Figure 4 shows the evolving patterns of these characteristics across models by month. The pixel count in the built-in camera improves substantially, the image quality in some of the high-end models can match a moderate compact digital camera. Screen size is larger, with a smaller margin around screen. Such improvements enhance the touch experience with more sensitive and accurate operation. While all characteristics improve over time, the most significant improvement
seems to be the dual-SIMs-dual-standby feature, with a little over 10% having space for an additional SIM card in the early periods, while it was almost a common feature by the end of the data. Such growth is not so much a technological advancement as a reflection of a shift in market share distribution among three wireless carriers. Even given that more and more consumers have switched from China Mobile to China Unicom for their 3G network service, China Mobile still possesses the largest share of mobile subscribers. Additionally, there are no free-minutes calls across networks, meaning an extra per minute charge will be incurred each time a phone call is made from the Unicom network to another network or the other way around, to a point where it is cheaper to have SIM cards for two different networks. The increasing share of phones that have such a feature indirectly suggests the
growing market share of China Unicom in 3G subscribers.

Figure 5 graphs the dynamics of sales and prices of the six large firms over time. I use 5% of the total market sales period as the threshold. Manufacturers with average total sales greater than this threshold are labeled as large firms, and the rest are labeled as fringe firms. Among the six large manufacturers, only iPhone and Samsung are foreign brands; the rest are all Chinese native brands. Prices are the monthly simple averages: the discounted sum of an upfront bundle price and a set of consecutive basic plan fees over the contract length. The one with the price weighted by sales generated similar results.\(^3\) Until the second half of 2011, iPhone was the leading force in bundle sales; however, it was matched by other manufacturers in the later periods. Particularly worth noticing, Coolpad and Lenovo became the new leading forces in bundle sales and continue to grow stronger over time. There is overall a converging trend on sales among the six manufacturers, and also a shift in the bundle sales market from foreign-dominated to native-dominated. Unlike bundle sales, prices present a more widely spread pattern and are consistent over time. Prices of bundles with a foreign manufacturer model are much higher than those of native manufacturers. In particular, iPhone bundles are priced two to three times higher than a non-iPhone bundle. Bundle prices of Samsung models rank as the second highest. Others with native brand models charge relatively uniform prices, all clustering around 2000RMB.

Figure 6: Wholesale price over time

Figure 6 shows the simple averages of manufacturer wholesale prices over time. The key strategy for identification of service, is that with service entering a retailer’s first-order condition and thus affecting manufacturer’s pricing behaviors, the variation in the wholesale prices provide the source of identification. As shown in Figure 6, manufacturers do vary wholesale prices from time to time. Furthermore, the variation pattern shares

---

\(^3\)See Figure 9 in Appendix for reference
similar pattern as the pattern with final prices. Again, weighting by sales does not alter the result much. Bundles with foreign manufacturers also have overall higher wholesale prices, while native-manufactured bundles stay constantly low.

The demographic dataset contains consumer information on living area, age group and education level. The data comes from the Sixth National Census Report, a nationwide census conducted in 2010, which is so far the most accurate source of statistics on population demographics in China. With the statistics in the report, I was able to construct the empirical joint distribution of demographics up to three dimensions. The demographic dataset is supplemented with the average demographics of smartphone purchasers, information also contained in the iiMedia 2012 survey report.

3 Empirical Model

In this section, I construct an empirical demand and supply model that later will be used to numerically estimate the magnitude of the service effect. In particular, the construction of the supply model provides us with a close look at how retailers’ optimization of service enters manufacturer’s first-order condition, a mapping from service to the wholesale price. I use a dynamic demand model for empirical estimation, assuming consumers are forward-looking. Purchasing a bundle is a dynamic decision for a consumer, not only because of the durability of the phone, but also because of the switching cost if a consumer switches to a different carrier, as well as a possible early termination fee if a consumer ends the contract before the legitimate termination date. A dynamic model is more appropriate in a situation that has to do with durable products and inter-temporal decision making. The model also allows consumers to have heterogeneous preferences over product characteristics. In particular, I allow consumers’ individual traits to interact with product characteristics, affecting the consumer’s preferences differently when a consumer carries different traits.

In addition, I assume consumers have no intention to terminate the contract early when purchasing and indeed stick with the contract throughout. This assumption is based on a discussion with an industry source suggesting that there are few early terminations overall. Phones being damaged or stolen does not necessarily lead to early termination, as a consumer can continue the contract by reporting damage and asking China Unicom to reissue a card associated with the bundle, however, the consumer needs to purchase a no-contract phone as an replacement. Therefore, most early terminations are due to other unexpected events. Due to the rarity of these situations, I assume consumers do not account for involuntary early terminations at the time of purchase. The lack of early termination and the length of the data collection period when combined give rise to the last assumption of the demand model, the single purchase assumption. A typical contract lasts for two years and my data length is close to three years; the combination of the two yield that it is most likely consumers will make at most one purchase within data length.

I assume that retailers pick one level of service for each manufacturer at a given time, such that products within the same manufacturer share the same amount of service. The

\footnote{See Figure 10 in Appendix for reference.}
level of service varies from time to time. Given that China Unicom was in an exclusive contract with Apple, I would expect there are more service efforts for iPhones than other products.

### 3.1 Demand Model

The construction of consumer demand follows a standard model of discrete choice (McFadden (1984); Berry (1994); Berry, Levinsohn & Pakes (1995); Nevo (2000b)), with modifications to accommodate a dynamic setting, as in the model by Gowrisankaran and Rysman (2012). As defined in the data section, a product is a “phone+network” bundle from China Unicom. Total market size consists of the consumers who choose to purchase a phone from a carrier rather than a phone retailer, either as a non-contract phone or as a bundle. The outside option is either a non-contract phone from any of three carriers or a bundle from a rival carrier. In each month \( t \), a representative consumer \( i \) who is currently without a contract has two decisions to make: to purchase a product or not and, if so, choosing which to purchase from the product set available at month \( t \), \( J_t \). Any choice other than a product from \( J_t \) is considered as an outside option. Conditional on purchasing a product, the gross flow utility that consumer \( i \) gets from product \( j \) in month \( t \) is given by

\[
\delta_{ijt} = x_{jt} \beta_i + \xi_{jt} + m + t + iPhone_{jt} + \lambda_t z_{mt} + \alpha_i f_{jt}
\]

where \( x_{jt} \) are the observed product characteristics, \( \xi_{jt} \) captures the product characteristics that are observed by the consumer but not by the econometrician, \( m \) represents the fixed effect of the phone manufacturer (a time invariant product characteristic), \( t \) identifies the unobserved time specific determinants of demand, \( z_{mt} \) represents carrier promotion service that is not observed by econometrician, with the subscript \( m \) representing manufacturer \( m \), and the subscript \( t \) means service is time dependent, \( iPhone_{jt} \) captures the effect of all other product promotions that is specific to iPhone but are not conducted by the retailer. The effect of retailer service is captured by \( \lambda_t \) and the impact varies over time. The last term \( f_{jt} \) is the monthly access fee for the network service provided by the carrier.

Theoretically, \( \lambda_t \) picks up the impact of service on market shares only in terms of carrier’s effort, whereas the retailer’s promotion effort for other manufacturers and Apple’s promotion effort is either captured by \( t \) or \( iPhone_{jt} \). Empirically, however, because service is not observed in data, it is impossible to separately identify \( \lambda_t \) from the other variables in the demand estimation. Specifically, retailer’s general service for all products will be captured in the time fixed effect and retailer’s special service for iPhone will be captured in the iPhone-time effect. Although not separated identified from the demand model, the service parameter meets the firm’s first-order condition, which we can draw inference from. Details about service will be discussed further in the supply model section.

The random coefficients \( \alpha_i \) denote the consumer’s marginal disutility of price, and \( \beta_i \) denote the unknown consumer taste parameters for different product characteristics other than price – parameters that vary by consumer but are constant over time. I assume that
the durability of the smartphone is one year longer than the length of its contract, and that then the consumer can choose either to keep the phone for another period or get rid of it with no cost. Utility from product characteristics other than monthly access fee, called the net flow utility, is represented by $\delta^n_{ijt}$, where $n$ denotes net value. Then the latent indirect utility of product $jt$ for consumer $i$ is the sum of the discounted utility of the product, whose expression is given by,

$$u_{ijt} = \sum_{\tau=1}^{T} r^{\tau-1} \delta^n_{ijt} + \alpha_i p_{jt} + \epsilon_{ijt}, \quad j = 1, \ldots, J_t,$$

where $p_{jt} = \sum_{\tau=1}^{T_c} r^{\tau-1} f_{jt} + p^b_{jt}$ denotes the overall price that a consumer needs to pay for bundle $j$, which is the discounted monthly access fee plus an one-time payment at the beginning of the contract. $T$ is the period length that consumer will hold the phone, $\epsilon_{ijt}$ is the i.i.d logit error distributed extreme value type I, capturing consumer’s idiosyncratic purchase experience.\footnote{In the paper of Sinkinson (2014), he omitted the i.i.d. assumption on logit draws and opts instead for only the random coefficients to rationalized consumer’s choice, as he argued that idiosyncratic shocks are dependent on consumer’s state holdings and thus is not random. My model, however, does not have such concerns: first is that the data starts at the beginning bundle sales, where no consumers have contracts; second is that I assume no repurchasing during the course of contract. Another reason has to do with the dynamics of the model, which make the i.i.d. assumption extra attractive, is because it gives a nice analytical form of market shares.}

Parameters $\alpha_i$ captures consumer’s disutility of price. The taste parameters $(\alpha_i, \beta_i)$ vary across consumers according to the following formula:

$$\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \bar{\alpha} \\ \bar{\beta} \end{bmatrix} + \Phi D_i + \Sigma \nu_i,$$

where $(\bar{\alpha}, \bar{\beta})$ are the mean of consumer tastes, the parameter matrix $\Phi$ captures the observed heterogeneity in consumers tastes, the deviation from mean tastes in population due to consumers’ various observed characteristics, while parameter matrix $\Sigma$ captures the unobserved taste heterogeneity due to unobserved consumer characteristics. Accordingly, $D_i$ contains the observed consumer characteristics such as residence area, educational achievement and age, and $\nu_i$ captures all other consumer characteristics that are not included in $D$. In the econometric model, $D_i$ has an empirical distribution $\hat{F}(D)$ from the demographic data, $\nu_i$ follows a normal distribution with mean zero and variance-covariance matrix $\Sigma$ to be estimated. For notational convenience, from onwards, the mean taste parameters other than that for price will be referred as linear parameters – denoted by $\theta_1 = \{\bar{\beta}\}$, the parameter of price and the those associated with consumers’ heterogeneous tastes will be referred to as the non-linear parameters – denoted by $\theta_2 = \{\bar{\alpha}, \Phi, \Sigma\}$. Such division matters in the sense that $\theta_1$ are the coefficients of the product-specific component, which does not vary with consumer characteristics, while as $\theta_2$ contains coefficients of the interactions between product-specific and consumer characteristics. Also computationally, given each guess of $\theta_2$, we first solve for $\delta^n(\theta_1)$ through nonlinear system of matching predicted market shares to observed market shares before we estimate $\theta_1$ linearly. In this sense, the mean price
parameter $\alpha$ is included in $\theta_2$ although it is the parameter of a consumer invariant product component.

Note that there is no continuation value in Equation (2), which follows the assumption that consumers believe they drop out of the market after purchase. This assumption fits reality; for example, before making a purchase, a consumer would know whether Apple will release a new model next month and adjust their purchasing decision accordingly. It may be the case that it is worth waiting one more month versus purchasing at the moment. However, once a purchase has been made, the consumer drops from the market and no longer cares how market evolves afterwards. With this said, the continuation value drops out of Equation (2).

If consumer $i$ instead chooses the outside option, he will get a flow utility of the outside option and a continuation value. As I cannot separately identify the value between each outside option, I normalize the value of flow utility to zero. Let $\Omega_t$ denote all the current product characteristics and any other factor that affect the future product characteristics. Let $V_i(u_{ijt}, \Omega_t)$ denote the value function at time $t$, then the indirect utility from choosing the outside option is written as

$$u_{i0t} = rE[EV_i(u_{ijt}, \Omega_{t+1})|\Omega_t],$$  \hspace{1cm} (4)$$

where $r$ is the monthly discounted rate equal 0.99, following existing literatures. $EV_i(u_{ijt}, \Omega_t)$ denotes the expectation of the value function such that $EV_i(u_{ijt}, \Omega_t) = \int \epsilon_{it} V_i(\epsilon_{it}, u_{ijt}, \Omega_t)dP_{\epsilon}$, where $\epsilon_{it} = \{\epsilon_{i0t}, \ldots, \epsilon_{iJ_t}\}$. $E[EV_i(u_{ijt}, \Omega_{t+1})|\Omega_t]$ is the continuation value at time $t$, an expectation of value of purchase at $t+1$ conditional on current value. Gowrisankaran & Rysman (2012) provides a method to construct this value using the logit inclusive value, which is the maximum discounted utility one can get from the current product set, and assume it is sufficient for consumers to predict the future values of the maximum discounted utility. Following Gowrisankaran & Rysman (2012), define logit inclusive value for consumer $i$ at month $t$ as

$$u_{it} = \log \left( \sum_{j \in J_t} \exp(u_{ijt}) \right),$$ \hspace{1cm} (5)$$

we can use logit inclusive value to replace the $\Omega_t$ and rewrite the continuation value function as $E[V_i(u_{i,t+1})|u_{it}]$, where

$$V_i(u_{i,t+1}) = \log \left( \exp(u_{i0t}) + \exp(u_{it}) \right).$$ \hspace{1cm} (6)$$

Now in order to solve the continuation value function, we only need to specify the probabilities of various states of $u_{i,t+1}$ that $u_{it}$ will transition into next period. Here I assume that $u_{it}$ follows an autoregressive model of first-order,

$$u_{i,t+1} = \varphi_{1i} + \varphi_{2i} u_{it} + \epsilon_{it},$$ \hspace{1cm} (7)$$

where $\epsilon_{it}$ is normally distributed with mean zero and $\varphi_{1i}$ and $\varphi_{2i}$ are parameters to be estimated.
Finally, given the previous set-up, the unconditional purchase probability for consumer $i$ to purchase product $j$ in time $t$ is given by

$$s_{ijt} = \frac{\exp(u_{it})}{\exp(u_{it}) + \exp(u_{it}) \times \exp(u_{ijt})},$$

and the market share of product $j$ in month $t$ is

$$s_{jt} = \int d \nu \int s_{ijt} dF(\nu)dF(D).$$

### 3.2 Supply Model

As mentioned earlier, we cannot directly estimate how consumers respond to service as service is not observed; however, we can infer service through supply chain first-order conditions. The goal of this section is to build one-to-one mapping between service and wholesale price. I will start by deriving a retailer’s optimizing conditions over final price and service and then use them to solve for manufacturer’s first-order condition over wholesale price. By the end of this section, I will be able to write down the wholesale price-cost margin as a function of service parameter.

Specifically, the way I compute price-cost margins follows the method conducted in Berry et al. (1995) – a static supply model that backs out price-cost margins by applying the firm’s first-order condition of price with demand estimates. One would naturally expect a dynamic supply model to follow a dynamic demand model. The essential idea of a dynamic supply model is to change the relevant marginal revenue from the current revenue to the present discounted value of the infinite stream of future revenue. An empirical implementation was conducted in a separate paper with Gowrisankaran and Rysman (Gowrisankaran, Rysman & Yu (2014)), where we show how to compute price-cost margins non-parametrically under dynamic settings. However, this method inevitably brings a substantial computational burden, as a price change for a given product leads to not only the need to recompute market shares for all the products in the current periods, but also for all the products in all the following periods.

To reduce the computational burden without compromising the validity of results, I instead focus on the main factor that determines the price other than product quality, which is the population shares of holdings. A price drop today increases holdings in the current period, thus eats away future demand, forcing the firm to further lower the price in the future. The magnitude of the impact on future price depends on how much market share the firm has. A firm with small market share is much less affected than a large firm as the reduction in future demand will be dissolved by rivals.\(^6\) As shown in the data section, the average market share of a firm is around 5%, which means the price-cost margins computed using the static model are not significantly different from those computed using dynamic.

\(^6\)In Gowrisankaran et al. (2014), we empirically confirmed this argument, showing that a firm (Sony) with more than 50% market share is heavily affected by pricing dynamics; conversely, firms with smaller market shares (around 20% or less, which included Canon, JVC and Panasonic) are affected very little, ignoring the dynamic consideration.
I will mainly focus on deriving the expressions for a linear pricing model with service element (referred as a “double-margin service model”). Theories have suggested that when it comes to vertical relationships such as exclusive territory, it is usually optimal to employ a nonlinear pricing scheme, in which a manufacturer charges a fixed fee and commits to a certain wholesale price. Although initially optimal, it is not posteriorly optimal; after the manufacturer has broken the commitment once, the best strategy is then to keep revising price every period. The data shows all manufacturers changed wholesale price multiple times, which favors the use of the linear pricing model.

**Simple linear pricing model.** In this scenario, the manufacturer sets the wholesale prices first and the carrier follows with retail prices and a certain level of service. Each party tries to maximize its own profit. Such a setting inevitably leads to two margins: a wholesale margin over manufacture cost and a retail margin over wholesale price and other carrier related costs, the so-called double-margin. To solve for the expression, I start with the carrier’s profit function. Based on the data structure, I observe one carrier that carries products from various manufacturers. Thus, I assume there are $M$ manufacturers each with a product set $J_m$, $m = 1, \ldots, M$, and one carrier whose product set is the size of all manufacturers’ product sets together, $J = \{J_m, m = 1, \ldots, M\}$. I index carrier by subscript $c$, carrier $c$’s profit function in month $t$ is,

$$
\pi_{ct} = \sum_{m=1}^{M} \sum_{j \in J_m} (p_{jt} - w_{jt} - \gamma^s z_{mt}) s_{jt}(p, z),
$$

(10)

where $p_{jt}$ denotes final price, $w_{jt}$ is the associated wholesale price paid to manufacturer by retailer, $z_{mt}$ represents the promotion service for manufacturer $m$ and $\gamma^s$ is the service cost parameter. Demand of bundle $s_{jt}$ is a function of both price and service. Also, I assume away carrier’s marginal cost other than wholesale price and those related to service because of two reasons. First is of the limited data source; second is the cost associated with bundle sales is most likely incurred by carriers on a store level rather than on product level. An additive fixed fee to the profit function would not alter carrier’s first-order condition.

The carrier has two variables to optimize: retail prices and promotion services for manufacturers. The first-order condition with respect to the price product $j$ is given by,

$$
s_{jt} + \sum_{m=1}^{M} \sum_{q \in J_m} (p_{qt} - w_{qt} - \gamma^s z_{mt}) \frac{\partial s_{qt}}{\partial p_{jt}} = 0.
$$

(11)

Let $\Delta_{ct}$ be the matrix that contains the first-order derivatives of market shares with respect to retail prices for the products that sold by carrier $c$ in time $t$, with entry $(j, q)$ entry equal $\frac{\partial s_{qt}}{\partial p_{jt}}$ and stack all the first-order conditions of the product in $J_t$, we obtain the carrier’s first-order condition with respect to price, written in vector form as the following,

\footnote{Here I omit the ownership matrix of carrier, $I_{ct}$ where entry $(i, j) = 1$ if both product $i$ and $j$ are sold by the same retailer and 0 otherwise. Since I only have data on China Unicom, all elements in $I_{ct}$ are ones. Note that this would not be the same case when it comes to manufacturer’s ownership matrix, as the carrier sells products from various manufacturers at the same time.}
\[ p_t - w_t - \gamma^s z_t = -\Delta_{ct}^{-1} s_t(p, z). \]  

(12)

Similarly, we can write down carrier’s first-order condition with respect to service \( z_{mt} \) as,

\[
\sum_{l=1}^{M} \sum_{q \in J_l} \left( p_{qt} - w_{qt} - \gamma^s z_{lt} \right) \frac{\partial s_{qt}}{\partial z_{mt}} - \gamma^s \sum_{q \in J_{mt}} s_{qt} = 0,
\]

(13)

and the vector form of the service first-order condition is given as, \( p_t - w_t - \gamma^s z_t = \Delta_{sz} \gamma^s \sum_{s \in J_{mt}} s_{mt}(p, z) \), where \( \Delta_{sz} \) is the market share response matrix to service.

Each manufacturer \( m \) in month \( t \) chooses wholesale price \( w \) to optimize its profit function. This is conditional on carrier’s reaction according to Equation (12) and (13). The manufacturer \( m \)’s profit function is given by,

\[
\pi_{mt} = \sum_{j \in J_{mt}} (w_{jt} - c_{jt}) s_{jt} (\hat{p}(w), \hat{z}(w)),
\]

(14)

where \( c \) is the manufacturer marginal cost and \( (\cdot) \) denotes carrier’s best response at the equilibrium given wholesale prices. As carrier bears full responsibility in providing service, manufacturer is free of service cost. Assuming a pure-strategy Nash equilibrium in the wholesale prices, manufacturer \( m \)’s first-order conditions are the following for a given wholesale price \( w_{jt} \),

\[
s_{jt} + \sum_{k \in J_t} \sum_{q \in J_{mt}} \left( w_{qt} - c_{qt} \right) \left( \frac{\partial s_{qt}}{\partial p_{kt}} \frac{\partial p_{kt}}{\partial w_{jt}} + \frac{\partial s_{qt}}{\partial z_{kt}} \frac{\partial z_{kt}}{\partial w_{jt}} \right) = 0, \quad \forall j \in J_{mt}.
\]

(15)

Because the matrix in the second parenthesis becomes very complicated when there are multiple products and manufacturers. To obtain this, it is easier to break down and compute piece by piece. We already have matrices of market share response to final price and service, which are \( \Delta_c \) and \( \Delta_{sz} \) respectively. There are two matrices left in parentheses yet to be defined. The One with the element \((j, q)\) equal \( \frac{\partial p_{jt}}{\partial w_{qt}} \), denoted as \( \Delta_p \), is the matrix of the derivatives of all the retail prices with respect to all the wholesale prices. The other one is the service response matrix with respect to wholesale prices, \( \Delta_z \), with element \((j, q)\) as \( \frac{\partial z_{jt}}{\partial w_{qt}} \). Note that other than matrix \( \Delta_p \), all three matrices have components of the first-order derivatives with respect to service, which make them functions of \( \lambda \). In particular, matrix \( \Delta_z \) is also a function of \( \gamma^s \). Now I proceed to the construction of these matrices.

Let us start with matrix \( \Delta_p \). Dropping time subscript to simplify notation, the total differentiation of carrier’s first-order condition on price (Equation (11)) for a given \( j \) with respect to all retail prices \( (p_k, \forall k \in J) \), all the services \( (z_l, l = 1, \ldots, M) \) and a wholesale price \( w_{f} \) is given by,
Let $G$ be the matrix with general element $g(j,k)$, $K$ denote the matrix with element $\kappa(j,l)$, where $\kappa(j,l)$ is a column vector of the length of product counts of manufacturer $l$, and $H$ denote the matrix with element $h(j,f)$ and its $f$th column $H_f$. Putting together all the products $j \in J$, we have $G \frac{dp}{dw_f} + K \frac{dz}{dw_f} = H_f$. I will derive the expression of matrix $\Delta_z$ shortly, let us assume for now that $\Delta_z$ is known, which $f$th column is $\frac{dz}{dw_f}$. After some arrangements, the $f$th column of matrix $\Delta_p$ is obtained as,

$$
\frac{dp}{dw_f} = G^{-1} \left(H_f - K \frac{dz}{dw_f}\right).
$$

Stacking all columns together, $\Delta_p = G^{-1} (H - K \Delta_z)$. The derivation of matrix $\Delta_z$ uses the same approach as $\Delta_p$, by totally differentiating carrier’s first-order condition on service (Equation (13)), only this time with respect to all services ($z_k, k = 1, \ldots, M$) and a wholesale price $w_f$. Again omitting the time subscription $t$, the expression is given as

$$
\sum_{k=1}^M \sum_{l=1}^M \sum_{q \in J_l} (p_q - w_q - \gamma^sz_l) \frac{\partial^2 s_q}{\partial z_m \partial z_k} - \gamma^s \sum_{q \in J_k} \frac{\partial s_q}{\partial z_m} - \gamma^s \sum_{q \in J_m} \frac{\partial s_q}{\partial z_k} \frac{dz_k}{dw_f} - \frac{\partial s_f}{\partial z_m} = 0.
$$

I use $S$ and $B$ to denote matrices with element $s(m,l)$ and $b(m,f)$ respectively. Element $s(m,k)$ is a block matrix with row length equal to the number of products of $m$ and column length equal the number of products of $k$; element $b(m,f)$ is a row vector of the length equal the number of products of manufacturer $m$. Putting all products together and completing some simple algebra yields

$$
\Delta_z = S^{-1} B.
$$

The expressions for the other two matrices $\Delta_c$ and $\Delta_{sz}$ are derived in Appendix B. With all the pieces gathered, using $\Delta_w$ to denote the market share response matrix to wholesale price with the form of $\Delta_w(\lambda, \gamma^s) = \Delta'_p \Delta_c + \Delta'_z \Delta_{sz}$, the wholesale price-cost margin written in vector form is derived as,

$$
\begin{align*}
wt - ct &= - (I_w \cdot \Delta_w(\lambda_t, \gamma^s))^{-1} st(p, z),
\end{align*}
$$

(20)
where $I_w$ is manufacturer’s ownership matrix with element $(i, j) = 1$ if both products $i$ and $j$ are produced by the same manufacturer. Equation (20) provides the source of identification of the service parameters. The identification of service is similar to the way that BLP identifies wholesale price. In Berry et al. (1995), they assume there is no role for service, and identify the wholesale price by using firm’s first-order condition of marginal revenue equal marginal cost, where marginal revenue is a function of price, market share, demand elasticity and wholesale price. Price and market share are directly observed in data, and the demand estimation provides the demand elasticities. In my case, I observe the wholesale price but allow for a more flexible setting of service, so service is the parameter that enters the demand elasticities and cannot be picked up by the demand estimation. In other words, there is a relationship between price, market share and markup that implies service parameter to be zero, which is the BLP relationship. However, I observe a different relationship in the data, and I try to find the service parameter that rationalizes such relationship.

Note that this model has embedded in the simple pricing model without service, a case that coincides with the scenario of simple linear pricing model in Villas – Boas (2007). To obtain the expressions for this model, one simply drops out the second term in $\Delta w$. The construction of $\Delta p$ becomes simpler too, as the second part of Equation (16) also drops out, the matrix becomes $\Delta p = G^{-1}H$, with general element $(j, q) = \frac{\partial p_j}{\partial w_q}$.

4 Estimation and Identification

4.1 Demand model estimation

The empirical estimation proceeds in two steps, given in backward order. Starting from the last stage, I estimate the demand model of the “phone+network” bundle using sales and demographic data. The estimation strategy uses the generalized method of moments (GMM) taken by Berry et al. (1995) and Petrin (2002), with my moments consisting of three sets. First two sets are the Berry et al. moments – the set of moments matches the model’s share predictions to those observed in the data and the set of moments related to the unobserved demand disturbances $\xi_j(\theta)$. Recalling that $\theta_2$ denotes the set of demand parameters that enter the mean utility function nonlinearly, the first set of moments are

$$s_j \left( \delta_j(\hat{\theta}_2), \hat{\theta}_2 \right) - s_j = 0, \quad j = 0, 1, \ldots, J.$$  \hspace{1cm} (21)

Solving this requires a fix point computation in which we start with a guess of the monthly flow utility $\hat{\delta}$ and cycle it through Equation (1) and (8) until a fixed point is found. To guarantee the identification of the model, one needs to prove that such point is unique. However, it cannot be proven in a dynamic setting unless we can show that all current and future products are substitutes, a condition shown in Berry et al. (1995) that is necessary for the uniqueness. I’ve tested with different initial values of $\hat{\delta}$ and it always converges to the same value and hence assume that for any vector of parameter $\theta_2$, there is a unique vector $\hat{\delta}$ such that $\hat{\delta} = \hat{\delta}(\theta_2, \theta_2)$.
The second set of moments are based on the moment condition that
\[ E \left[ \xi_j(\theta) \Big| Z^d_j \right] = 0, \tag{22} \]

where \( Z^d_j \) is a vector of exogenous variables, performing as instruments for demand model. The choice of the set of instruments is discussed in the subsection “Identifications and instruments”.

The last set of moments are the micro moments, aiming to provide more accurate demand estimates when consumer-level data is not observable. According to Petrin (2002) and Berry, Levinsohn & Pakes (2004), information that relates demographic averages of consumers to the products they purchase plays the same role as consumer-level data. Supplementing market-level data with such information helps to precisely identify the relevant substitution patterns when price and product characteristics are not sufficient to do so. The China Statistical Yearbook 2010 (CSY 2010) contains information to construct a joint distribution of population by living area, educational achievement and age group, and the IIMedia China Smart Phone Market Annual Research Report 2012 provides information on corresponding demographic averages of purchasers of new smart-phones. Combining these pieces of information, I can construct the third set of moments – micro-moments, matching model predictions to the following averages:

\[ E \left[ \bar{d} - E \left[ d_i \Big| \{i \text{ purchases new model}\} \right] \right] = 0, \tag{23} \]

where \( d \in \{\text{area, education, age}\} \) and \( \bar{d} \) is the according averages conditional on purchasing observed in data. To be more specific, the first micro-moment matches to the average probability of living in different areas (city, outside the city) conditional on new model purchase; the second matches the the share of population with different education levels (junior, senior high and equivalent, bachelor) conditional on new model purchase; and the last matches to the probability of being in different age groups (18-34, 35-54 and >54) conditional on new model purchase. More details on the construction of micro-moments are discussed in Appendix A.

Stacking all sets of moments together and using \( G_1(\theta) \) to denote the one matching market shares, \( G_2(\theta) \) to denote the mean independent conditions and \( G_3(\theta) \) to denote the one matching consumer demographics, we have the population moments

\[ E \left[ G(\theta_0) \right] = E \begin{bmatrix} G_1(\theta_0) \\ G_2(\theta_0) \\ G_3(\theta_0) \end{bmatrix} = 0. \tag{24} \]

Hansen (1982) shows that the optimal two-step GMM estimator takes the following form:

---

8The 2010 Yearbook contains the results from the sixth nationwide population census. Its statistics are the most comprehensive and accurate among all recent yearbooks, and it is the only yearbook that has information on joint distribution of the three demographic variables.

9With in my data length, the variation in demographics are negligible, I use the statistics from the CSY 2010 and IIMedia Report 2012 to approximate the corresponding demographic averages within the data length.
\[ \theta^* = \arg \min_{\theta \in \Theta} \hat{G}(\theta)'W\hat{G}(\theta), \] (25)

where \( \hat{G}(\cdot) \) is the sample analogue of \( G(\cdot) \); \( W \) is the weight matrix. In the first step, use \( W = (Z^dZ^d)'^{-1} \) to obtain the preliminary consistent estimate of the true value, \( \tilde{\theta} \), and at the second step substitute \( W \) with the optimal weighting matrix, \( \tilde{S}^{-1} = E \left[ GG' \bigg| \tilde{\theta} \right] \), the inverse of the asymptotic variance-covariance matrix of moments evaluated at \( \tilde{\theta} \). Let \( \Gamma(\tilde{\theta}) = E \left[ \frac{\partial G}{\partial \theta} \bigg| \tilde{\theta} \right] \), denote the first-order derivation of moments with respect to parameters evaluated at \( \tilde{\theta} \), and the asymptotic variance of the two-step GMM estimator with optimal weighting matrix is,

\[ V(\theta^*) = N^{-1} \left( \Gamma'\tilde{S}\Gamma \right)^{-1} \] (26)

### 4.2 Service estimation

I start by writing down the cost functions. For each guess of service parameters, marginal cost can be computed as a function of \( (\lambda, \gamma^s) \) using the method described in the empirical model supply section. Since I cannot separately identify \( \lambda_t \) and \( \gamma^s \), I normalized \( \gamma^s = 1 \) and \( \lambda_t \) captures the benefit of service on demand relative to service cost. Similar to product characteristics in the demand model, cost characteristics are decomposed into one subset which is observed by econometricians and an unobserved component. For a given product \( j \) in time \( t \), using \( x^c_j \) to denote the cost vector and \( \omega_j \) to denote the unobserved cost component, the marginal cost \( c_j \), is written as

\[ c_j(\lambda_t) = x^c_j\gamma + \omega_j + \varepsilon_j, \] (27)

where \( \gamma \) is the vector that contains the cost parameters other than service. I include all the product characteristics \( x \) are in \( x^c \), and expect \( \omega \) to be correlated with \( \xi \). This is because whichever unobserved product characteristics improve the product quality might also be costly to produce. Given Equation (20), substituting the expression for marginal cost and we have the following,

\[ w_t + (I_w \cdot \Delta_w(\lambda_t))^{-1}s_t(p, z) = x^c_t\gamma + \omega_t + \varepsilon_t. \] (28)

Note that the price-cost margins are a function of demand model parameters and both retail and wholesale prices. According to Equation (20), wholesale price \( w \) is a function of \( \omega \), together with the correlation between \( \xi \) and \( \omega \), I cannot assume the variables on the left-hand side of Equation (28) are uncorrelated with \( \omega \). Just as in estimating the demand model, I can construct supply model moment conditions by assuming orthogonality between \( \omega \) and appropriate instruments. Letting \( Z^s \) denote the instrument set for supply model, we have the moment condition for supply model written as,

\[ E \left[ \omega_j(\lambda_t)|Z^s_j \right] = 0. \] (29)
The remaining estimation is the same as the demand model; the only difference is that the population moments are consisted solely by the independent moment conditions.

Last but not least, to capture the time variant feature of the retailer service, I assume $\lambda_t$ is a linear function of time with the form given as

$$\lambda_t = \psi_0 + \psi_1 t,$$

(30)

where $\psi_1$ captures the time variant effect. Therefore, the service estimation is in search for the optimal parameters $\psi^* = \{\psi_0^*, \psi_1^*\}$ that minimize the GMM objective function of the supply model.

### 4.3 Identifications and instruments

There are two sets of demand parameters: those that enter the demand model nonlinearly $\theta_2 = \{\alpha, \Phi, \Sigma\}$, and those that enter linearly $\theta_1 = \{\beta^k, k \in K\}$. Micro-data allows me to identify the parameters in $\theta_2$, even without assumptions on the distribution of $\xi$. Given a set of parameter values $\hat{\theta}_2$, consumers with different demographic attributes (both observable and unobservable) interacting with product characteristics will yield different product choices and thus different average attributes conditional on product purchase. By requiring the model parameters to generate predictions of average attributes matching those observed in data, I obtain the source of identification of $\theta_2$. The interaction between consumer attributes and product characteristics also determines substitution patterns. As the variance of consumer attributes $\{\Phi, \Sigma\}$ for the $k$ product characteristic increases, products with this characteristic become better substitutes, and there is disproportionally more share-shifting toward these products when the current product is no longer available. The dynamic model provides substitution patterns across periods in addition to within periods. These together would provide more accurate estimation for demographic coefficients. The identification of the remaining parameters $\theta_1$, the mean preferences of product characteristics other than price, requires the market-level data. They are identified by the change of product $j$’s market share associated with change in the product characteristic.

The service parameter is identified by the change in manufacturer’s wholesale prices associated with a change in retailer’s optimization over service. The time trend coefficient of service will be further identified by the variations of wholesale prices within the manufacturer across periods.

As is standard in existing literature, retail price is endogenously correlated with unobserved demand disturbance $\xi$. The other observed products characteristics are exogenous, therefore, observed product characteristics other than price are the first to be included as instruments. Inspired by Berry et al. (1995), I construct the remaining set of instruments for the demand model by first exploiting factors that affect a firm’s market power. I include average characteristics of the products that produced by the same firm and those produced by rival firms, the counts of own and rivals’ products; as each month is a separate market, these attributes are at monthly base. I also includes dummies for the six large manufacturers.

---

10 The six manufacturers, ranked by market share from high to low as: Apple, Lenovo, Coolpad, ZTE, Huawei and Samsung.
These variables would either raise or lower a firm’s market power. For example, market averages tend to lower a firm’s market power and the firm’s indicator tends to raise its market power. Other than the variables that affect a firm’s market power, cost shifters are also part of the instruments set. For this, I include the wholesale price for each model and the averages for products that produced by the same firm and those produced by rival firms. All of these variables are correlated with price and mean independent of the demand disturbance, thus are appropriate instruments for a demand model.

The choice of instruments for the supply disturbance $\omega$ is as that for demand except that I drop wholesale price and the related brand and market averages because wholesale price is endogenous to $\omega$.

5 Empirical Results

5.1 Demand estimation

The parameter estimates of the demand model are presented in Table 2. The Table contains demand estimates of five specifications. All five specifications are under the dynamic single purchase setting, assuming that consumers look forward into the future waiting for the best purchase and will drop out of the market once a purchase has been made. The first two columns present the results of the simple logit model with only mean taste parameters, a model considered by Melnikov (2013), where the evaluation of the market per period, a term similar to logit inclusive value, is captured by time dummies. Column 1 shows the results from the simple OLS logit model, ignoring the possible effect of endogenous price, while column 2 takes into account the endogeneity issue. Theoretically, if some unobserved product quality is positively correlated with price, then an estimation that doesn’t address the issue properly would lead to over-estimated parameters. As shown, the price coefficient in the OLS logit model has been substantially over-estimated, compared to the one estimated using the instrument: $-1.73$ versus $-1.44$. For other coefficients, OLS logit and IV logit generate similar results, although with IV logit has overall higher significance. As expected, price contributes negatively to consumers’ utility, with a mean coefficient of $-1.73$. The magnitude of price suggests that consumers’ demand for a bundle is elastic. Consumers obtain a negative gross utility(constant term in mean coefficients) from the bundle when all other characteristics equal zero, a comparison between a product in the market and outside option. Negative value is a simple reflection of the market share of product compared to the greater share of the outside option. Constant terms together with time dummies also capture both any change in competition among rivals and demand shocks that are common to the industry. Note that all parameters are estimated based on the discounted utility.

All parameters of the bundle characteristics, including screen size, battery life, CPU speed, dual-SIMs-dual-standby and whether the smartphone model is an imported brand, have expected signs and most of them are significant.\[^{11}\] Among all the product characteris-\[^{11}\] I excluded camera pixels from demand estimation, due to its highly colinearity with screen size. A manufacturer would not build a model that has a huge gap between camera pixels and screen size, as only when they have similar quality level would they together provide the best performance. I further tested by including camera pixels in demand estimation, and as I expected, none of the coefficients of camera pixels are significant in any model specifications.
Table 2: Parameter Estimates of Dynamic Demand Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>OLS Logit</th>
<th>Instrumental Variable Logit</th>
<th>Random Coefficients</th>
<th>Random Coefficients with Micro-moments</th>
<th>(4) with iPhone-time Dummy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>−4.79 (2.19)**</td>
<td>−3.97 (1.73)</td>
<td>−6.1 (1.16)**</td>
<td>−9.45 (1.18)**</td>
<td>−15.73 (1.19)**</td>
</tr>
<tr>
<td>Log price</td>
<td>−1.44 (0.23)**</td>
<td>−1.73 (0.38)**</td>
<td>−2.01 (0.63)**</td>
<td>−2.13 (0.6)**</td>
<td>−1.91 (0.61)**</td>
</tr>
<tr>
<td>Log ss</td>
<td>1.21 (0.45)**</td>
<td>1.41 (0.48)**</td>
<td>0.95 (0.23)**</td>
<td>1.04 (0.23)**</td>
<td>0.86 (0.36)**</td>
</tr>
<tr>
<td>Log battery</td>
<td>0.94 (0.35)**</td>
<td>0.99 (0.35)**</td>
<td>0.61 (0.19)**</td>
<td>0.68 (0.2)**</td>
<td>0.71 (0.37)**</td>
</tr>
<tr>
<td>Log CPU</td>
<td>0.54 (0.18)**</td>
<td>0.6 (0.18)**</td>
<td>0.41 (0.08)**</td>
<td>0.44 (0.09)**</td>
<td>0.32 (0.11)**</td>
</tr>
<tr>
<td>DSIDS</td>
<td>0.33 (0.11)**</td>
<td>0.3 (0.11)**</td>
<td>0.14 (0.06)**</td>
<td>0.11 (0.06)**</td>
<td>0.17 (0.11)**</td>
</tr>
<tr>
<td>Imported</td>
<td>0.95 (0.17)**</td>
<td>1.04 (0.19)**</td>
<td>0.66 (0.08)**</td>
<td>0.71 (0.08)**</td>
<td>0.74 (0.08)**</td>
</tr>
</tbody>
</table>

Measure of fit:

<table>
<thead>
<tr>
<th></th>
<th>Adjusted R²</th>
<th>F-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>First stage:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.95</td>
<td>51.49**</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>51.5**</td>
<td>51.5**</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses. All models include manufacturer dummies and time dummies. There are 883 observations. Significance level indicator: **p < 0.05; ***p < 0.01; ****p < 0.001.

...an imported brand seems to enter consumers’ utility function with the greatest impact, increasing the utility by 1% with the indicator equal one. This impact is conditional on having controlled for other characteristics, meaning that with the same characteristics, a smartphone of a foreign brand is more popular than a native brand. This could be the original design, or operating system (each manufacturer usually is dedicated to one operating system at a time), or some other quality enhancers that are not measurable. Last but not least, the parameters of characteristics are smaller in absolute values compared to the constant term, suggesting that the product differentiations among bundles are important but smaller than the differentiation from the outside options.

For the test of validity of instruments, the “First stage” section in Table 2 shows that both values of the first stage $R^2$ and $F$ statistic are high, implying that the instruments used here are of high relevance to the endogenous price and therefore are valid.

Throughout column 3 to column 5 other than the price related coefficients, parameters are estimated based on the monthly flow utility of the bundle. To make the results comparable to the first two columns, magnitudes of coefficients other than price (including mean and random coefficients) from column 3 to column 5 are multiplied by the discounted factor, which is $(1 - \beta^{T-1})/(1 - \beta)$. Column 3 presents the estimates using the dynamic
model described in section 4.1, a base line model with only two random coefficients, one on constant term and one on price. The random coefficients on the constant term and price are both positive and significant, suggesting there are substantial variations among consumers in the utility from a bundle as well as distaste on price. In particular, the magnitude of the random coefficient on price is higher than that on the constant term, indicating consumers are more sensitive to the price of the product relative to quality. Adding random coefficients to help explain the variation in price, the mean parameter of price is lower than the simple logit model.

Column 4 provides the estimates using the same model as column 3 but with a richer setting, using additional micro-moments to infer the effect of consumers’ demographic characteristics on the bundle utility. The implications of demographic characteristics are helpful to further explain the variations in consumers’ preferences over the mean utility and price. Information on consumer’s income is helpful to explain consumers’ various tastes on price. As I do not observe the distribution of consumers’ income, I use educational achievement as an approximation for income, assuming that consumers with higher educational achievements generally have more income. The indicator of being a city resident and in a certain age group help us understand the variation in consumers’ evaluations over the mean utility. As the results show, a city resident values the bundle more than a non-city resident by 19%; young adults value the bundle more than consumers from other age groups. On the price front, having a senior high school degree does not make a significant difference in price sensitivity compared to those who are less educated; on the other hand, having a bachelor degree makes a consumer 13% less price sensitive than those who have a lower educational achievement. Mean coefficients estimates show similar magnitude between column 3 and column 4.

Finally, column 5 shows the estimates of adding the iPhone-time interaction term on top of the model setting in specification (4). Specification (4) assumes the level of service that the retailer picks for each manufacturer is not substantially different from each other, so the effect of service varies only by time but not across manufacturers. As a result, the time dummies capture the effect of retailer’s service together with other unobserved time specific demand shocks. Specification (5) further allows for the retailer to differ the service between manufacturers. Under the impression that an exclusive contract gives the retailer incentive to provide extra service for the manufacturer, the iPhone-time dummies will pick up the special treatment that the retailer offers specifically for iPhone. Also captured by the iPhone-time dummies are any iPhone-time specific shocks, which includes the promotion effort from the manufacturer. As shown, most of the parameters have similar values in comparison to column (4) with the exception of the random coefficient of city residence. The price coefficient is slightly lower than the one in columns (3) and (4). The overall significances for the mean coefficients are lower than (3) and (4), which is probably caused by adding the extra iPhone-time dummies. Given that results do not differ much between columns (4) and (5), and given the concern of over identification, I will use the results from column (4) in the service estimation.

Continuing with demand estimates, Table 3 presents the value of manufacturer’s fixed effect under four specifications. The five specifications present similar results, bearing in mind that results from the last three columns are multiplied by the discount vector to be comparable across specifications. Consumers show strong preference towards iPhone
Table 3: Manufacturer fixed effect estimates

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>OLS Logit</th>
<th>Instrumental Variable Logit</th>
<th>Random Coefficients</th>
<th>Random Coefficients with Micro-moments</th>
<th>(4) with iPhone-time Dummy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coolpad</td>
<td>0.84 (0.16)**</td>
<td>0.86 (0.15)**</td>
<td>0.49 (0.08)**</td>
<td>0.5 (0.08)**</td>
<td>0.76 (0.15)**</td>
</tr>
<tr>
<td>Huawei</td>
<td>0.72 (0.16)**</td>
<td>0.72 (0.16)**</td>
<td>0.43 (0.09)**</td>
<td>0.45 (0.09)**</td>
<td>0.67 (0.15)**</td>
</tr>
<tr>
<td>Apple</td>
<td>2.35 (0.2)**</td>
<td>2.53 (0.28)**</td>
<td>1.49 (0.08)**</td>
<td>1.58 (0.08)**</td>
<td>4.01 (1.06)**</td>
</tr>
<tr>
<td>Lenovo</td>
<td>0.57 (0.15)**</td>
<td>0.56 (0.15)**</td>
<td>0.33 (0.08)**</td>
<td>0.33 (0.08)**</td>
<td>0.59 (0.15)**</td>
</tr>
<tr>
<td>Samsung</td>
<td>-0.01 (0.13)</td>
<td>-0.02 (0.13)</td>
<td>0.00 (0.07)</td>
<td>0.01 (0.07)</td>
<td>0.06 (0.12)</td>
</tr>
<tr>
<td>ZTE</td>
<td>0.85 (0.17)**</td>
<td>0.86 (0.16)**</td>
<td>0.5 (0.09)**</td>
<td>0.51 (0.09)**</td>
<td>0.76 (0.16)**</td>
</tr>
</tbody>
</table>

**Note:** Standard errors are in parenthesis. Significance level indicator: *p < 0.05; **p < 0.01; ***p < 0.001.

relative to other models and value the bundle with iPhone on average 3 times more than the bundles with other phones. For all native manufacturers, consumers do not seem to distinguish much among brands. Each of the six large manufacturers except Samsung are significantly more popular than other fringe manufacturers (baseline). Part of the reason Samsung does not stand out may be its wide-spread product spectrum. Manufacturers other than Samsung have particular marketing strategies, eg., Apple markets itself as high-end luxurious product and other native brands target the major population with economy models – less advanced configurations with very affordable prices. Samsung’s product line reaches to both ends. For example, on the high-end, Samsung Galaxy S II GT-I9100 is a relative close substitute to the iPhone with bundle list price 4938RMB, whereas on the low end, Samsung Corby S3370 only costs 1240RMB.

In section 3.1, the logit inclusive value is defined as the value of purchasing as opposed to the value of the outside option, where the value of purchasing in my case is the value of purchasing a bundle from China Unicom as opposed to other alternatives. Figure 7 describes the evolution of this value over time. The line LIV (4) denotes the logit inclusive value computed from specification (4) – logit model with random coefficients and micro-moments. The time dummy in Melinkov’s model (specification (2)), denoted by LIV (2) is equivalent to the logit inclusive value and therefore can be used as a reference. I also include the time fixed effect and the iPhone-time fixed effect, in which the former captures the time specific demand shock that is common to all products, including the carrier’s general services for all products, and the latter captures the retailer’s special service for iPhone and other product promotions conducted by Apple. Logit inclusive value from specification (2) and (4) overall presents similar upward trends over time, although the values at the early periods are higher in specification (2) than (4). In particular, both lines go flat in the first half of 2012, which could indicate the impact of Apple forming a partnership with China Telecom. Also, there is an “M” shape pattern around January every year, which is the time right before the most significant Chinese holiday – the Spring Festival. The first spike could reflect the consumer’s increasing need to shop for the holiday and the retailer’s extra promotion effort in catering the increasing demand. Most firms will close business during the holiday which may explain the drop in the value right after the first spike and before they resume business.

27
Figure 7: Value of purchasing from China Unicom after the holiday. The time fixed effect does not have much fluctuation other than the “M” shape holiday effect, suggesting the retailer’s general service effort is relatively constant over time. However, the iPhone-time fixed effect displays more fluctuations and a slight drop over time. In particular, each spike on the curve happens during the time where a new iPhone model was being released, which are September 2010 (iPhone 4), April 2012 (iPhone 4S) and December 2012 (iPhone 5). Another interesting feature of the iPhone-time fixed effect is that as opposed to the other curves, there is a significant drop between February 2012 to March 2012. Such a drop is more likely due to the reduction of the retailer’s special service for iPhone; it would not otherwise make sense for Apple to reduce advertising just before releasing a new model. Apple formed a partnership with China Telecom in March 2012, which is probably the cause of China Unicom’s drop in special service as a response to the event.

For a robustness check, I also estimate the static version of the demand model. As opposed to the dynamic model where bundle is a durable good and consumers only purchase once in a life time, a static model assumes the bundle contract only last for one period and consumers are free to purchase a different bundle the next period. The estimation results are shown in Table 4. The two specifications here correspond to the random coefficients and random coefficients with micro-moments in the dynamic setting respectively. For the mean coefficients other than price, the results from the static model do not differ much from the dynamic model. However, they are significantly different from the dynamic model in the random coefficients, particularly the specification with micro-moments. The mean coefficient on price in the static model is slightly lower than that in the dynamic setting, −2.05 versus −2.01 for the model with only random coefficients, however higher in the specification with extra micro-moments. In addition, all of the demographic coefficients other
Table 4: Parameter Estimates of Static Demand Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Static Random Coefficients</th>
<th>Static Random Coefficients with Micro-moments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>Random Coefficients</strong></td>
<td>{Φ, Σ}</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.22 (0.4)**</td>
<td>0.26 (0.42)</td>
</tr>
<tr>
<td>Price</td>
<td>0.2 (0.05)**</td>
<td>0.08 (0.11)</td>
</tr>
<tr>
<td>City</td>
<td></td>
<td>0.48 (0.36)</td>
</tr>
<tr>
<td>Senior high</td>
<td>-0.16 (7.35)</td>
<td></td>
</tr>
<tr>
<td>Bachelor</td>
<td>3.77 (0.42)**</td>
<td></td>
</tr>
<tr>
<td>Age: 18 – 34</td>
<td>5.61 (3.57)</td>
<td></td>
</tr>
<tr>
<td>Age: 35 – 54</td>
<td>4.41 (3.15)</td>
<td></td>
</tr>
<tr>
<td><strong>B. Mean Coefficients</strong></td>
<td>{α, β}</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-6.161 (1.17)**</td>
<td>-11.211 (1.17)**</td>
</tr>
<tr>
<td>Log price</td>
<td>-2.05 (0.01)**</td>
<td>-2.06 (0.04)**</td>
</tr>
<tr>
<td>Log ss</td>
<td>0.81 (0.23)**</td>
<td>0.89 (0.22)**</td>
</tr>
<tr>
<td>Log battery</td>
<td>0.57 (0.19)**</td>
<td>0.59 (0.19)**</td>
</tr>
<tr>
<td>Log CPU</td>
<td>0.35 (0.09)**</td>
<td>0.38 (0.09)**</td>
</tr>
<tr>
<td>DSDS</td>
<td>0.16 (0.06)**</td>
<td>0.15 (0.06)**</td>
</tr>
<tr>
<td>Imported</td>
<td>0.59 (0.08)**</td>
<td>0.62 (0.08)**</td>
</tr>
</tbody>
</table>

First stage:
R²: 0.94  0.95
F-Statistic: 51.5***  51.5***

*Note: Standard errors are in parenthesis. All models include manufacturer dummies and time dummies. There are 883 observations. Significance level indicator: *p < 0.05; **p < 0.1; ***p < 0.001.*

than “bachelor” show more variations in the static model. The overall significance level in the static model is lower than the dynamic model. To conclude, static and dynamic models generate results with substantial differences, and the dynamic model is more appropriate here knowing consumers are forward-looking.

5.2 Service Estimation

The estimates of the service parameter and other cost parameters are presented in Table 5. The demand estimates used in service estimation are from the specification of random coefficients with micro-moments. Table 5 contains two sets of results: the first set is the parameter of the linear function of λₜ – the constant term and the time trend coefficient; the second set contains manufacturer cost shifters. Both of the service parameters are significant. A positive ψ₀ and a negative ψ₁ suggest that service increases the demand, however, its impact subsides over time. For a new product like the iPhone, special service matters more in the early periods in the sense that consumers rely more on service for information about the product when it is relatively new in the market, and less so when the product is more familiar to consumers. Given the two coefficients, the service parameter
can be constructed using Equation (30). Note that service enters the flow utility, so the overall effect of service on demand should be the discounted sum of the effect over time. Using superscript $^{dr}$ as the notation of discounted value, the benefit of service on demand relative to service costs is declining over time with a mean value $5.71e^{-5}$.

Most of the cost coefficients – except for DSDS – have expected signs. Screen size, battery and CPU all have a significant effect on increasing costs, and screen size seems to have the most impact among all the cost drivers, increasing cost by 0.71% with a 1% increase in screen size. This matches the results of a bill-of-materials analysis for an smartphone, in which screen takes up the biggest proportion of the manufacturer cost. It is surprising to see the DSDS turn out to be an cost reducer, and I could not explain why the coefficient is negative. Having controlled for the hardware costs, marginal costs are not substantially different between the six large manufacturers except for Apple. First stage statistics show high $R^2$ and $F$ – Statistics, suggesting the instruments used here are valid. Figure 8 shows the recovered wholesale price-cost markup ratio given the service parameters. The markups overall are constant with a slight decline over time, and are more volatile for native manufacturers than foreign. iPhone has the highest wholesale markup, with an average 45% markup over marginal costs; Samsung has the second highest wholesale markup, which is 7% lower than iPhone. Native brands have an average markup around 20%, in particular,
ZTE is the lowest of all with an average markup of 13%.

Given the fact that service has a significant effect on increasing demand, this naturally leads to a discussion on how much we would miscalculate manufacturer costs if we ignored the service factor. Table 6 provides the distribution of manufacturer costs under two scenarios: with service consideration and without.\(^2\) The first block in the table presents the costs distributions by manufacturer when taking service into account, and the second block shows the results when we ignore the service effect. Let us start with the first block. Apple has the highest costs overall of all smartphone manufacturers. Native brands Coolpad, Huawei and ZTE share similar costs (all cheaper than Apple, Lenovo and Samsung) and amount to only 1/3 the costs of iPhones. In addition, Lenovo and Samsung have more product variety than the other manufacturers. Since all six manufacturers have at least one product on the market each period, more observations suggest more model varieties. Results from the second block show that ignoring the service effect leads to an underestimation of the marginal costs, however, the degree varies greatly between manufacturers. Neglecting the service factor would not substantially bias the estimated costs for Coolpad, Huawei, Lenovo and ZTE, however, it will lead to underestimating the marginal costs for Samsung by almost 2%. Failing to include the service factor will also cause a significant downward bias for iPhone marginal cost by 7.77%.

\(^2\)Half of the recovered marginal costs found using Equation (20) are negative. This is because consumers would not know how much data they actually need before the contract. Because China Unicom does not allow for downgrading plans later, it is better to start with the introductory plan and upgrade as needed. Here I assume all consumers upgrade their data plan by one level after signing the contract. This is probably an underestimation compared to the upgrading pattern in reality. Because this extra charge happens after consumers purchase the bundle, it is not included in the retail price and thus would not affect the price coefficient. Therefore, I add the extra charges to the recovered marginal costs using manufacturer’s first-order
Table 6: Recovered Costs and Wholesale Markups

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manufacturer Costs (RMB)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coolpad</td>
<td>92</td>
<td>638</td>
<td>189</td>
<td>151</td>
<td>903</td>
</tr>
<tr>
<td>Huawei</td>
<td>94</td>
<td>712</td>
<td>195</td>
<td>443</td>
<td>1,247</td>
</tr>
<tr>
<td>Apple</td>
<td>79</td>
<td>2,454</td>
<td>573</td>
<td>980</td>
<td>3,431</td>
</tr>
<tr>
<td>Lenovo</td>
<td>102</td>
<td>836</td>
<td>467</td>
<td>318</td>
<td>1,895</td>
</tr>
<tr>
<td>Samsung</td>
<td>116</td>
<td>1,124</td>
<td>648</td>
<td>433</td>
<td>2,649</td>
</tr>
<tr>
<td>ZTE</td>
<td>74</td>
<td>649</td>
<td>180</td>
<td>320</td>
<td>1,137</td>
</tr>
<tr>
<td><strong>Wholesale Markup Reduction %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coolpad</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td>0.86</td>
</tr>
<tr>
<td>Huawei</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td>1.54</td>
</tr>
<tr>
<td>Apple</td>
<td>5.37</td>
<td></td>
<td></td>
<td></td>
<td>0.9</td>
</tr>
<tr>
<td><strong>Bias of Costs if Ignore Service %</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coolpad</td>
<td>-0.41</td>
<td></td>
<td></td>
<td></td>
<td>-0.62</td>
</tr>
<tr>
<td>Huawei</td>
<td>-0.59</td>
<td></td>
<td></td>
<td></td>
<td>-1.8</td>
</tr>
<tr>
<td>Apple</td>
<td>-7.77</td>
<td></td>
<td></td>
<td></td>
<td>-0.63</td>
</tr>
</tbody>
</table>

*Note:* All recovered costs include the extra plan charge, which is the discounted sum of upgrading the data plan by one level. Underestimation percentage represents the mean of percentage of marginal costs being underestimated if one ignores service effect.

On the flip side of the coin, the differences between manufacturer costs with and without the service factor also reflect the different pricing behaviors when a manufacturer wants to encourage more service from retailers. Theoretically, increasing the wholesale price will affect service negatively; a higher wholesale price will lead to a higher retail price, which in turn causes a reduction in sales. Retailers have less incentive to provide service as the marginal increase in sales from service would not be as high as if the sales were high. As the results show, a unit increase in wholesale price will cause an average 0.15 reduction in service, and the degree is more for manufacturers with larger market shares. In particular, such an increase in wholesale price in the early periods will cause the service for iPhones to reduce by 0.28. With this said, a manufacturer who wants more service and has a relatively larger market share should reduce wholesale price more than the manufacturer who does not. Comparing the two sets of wholesale markups using the recovered marginal costs under the two scenarios, taking into account retailer’s service optimization will lower the wholesale markup of iPhone by 5.4%.

One thing worth mentioning is that, if it is given that the manufacturer will lower the wholesale margin to increase retailer service and that service has less impact on demand as time goes by, with all else holding constant, we should expect the wholesale markup ratio to rise for iPhone and Samsung over time. However, from what is shown in Figure 8, the trend for the wholesale markups is rather flat for both manufacturers. One probable explanation is the increased competition from the rival native manufacturers. Recall in condition, and all the marginal costs are positive after doing so.
Figure 5, close to the end of the data period, the sales of native brands surpassed the sales of iPhone, and the increasing competitive pressure from rival manufacturers prevents Apple from increasing the markup.

The results show that Apple values Unicom’s service much more than any other manufacturer and is willing to substantially lower the wholesale price to induce more special service for the iPhone. In this sense, the initial choice of exclusive contract with Unicom is optimal for Apple; it increases Unicom’s sales on the iPhone and at the same time eliminates the possibility of rivals free-riding on Unicom’s service. Both of these advantages give Unicom extra incentive to exert more effort into special service for the iPhone. However, the benefit of service on demand subsides over time and when service can no longer significantly increase the demand, continuing the exclusive contract would no longer be beneficial. If the main purpose of inducing retailer service is to increase the sales, at this stage, assigning more agents would be more helpful. This could be one potential explanation as to why Apple had an exclusive contract with China Unicom for two years and six months before forming a second partnership with China Telecom.

6 Conclusion

This paper develops a method to evaluate the role of retailers in determining service, even when service is not observed. The identification of service comes from the variation of manufacturer’s wholesale prices, in the sense that retailer’s optimization of service enters the manufacturer’s first-order condition and thus alters its price setting. First I construct and estimate a demand model using a panel dataset of price, quantity and characteristics of the “phone+network” bundle offered by China Unicom, the second largest wireless carriers in China. The data also provide the wholesale prices of the phones charged by manufacturers to China Unicom. The demand model allows for consumers to be forward-looking and with persistent heterogeneous preferences regarding product characteristics. The estimates of the demand model are then used in a structural supply model to estimate the service parameters. In the supply model, the retailer sets both final price and service level after manufacturer determines the wholesale price.

The demand estimates suggest substantial variations in consumer preferences, and adding micro-moments yields more sensible results. A static analysis using the same data provides with less realistic results, suggesting that the use of a dynamic demand model is important in analyzing consumer purchases of durable goods.

In the service estimation, I find that service has a positive significant effect on demand, however, the impact subsides over time. Among all the manufacturers, Apple values retailer’s service the most and willing to lower wholesale markup by 5% to induce more retailer service. Other than Apple, China Unicom does not seem to provide special service for other phone manufacturers. Using the recovered marginal costs, I find that the iPhone has the greatest wholesale margin given that Apple has lowered the margin for the service purposes. I also find that ignoring the service effect will not significantly bias the estimated marginal costs for the native manufacturers, however, it does cause a substantial downward bias for Apple and Samsung. The pattern in service thus provides a potential explanation to Apple’s initial exclusive contract with China Unicom and the subsequent contract arrangement.
References


Appendix

A. Micro-Moments

The idea of micro-moments is essentially to match the model predictions for the attributes of consumer that purchases a new model to the counterpart observed in the data. To form the sample moments, take the attribute of consumer being a city resident for example, the population moment is,

\[ E[\bar{z} - E[z_i|y_1^j, \theta_2]] = 0, \]

where

\[ z = 1\{\text{city}\} \text{ and } y_1^j = 1\{\text{product } j}\]

I cannot calculate the \( E[z_i|y_1^j, \theta_2] \) exactly, however, it can be approximated using Bayes’ rule:

\[ E[z_i|y_1^j, \theta_2] = \int z Pr(z|y_1^j, \theta_2) = \frac{\int z \int v Pr(y_1^j|z, v, \theta_2) P(dv, dz)}{Pr(y_1^j, \theta_2)} \approx \frac{1}{ns} \frac{1}{s_j} \sum_r z_r Pr(y_1^j|z_r, v_r, \theta_2, \delta_i(\theta_2)), \]

where \( \delta_i(\theta_2) \) is the utility level, given demand parameter \( \theta_2 \), that rationalizes the observed market share \( s_j \). Then substitute (32) into (31) to obtain the sample moments.

It is worth discussing the validity of the use of micro-data. All the data averages attributes of consumers that purchase a new model are measurements on nationwide basis, including all possible purchasing channels (three carriers and other model retailers). However, model predictions contain only one carrier: China Unicom. To establish an apple-to-apple match, I need to drop off the signs of the effect of average attributes that are sensitive to different purchasing channels, since it is likely to bias the corresponding micro-moments. I proceed with the filtering by first dividing consumers into a carrier group and regular retailer group, the former preferring to obtain the smartphone through a “phone+network” bundle from a carrier and the other group preferring to get the smartphone and network service through separate transactions. Then, I further divide the consumers in the carrier group into three subgroups, each preferring his own carrier to the other two. Within each layer, I will discuss whether the demographic variables, living area, age and education level would affect some group in the opposite direction.

I start with the broader layer. Price and model are the two major factors that affect consumers choice between bundle and the phone by itself: which has the desirable model and which has the overall better price. With no doubt, a rational consumer will pursue the channel that provides a cheaper price for his/her desired model, hence, it is not likely that any of the demographic variables would have an opposite effect on the consumers in either group. Moving on to the finer layer, groups that each prefer a different carrier. If say one of the carriers, China Mobile, has better signal coverage in rural area, which is likely to be true, then consumers that are rural residents would prefer China Mobile to the other.

36
two carriers. This would lead to a sign flip between China Mobile group and the other two
groups. A minor concern arises from the various level of stickiness to the original carrier
among age groups. China Mobile is the oldest and also the largest carrier under the 2G
network; the age structure of its 2G subscribers skew towards the senior consumers relative
to other two carriers. If the stickiness is increasing with age, then the effect of age on carrier
choice for the consumers in senior age group is likely to be positive for China Mobile and
negative for the other two.

B. Derivations of matrix $\Delta_n$, $\Delta_z$ and $\Delta_{sz}$

This section provides the expressions for matrix $\Delta_c$, $\Delta_{sz}$, and components required for $\Delta_p$
and $\Delta_z$. In the following derivations, I omit all the time subscripts for simple notations.
To proceed, I start with market share response matrix with respect to retail price $\Delta_c$. This
simply follows the routine described in Berry et al. (1995), such that expression of element
$(j, q)$ in $\Delta_n$ is given by,

$$
\frac{\partial s_q}{\partial p_j} = \begin{cases} 
\sum_{i=1}^{n_s} g_i \alpha_i s_{ij} (1 - s_{ij}), & \text{if } j = q; \\
- \sum_{i=1}^{n_s} g_i \alpha_i s_{iq} s_{ij}, & \text{if } j \neq q,
\end{cases}
$$

(33)

where $\alpha_i$ denotes consumers' heterogeneous preferences over price with subscript $i$ represent-
ing each consumer type, and $g_i$ is the importance weight assigned to $i$. Furthermore,
the second order derivatives of market share with respect to price can be derived as

$$
\frac{\partial^2 s_q}{\partial p_j \partial p_k} = \begin{cases} 
\sum_{i=1}^{n_s} g_i \alpha_i^2 s_{ij} (1 - s_{ij}) (1 - 2s_{ij}), & \text{if } j = q = k; \\
\sum_{i=1}^{n_s} g_i \alpha_i^2 s_{ij} s_{ik} (2s_{ik} - 1), & \text{if } j = q \text{ and } q \neq k; \\
\sum_{i=1}^{n_s} g_i \alpha_i^2 s_{iq} s_{ij} (2s_{iq} - 1), & \text{if } j \neq q \text{ and } q = k; \\
2 \sum_{i=1}^{n_s} g_i \alpha_i^2 s_{iq} s_{ij} s_{ik}, & \text{otherwise.}
\end{cases}
$$

(34)

Matrix $\Delta_{sz}$ is the sales response matrix with respect to service. Define $\tilde{\lambda} = \sum_{\tau=1}^{T_c} \tau^{\tau-1} \lambda$, which captures the service effect on discounted utility. Element $(m, q)$ in matrix $\Delta_{sz}$ is a row
vector, with a length equal the number of products of manufacturer $m$, and its expression
is given by

$$
\frac{\partial s_q}{\partial z_m} = \begin{cases} 
\tilde{\lambda} \sum_{i=1}^{n_s} g_i s_{iq} (1 - \sum_{k \in \mathcal{J}_m} s_{ik}), & \text{if } q \in \mathcal{J}_m; \\
- \tilde{\lambda} \sum_{i=1}^{n_s} g_i s_{iq} (\sum_{k \in \mathcal{J}_m} s_{ik}), & \text{otherwise.}
\end{cases}
$$

(35)
Together with Equation (33) and Equation (35), the expression for the second order derivatives of market share with respect to price and service in $\kappa(j,l)$ in Equation (16), is shown as the following,

$$\frac{\partial^2 s_q}{\partial p_j \partial z_l} = \begin{cases} 
\tilde{\lambda} \sum_{i=1}^{n_s} g_i \alpha_i s_{ij}(1 - 2s_{ij})(1 - \sum_{k \in J_l} s_{ik}), & \text{if } j = q \text{ and } \{j,q\} \subset J_l; \\
\tilde{\lambda} \sum_{i=1}^{n_s} g_i \alpha_i s_{ij}(2s_{ij} - 1)(\sum_{k \in J_l} s_{ik}), & \text{if } j = q \text{ and } \{j,q\} \not\subset J_l; \\
2\tilde{\lambda} \sum_{i=1}^{n_s} g_i \alpha_i s_{ij} (\sum_{k \in J_l} s_{ik} - 1), & \text{if } j \not= q \text{ and } \{j,q\} \subset J_l; \\
\tilde{\lambda} \sum_{i=1}^{n_s} g_i \alpha_i s_{iq} (\sum_{k \in J_l} s_{ik} - 1), & \text{if } j \not= q \text{ and } j \not\in J_l \text{ or } q \not\in J_l; \\
2\tilde{\lambda} \sum_{i=1}^{n_s} g_i \alpha_i s_{ij} (\sum_{k \in J_l} s_{ik}), & \text{otherwise}. 
\end{cases}$$

(36)

Given the market share response matrix to service, its second-order derivatives can be derived for in $(j,k)$, where $\{j,k\} \subset J_A$, as,

$$\frac{\partial^2 s_q}{\partial z_m \partial z_l} = \begin{cases} 
\hat{\lambda} \sum_{i=1}^{n_s} g_i s_{iq}(1 - \sum_{k \in J_m} s_{ik})(1 - 2\sum_{k \in J_m} s_{ik}), & \text{if } q \in J_m \text{ and } m = l; \\
\hat{\lambda} \sum_{i=1}^{n_s} g_i s_{iq}(\sum_{k \in J_m} s_{ik})(2\sum_{k \in J_m} s_{ik} - 1), & \text{if } q \in J_m \text{ and } m \not= l; \\
\hat{\lambda} \sum_{i=1}^{n_s} g_i s_{iq}(\sum_{k \in J_m} s_{ik})(2\sum_{k \in J_l} s_{ik} - 1), & \text{if } q \not\in J_m \text{ and } m = l; \\
\hat{\lambda} \sum_{i=1}^{n_s} g_i s_{iq}(\sum_{k \in J_m} s_{ik})(\sum_{k \in J_l} s_{ik}), & \text{if } q \not\in J_m \text{ and } m \not= l; \\
2\hat{\lambda} \sum_{i=1}^{n_s} g_i s_{iq}(\sum_{k \in J_m} s_{ik})(\sum_{k \in J_l} s_{ik}), & \text{otherwise}. 
\end{cases}$$

(37)
### Table 7: Manufacturer Summary Statistics

<table>
<thead>
<tr>
<th>Name</th>
<th>Import</th>
<th>O.S.</th>
<th>DSDS</th>
<th>Market shares (10^-4)</th>
<th>Retail price (RMB)</th>
<th>Product counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Std.</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Big Six</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apple</td>
<td>Yes</td>
<td>iOS</td>
<td>0</td>
<td>2.46</td>
<td>1.06</td>
<td>4,874.4</td>
</tr>
<tr>
<td>Lenovo</td>
<td>No</td>
<td>Andr</td>
<td>0.59</td>
<td>2.01</td>
<td>1.89</td>
<td>1,275.5</td>
</tr>
<tr>
<td>Coolpad</td>
<td>No</td>
<td>Andr</td>
<td>0.44</td>
<td>1.93</td>
<td>1.83</td>
<td>997.7</td>
</tr>
<tr>
<td>ZTE</td>
<td>No</td>
<td>Andr</td>
<td>0.36</td>
<td>1.23</td>
<td>0.85</td>
<td>855.7</td>
</tr>
<tr>
<td>Huawei</td>
<td>No</td>
<td>Andr</td>
<td>0.33</td>
<td>1.09</td>
<td>1.05</td>
<td>1,020.2</td>
</tr>
<tr>
<td>Samsung</td>
<td>Yes</td>
<td>Andr, Oths</td>
<td>0.26</td>
<td>0.95</td>
<td>1.10</td>
<td>2,150.2</td>
</tr>
<tr>
<td><strong>Fringe Brands</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xiaomi</td>
<td>Yes</td>
<td>Oths</td>
<td>0</td>
<td>0.68</td>
<td>0.39</td>
<td>1,885.1</td>
</tr>
<tr>
<td>Nokia</td>
<td>Yes</td>
<td>Sym, Win</td>
<td>0</td>
<td>0.59</td>
<td>0.29</td>
<td>1,636.3</td>
</tr>
<tr>
<td>HTC</td>
<td>Yes</td>
<td>Andr</td>
<td>1</td>
<td>0.55</td>
<td>0.28</td>
<td>2,123.6</td>
</tr>
<tr>
<td>Hisense</td>
<td>No</td>
<td>Andr</td>
<td>0.89</td>
<td>0.43</td>
<td>0.21</td>
<td>667.8</td>
</tr>
<tr>
<td>Ktoucch</td>
<td>No</td>
<td>Andr, Win, Oths</td>
<td>0.83</td>
<td>0.35</td>
<td>0.34</td>
<td>961</td>
</tr>
<tr>
<td>Meizu</td>
<td>No</td>
<td>Oths</td>
<td>0</td>
<td>0.32</td>
<td>0.12</td>
<td>2,299</td>
</tr>
<tr>
<td>Motorola</td>
<td>Yes</td>
<td>Andr</td>
<td>0.28</td>
<td>0.25</td>
<td>0.36</td>
<td>1,691</td>
</tr>
<tr>
<td>Gionee</td>
<td>No</td>
<td>Andr</td>
<td>1</td>
<td>0.24</td>
<td>0.10</td>
<td>1,149</td>
</tr>
<tr>
<td>Philips</td>
<td>Yes</td>
<td>Andr</td>
<td>1</td>
<td>0.21</td>
<td>0.11</td>
<td>878</td>
</tr>
<tr>
<td>Malata</td>
<td>No</td>
<td>Andr</td>
<td>0</td>
<td>0.18</td>
<td>0.05</td>
<td>701.9</td>
</tr>
<tr>
<td>Haier</td>
<td>No</td>
<td>Andr</td>
<td>1</td>
<td>0.21</td>
<td>0.05</td>
<td>799</td>
</tr>
<tr>
<td>Changhong</td>
<td>No</td>
<td>Andr</td>
<td>1</td>
<td>0.13</td>
<td>0.03</td>
<td>665.7</td>
</tr>
<tr>
<td>Sony Ericsson</td>
<td>Yes</td>
<td>Andr, Win</td>
<td>0</td>
<td>0.13</td>
<td>0.03</td>
<td>1,757.5</td>
</tr>
<tr>
<td>Amoi</td>
<td>No</td>
<td>Andr</td>
<td>1</td>
<td>0.07</td>
<td>0.01</td>
<td>1,199</td>
</tr>
<tr>
<td>LG</td>
<td>Yes</td>
<td>Andr</td>
<td>0</td>
<td>0.05</td>
<td>0.02</td>
<td>999</td>
</tr>
</tbody>
</table>

*Notes:* column 3 reports the operating systems each brand equipped with, “Andr”, “Sym”, “Win” are short for “Android”, “Symbian” and “Windows”. Column 4 reports the share of models in each brand that equipped with feature dual-cards-dual-standby.
Figure 9: Prices over time (sales weighted averages)

Figure 10: Wholesale prices over time (sales weighted averages)