Consumer Equilibrium, Demand Effects, and Efficiency in Group Buying

Abstract

Problem definition: We explore consumer equilibrium and efficiency in group buying events, in which the unit price for a good or service decreases with higher number of customer sign ups. Specifically, we study the following questions: (i) How does the dynamic consumer sign up equilibrium evolve during these events? (ii) Is there empirical evidence of employing group buying events improving demand? If so, by how much? (iii) What are the profit gains from employing this mechanism?

Academic / Practical Relevance: Group buying events are becoming increasingly popular, especially in emerging markets. Our study contributes to better understanding of this innovative pricing mechanism and its effects on demand and profits, and can help with improved and more effective utilization of the method in practice.

Methodology: We build a dynamic game-theoretical model to study consumer behavior and solve for its equilibrium. We then apply it to data obtained from group buying events employed by a large appliance retailer, and utilize structural regression methods and data clustering to estimate and evaluate demand and profit effects.

Results: We demonstrate that the equilibrium we derived closely approximates the observed consumer behavior. Using the estimation from this equilibrium outcome, we estimate that employing group buying on average improved the retailer’s product demand by 15.4%, and profits by 10.9%, yielding monetary gains of approximately $4.3 Million.

Managerial Implications: Our theoretical model provides insights into dynamic customer behavior during group buying events, and can be a basis for researchers and practitioners in studying and designing group buying events. We also demonstrate significant demand and profit boosting effects of group buying, providing concrete support for managers on benefits of employing the mechanism. Further, we provide statistical evidence and insights for pricing patterns that are most impactful in improving profits when employing group buying.

Key Words: Group buying, dynamic equilibrium, empirical analysis.
1 Introduction

Rapid development of technology and availability of multiple consumer channels that give many alternatives to consumers have put retailers in an unprecedentedly competitive business environment in recent years. With the proliferation of commerce on the Internet, many firms have adopted innovative techniques to attract and retain customers, while struggling to maintain and improve shrinking margins. One of the popular techniques employed recently is Group Buying, which has been pioneered on the Internet more than a decade ago by companies such as Mobshop and Groupon, and which, in recent years is becoming a fast growing retailing method, especially in emerging economies like China.

The idea behind group buying is simple: in the concept’s most basic form, a product or service is made available for purchase during a pre-announced sign-up window, i.e., the group buying event, usually lasting less than a day. If the number of customers who sign up to purchase during the event exceeds a certain threshold, a deal, meaning a discount on the unit retail price, will be available for everyone who signed up. If the number of sign ups does not exceed the pre-specified threshold, then the deal is not offered, and the product or service is available only at a higher base price. Even though the concept initially was mostly popular with services such as restaurant meals, activities, and contractor services, in certain markets it is also recently increasingly becoming popular for deals on goods and products such as electronics, household products, and clothing. In 2016, one of China’s most popular Internet commerce platform, Taobao, which is owned by Alibaba, hosted more than an estimated 200,000 individual group buying events for third-party retailers, with revenues totaling more than 32 Billion Chinese Yuan (CNY) (approximately 4.8 Billion USD). The exact group buying mechanism employed can include a number of variations, such as multiple threshold levels corresponding to multiple discount levels, and mandatory deposits customers have to pay for signing up before the deal becomes certain. For instance, Taobao employs a particular structure for all retailers for whom it hosts on its platform. The platform has two potential deals and a mandatory sign-up deposit. The required number of customers for the two deals and the deposit amount are fixed by Taobao. Platform rules also require a fixed deal price discount amount, which is chosen by the retailer along with the base price. If the first and second deals materialize, they reduce the base price by one and two price discount amounts, respectively.

Considering the price discounts given by the retailers to all customers who make a final purchase, in order for a group buying mechanism to be successful, it has to attract increased consumer demand. In a group buying event, there are two components for increased purchases: First, with the hope of lower prices when a deal happens, some customers who would not purchase at the base price may become interested in purchasing the good. This type of increase in sign ups can happen during the event in a dynamic manner before or after a deal threshold is reached. Therefore, customer incentives based on their valuations of the
product and their expectations on the sign-up behavior of other customers are critical in determining the final number of units sold, and the resulting equilibrium sign-up behavior should be studied rigorously to determine the effects of discounts on purchases. The second type of demand increase comes from network effects, namely dissemination of information about the group buying event by customers themselves to attract other customers to sign up, which can ultimately increase the likelihood of a deal happening (cf. Jing and Xie 2011, Chen and Lu 2015). In particular, when there is a group buying event, customers who are aware of the event have incentives to spread the word and inform others to consider signing up, since that increases the probability of reaching a threshold for a deal and lowering the price, in essence, acting like unpaid “sales agents” for the retailer. This behavior has an effect of increasing the demand, i.e., consumer arrival rate during the event, and can boost the number of purchases. An immediate question that arises here is whether one can quantify and test the magnitude of this effect.

Given these dynamics that group buying events introduce and the growing popularity of the concept, a number of interesting research questions emerge. Specifically, what are the equilibrium dynamics of the consumer sign-up process during the group-buying event? How do the consumers’ likelihood of signing up and the probability of the deals materializing evolve during the event? How much do the consumers’ networking and information dissemination improve demand in group buying events compared to traditional single-price sales? Taking into account the price discounts offered in group buying and the demand increases they induce, do retailers have direct net profit gains from group buying, and if so how much? Finally, what are some empirically verifiable suggestions for patterns of deal discounts in order to improve profitability of group-buying events? In this paper we aim to address these questions, by first building a theoretical model to capture customer economic behavior, and then applying it empirically to group buying and traditional single-pricing sales data obtained from Taobao.

2 Literature Review

Pricing of retail goods and services has long been studied in the literature (cf. Schmalensee and Willig 1989, Wilson 1993, Talluri and Van Ryzin 2006). Studies on pricing a retail product through a group buying mechanism, however, are relatively new, and many aspects of when and how group buying and related consumer discount based mechanisms are profitable for a retailer are still in the process of being disentangled by an increasing number of theoretical and empirical studies. One stream of theoretical literature with early roots employ batch consumer sign ups in analyzing the effectiveness of group buying. Anand and Aron (2003) derive a monopolist’s optimal group-buying schedule under different kinds of demand uncertainty and study the impact of production postponement on a group buying strategy. Their results show that the effectiveness of group buying mechanism over traditional single-pricing relies on the
nature of uncertainty on the demand curve. Hu et al. (2013) consider the case where consumers make
decisions simultaneously in group buying as a batch and show that a sequential sign-up mechanism leads
to higher deal success rates. Marinesi et al. (2016) also consider batch customer sign-ups. They show
that employing group buying allows firms better utilize their capacity. They further demonstrate that the
presence of strategic customers can be advantageous when employing group buying, and the mechanism
can help generate significant profit gains. Differing from and complementing this stream of literature, we
model a continuous time, stochastic consumer arrival and sign-up process, and find the dynamic consumer
equilibrium. Our approach allows us to study the evolution of customer sign-up patterns throughout the
event window, and enables detailed structural estimation of the parameters with our group buying data
that includes sign-up times for all consumers in each event.

Another stream of literature studies the effects of threshold pricing structure on consumer sign ups in
group buying mechanisms. Kauffman and Wang (2001) study group-buying data from one of the earliest
group buying companies, Mobshop.com, and find that number of existing sign ups and approaching sign-up
thresholds both have positive effects on new orders placed. Subramanian (2012) and Liang et al. (2014)
find that sharing information on the number of customer sign ups increase the deal success rate and
consumer surplus but reduce the profitability of the mechanism for the seller. Wu et al. (2014) explore the
possibility of increased consumer sign-up rate effects before and after the thresholds are reached, and find
empirical evidence for before threshold effects in all products, while showing that after-threshold effects
exist only for some products. Our study contributes to this branch by empirically demonstrating that
strategic consumer dynamics can explain the sign-up patterns in group buying events better compared to
other behaviors such as waiting and signing up after thresholds are crossed.

A number of papers study the profitability of group buying. Chen et al. (2002, 2007) explore optimal
bidding for risk neutral and symmetric buyers in group-buying events with fixed numbers of goods and
customers, and compare the profitability of group-buying events with traditional single-pricing. They find
that group buying can outperform the fixed price mechanism only under economies of scale or risk-seeking
sellers. Chen and Zhang (2014) find conditions, under which group buying can maximize profits, and show
that the profitability of the mechanism depends on the nature of uncertainty in the market. Chien-Wei
and Hsien-Hung (2016) argue that when customers are heterogenous in group buying costs, employing
group buying may be preferable to non-discriminated pricing. Deviating from most of the literature that
focuses on one retailer, Chen and Roma (2011) study group buying in a two-level distribution channel
with one manufacturer and two competing retailers. They show that group buying is beneficial for the
smaller of the two retailers but can hurt the larger one, while increasing supplier revenues. In our paper,
we empirically demonstrate that combining careful pricing and consumer network effects, a retailer can
increase her profits significantly with group buying events compared to single-pricing.

The genesis of group buying is retailers’ providing discounts and deals, and given the increased application of discount variants in recent years, a number of studies explore the profitability of such strategies empirically. Wu et al. (2015) analyze daily deals provided by Chinese retailers and find that merchants in fact experience losses from discounts during promotion periods, but they make profits through increased future purchases. Cao et al. (2015) quantify the impact of discount percentage on sales, and conclude that a larger discount percentage may reduce sales by being perceived as a signal of low product quality. Edelman et al. (2016) find that online discount vouchers tend to be more profitable for relatively unknown firms while being not likely to increase profits for better-known ones. Overall, existing theoretical and empirical studies in the literature paint a mixed picture on the profitability of group buying, and its desirability for sellers. One dimension we aim to contribute to the debate on this front is the consumer network effects and incentives to recruit other customers group buying creates. Jing and Xie (2011) present a theoretical model to study the effect of consumer social interactions, i.e., using a discounted price to motivate consumers to work as “sales agents” to acquire other consumers. They argue that the demand increase brought about by such social interactions can make offering discounts in group buying events profitable, and efficient interpersonal communication makes the mechanism more profitable to firms. Zhou et al. (2013) empirically study the information diffusion process in group buying, and find that mass media communication and interpersonal communication stimulate the sales at the start of the process while reducing the sales at the end. Zhang and Gu (2015) and Chen and Lu (2015) find that social factors including online interactions, media and personal recommendations positively affect consumers’ group buying intentions and social influence. On the other side of the argument, Gwee and Chang (2013) and Zhang and Tsai (2015) claim that purchases at group buying websites are usually impulsive rather than planned. Hu and Winer (2016) argue that the existence of the deal threshold does not necessarily stimulate customers to inform others, but information about tipping points may accelerate customer sign ups.

Our paper contributes to this debate by directly empirically testing demand improvement effects in group buying events and measuring them compared to losses from discounts. We provide evidence that, in fact, firms can make direct profits in group buying events despite giving significant discounts to consumers in many cases, because of the incentives group buying events create for customers to spend effort in networking and recruiting others. In sum, our study not only provides a theoretical explanation for dynamic consumer behavior in group buying events, but also provides novel empirical evidence on the sources of profitability for retailers from these events.
3 Theory

Many online retailers and platforms are developing and employing different pricing mechanisms to implement group buying. The one we will study in our paper, is the one employed by Taobao, the largest online retailing platform in China, as our data comes from the events hosted on this platform. In this mechanism, there are two possible deals, and including the base (no deal) unit price, there are three possible prices that can materialize. There is also a required non-refundable deposit for signing up. Given the platform’s fixed deal sign-up thresholds and deposit amount, the retailer chooses the base unit price and a deal discount on the unit price that applies to both deals. We first model the consumer behavior and equilibrium in this setting theoretically.

3.1 Model Description

Consider a retailer selling a product to consumers on a third-party platform. The retailer will hold a group buying event on a continuous time window indexed \([0, T]\). During this period, customers arrive following a Poisson process. (For ease of exposition, we will refer to the retailer as “she”, and each customer as “he” throughout the paper.) Each customer has a unit demand for the product, and his reservation value \(u\) has a c.d.f. denoted by \(F\) with a p.d.f. \(f\). Upon arrival, knowing the pricing pattern announced by the buyer as described below and the number of customers who have already joined up to that point, each customer decides to join or leave. If the customer joins, he pays a non-refundable deposit \(d > 0\), which, following Taobao’s process, is set fixed by the platform that hosts the event. Denote the total number of customers who joined by the end of the event window by \(N\).

The pricing of product is in the form of two-threshold group-buying discounts. Specifically, given sign-up thresholds \(M_1\) and \(M_2\), \(0 < M_1 < M_2\), fixed by the platform, the retailer announces three prices: the base price, \(p_0\), and the first and second deal prices, \(p_1\) and \(p_2\) respectively, where \(p_0 \geq p_1 \geq p_2 \geq d\). If the total number of customers who joined, \(N\) is less than \(M_1\), no group deal materializes, and the unit price of the product, \(p\), is set to the base price \(p_0\). If \(M_1 \leq N < M_2\), then the first deal will be on and the unit price will be set at \(p = p_1\). Finally, if \(N \geq M_2\), then the second deal is on, and the unit price will be set at \(p = p_2\). Following Taobao’s process, the price decrement (the deal discount) is constant and set at \(\delta \geq 0\). That is \(p_0 = p_1 + \delta\) and \(p_1 = p_2 + \delta\). At the end of the event window, i.e., at \(t = T\), after observing the total number of customers who joined, each customer makes a decision to stay in the deal or drop out. If a customer stays, he pays the balance of the price, \(p - d\), and commits to buying the product. If he drops out, however, he forfeits the deposit \(d\). Finally, denoting the number of customers that still stay after the drop outs by \(Q\), the retailer produces \(Q\) units to be delivered to the staying customers at unit cost \(c\).

The event-window itself usually has a length of 12 hours. However, the retailer announces the event
Figure 1: Timeline of Group Buying activity on Taobao.com.

and posts the prices and the deal discounts about one-week in advance. During this time before the
event, customers may network and recruit others in order to increase participation and hence increase the
probability of one of the deals happening. We call this period Phase I and the event-window Phase II.
Let $\lambda_0$ denote the arrival rate of customers during the event if the product were sold through traditional
single-pricing. Due to customer efforts in recruiting others, the arrival rate of customers during the event
increases to $\lambda_g > \lambda_0$. Figure 1 summarizes the timeline.

In order to empirically measure demand increase from the network effects, we will need to estimate
the consumer arrival rates, $\lambda_g$, for the group buying events from the sign up data. To be able to do that,
we will first have to study the dynamic consumer equilibrium behavior in Phase II, i.e., during the Group
Buying Event. In the rest of this section, we will study this equilibrium.

3.2 Consumer Equilibrium during the Group Buying Event

We start with the customers’ decision after the event. Consider a customer, who arrived and signed up at
time $t \in [0, T]$, with utility $u$. After the event, when the price $p$ is determined, he needs to decide whether
to stay or drop out. If the customer drops out he loses his deposit, and his overall payoff will be $-d$, while
if he stays, his payoff will be $u - p$. The customer will choose the larger at that point and his surplus will
be $\max\{u - p, -d\}$.

Given this post-event behavior, each consumer that arrives at time $t \in [0, T]$ observes the price structure,
the two deal thresholds, $M_1$ and $M_2$, and the total number of sign ups up to that point, (i.e., on $[0, t]$),
$N_t$. Projecting the shaping up of the rest of the event and his decision at the end of the event contingent
on the realization of the deals, he makes a decision on whether to sign up or not. Define \( N_t^+ \) as the number of sign ups on \([0, t]\), i.e., including the sign-up decision of the customer who arrives at \( t \). That is, if the customer that arrives at time \( t \) decides to join, then \( N_t^+ = N_t + 1 \), otherwise \( N_t^+ = N_t \). For any \( t \in [0, T] \), define \( \pi_k^1(t) \) as the time \( t \) probability that only the first deal happens given that \( N_t^+ = k \), i.e.,

\[
\pi_k^1(t) = \text{Pr}\{M_1 \leq N < M_2 | N_t^+ = k\}.
\]

Similarly define \( \pi_k^2(t) \) as the time \( t \) probability that the second deal happens given that \( N_t^+ = k \), i.e.,

\[
\pi_k^2(t) = \text{Pr}\{N \geq M_2 | N_t^+ = k\}.
\]

Note that the following boundary conditions hold:

(i) \( \pi_k^1(T) = 0 \) for \( 0 \leq k < M_1 \), and \( \pi_k^2(T) = 0 \) for \( 0 \leq k < M_2 \).

(ii) \( \pi_k^1(t) = 1 - \pi_k^2(t) \), for \( M_1 \leq k < M_2 \).

(iii) \( \pi_k^1(t) = 0, \pi_k^2(t) = 1 \) for \( k \geq M_2 \).

Finally, for \( k \geq 0 \), define \( H_k(t) \) as the probability that a consumer arriving at time \( t \) signing up, given that \( N_t = k \). Notice that for all \( t \in [0, T] \), and \( k \geq 0 \), \( H_k(t) = 1 \) if \( u_t \geq p_0 \), \( H_k(t) = 0 \) if \( u_t < p_2 \). Further, for \( k \geq M_2 - 1 \) the boundary condition \( H_k(t) = 1 - F(p_2) \) holds.

Now, consider the decision of a customer with valuation \( u \) who arrives at time \( t \). For \( N_t = k \geq 0 \) denote this customer’s expected utility of signing up by \( V_k(u, t) \). Then

\[
V_k(u, t) = \pi_k^1(t) \max\{u - p_1, -d\} + \pi_k^2(t) \max\{u - p_2, -d\} + \max\{u - p_0, -d\} (1 - \pi_k^1(t) - \pi_k^2(t)),
\]

and he would choose to sign up if and only if \( V_k(u, t) \geq 0 \). This implies

\[
H_k(t) = \text{Pr}\{V_k(u, t) \geq 0\},
\]

where the consumer’s utility of not joining is normalized to 0. Utilizing (1), we can derive the characterization of a consumer’s sign-up decision based on his arrival time, \( t \), and the number of sign ups up to that point \( N_t \). The following lemma states the structure of this decision.

**Lemma 1** For each \( t \in [0, T] \), given \( N_t = k \geq 0 \), there exists a threshold \( \bar{u}_{k,t} \in [p_2, p_0] \) such that a customer who arrives at time \( t \) with reservation utility \( u \) signs up if and only if \( u \geq \bar{u}_{k,t} \). \( \bar{u}_{k,t} \) is characterized as

\[
\bar{u}_{k,t} = \begin{cases} 
  p_0 - 2d - d \left( 1 - \frac{1}{\pi_{k+1}^1(t)} \right) & \text{if } 0 \leq d \leq \delta \pi_{k+1}^2(t), \\
  p_0 - d + \frac{d(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t))}{\pi_{k+1}^1(t) + \pi_{k+1}^2(t)} & \text{if } \delta \pi_{k+1}^2(t) < d \leq \delta \pi_{k+1}^1(t) + 2\delta \pi_{k+1}^2(t), \\
  p_0 - \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) & \text{if } d > \delta \pi_{k+1}^1(t) + 2\delta \pi_{k+1}^2(t).
\end{cases}
\]
Based on Lemma 1, we can now derive the consumer equilibrium by solving for the continuous time equilibrium evolution of the consumers’ sign up probability, $H_k(t)$, and the probabilities of the two deals happening, $\pi_1^k(t)$ and $\pi_2^k(t)$, respectively. Starting with $\pi_2^k$, given that at time $t \in [0,T]$, $N_t = k$, in order to calculate the time $t$ probability of second deal happening, we can condition on the arrival time of the next customer. For $x \in (0,T-t]$, suppose that the next customer arrives at time $t+x$. Then since the customer arrival process is Poisson with rate $\lambda_g$, the distribution of $x$ is the interarrival distribution for this process, i.e., Exponential with rate $\lambda_g$ and p.d.f. $\lambda_g e^{-\lambda_g x}$. At the time of his arrival, $t+x$, the next customer observes that $N_{t+x} = k$, and decides to join with probability $H_k(t+x)$. If he joins, then $N_{t+x}^+ = k+1$, and otherwise $N_{t+x}^+ = k$. Consequently, the time $t+x$ probability of the second deal happening is $\pi_2^{k+1}(t+x)$ if he joins, and $\pi_2^k(t+x)$ otherwise. We can then write the a dynamic recursive equation for the probability that the second deal will happen, $\pi_k^2(t)$ by taking the conditional expectation on the arrival and the decision of the next customer as

$$\pi_k^2(t) = \int_0^{T-t} (H_k(t+x)\pi_{k+1}^2(t+x) + (1-H_k(t+x))\pi_k^2(t+x) ) \lambda_g e^{-\lambda_g x} dx ,$$

with boundary conditions $\pi_k^2(t) = 1$ for $k \geq M_2$, and $H_k^2(t) = 1 - F(p_2)$ for $k \geq M_2 - 1$. In a similar manner, for $\pi_k^1$, we can obtain

$$\pi_k^1(t) = \int_0^{T-t} (H_k(t+x)\pi_{k+1}^1(t+x) + (1-H_k(t+x))\pi_k^1(t+x) ) \lambda_g e^{-\lambda_g x} dx ,$$

Figure 2: Consumer utility ranges for sign-up and staying decisions at time $t \in [0,T]$ and $k \geq 0$ existing sign ups, when $p_2 \leq \bar{u}_{k,t} < p_1$. 

Figure 2 shows the structure of consumer sign-up behavior on the utility axis $(u)$ for a customer who arrives at time $t$ for the case $p_2 \leq \bar{u}_{k,t} < p_1 - d$. As stated in Lemma 1, the customer will sign up if and only if $u \geq \bar{u}_{k,t}$. However, as we also discussed above, after the event, a customer who signs up will drop out if $u < p - d$. Therefore, as can also be seen in the figure, customers with utility values lower than $p_1 - d$ who signed up will drop out after the event if the first deal does not happen (i.e., $N < M_1$), and customers with utility values lower than $p_2 - d$ who signed up will drop out after the event if the second deal does not happen (i.e., $N < M_2$). The details of the consumer behavior when $\bar{u}_{t,k}$ falls into other intervals follow with similar logic.
Figure 3: Equilibrium $\pi^1_k$, $\pi^2_k$ and $H_k$ functions for varying number of existing sign ups. Panels (a) and (b) illustrate the probability of only the first deal materializing, and the probability of only the second deal materializing, $\pi^2_k(t)$, respectively; and panel (c) shows the consumer sign-up probability $H_k(t)$, for selected values of the existing number of consumer sign ups ($k$). For all panels, $T = 1$, consumer utility distribution is Uniform on $[0,1]$, $\lambda_g = 30$, $M_1 = 20$, $M_2 = 50$, $p_0 = 0.7$, $p_1 = 0.6$, and $p_2 = 0.5$, and $d = 0.02$.

with boundary conditions $\pi^1_k(t) = 0$ for $k \geq M_2$, and $\pi^1_k(t) = 1 - \pi^2_k(t)$ for $M_1 \leq k < M_2$. Solving the differential equation system that arises from (4) and (5) on $t \in [0,T]$ and $k \geq 0$ recursively with the above boundary conditions, we can obtain the consumer equilibrium characterized by $(\pi^1_k, \pi^2_k, H_k)$, $k \geq 0$. The following proposition states the result.

**Proposition 1** For any $\lambda_g, d > 0,$ and $0 < \delta < p_0/2$, there exists a unique dynamic consumer equilibrium. In equilibrium, a customer with utility $u$ arriving at time $t$ with $N_t = k$ signs up if and only if $u \geq \bar{u}_{k,t}$, where $\bar{u}_{k,t}$ is as defined in (3). The equilibrium is characterized by the solution to the dynamic recursive equation system

$$\pi^i_k(t) = \lambda_g \int_t^T e^{-\lambda_g \int_s^t H_{k+1}(v)dv} H_{k+1}(s) \pi^i_{k+1}(s) ds, \quad \text{for} \quad 0 \leq k < M_i, \ i = 1, 2,$$

with boundary conditions $\pi^1_k(t) = 0$, $\pi^2_k(t) = 1$, for $k \geq M_2$, $\pi^1_k(t) = 1 - \pi^2_k(t)$ for $M_1 \leq k < M_2$, and where

$$H_k(t) = 1 - F(\bar{u}_{k,t}), \quad \text{for} \quad k \geq 0. \quad (7)$$

Figure 3 demonstrates the consumer equilibrium outcome as a function of time $t$ for deal thresholds $M_1 = 20$ and $M_2 = 50$. As can be seen in Panel (b), for any given time point $t$, the probability of the second deal happening $\pi^2_k(t)$ is higher as the number of sign ups on $[0,t]$, i.e., $k$, becomes higher. Further, for any fixed number of sign ups, as the time progresses, the second deal probability decreases since less
time is left for the remaining $M_2 - k$ customers to sign up for the second deal to happen. The patterns for the probability that only the first deal happens, $\pi_k^1$, are more subtle as can be seen in Panel (a) of Figure 3. First, for a small number of existing sign ups, e.g., $k < 10$, for any given $t$, as the number of sign ups increases $\pi_k^1$ increases, since with a higher number of customers already in, it becomes more likely for the first deal to happen, and $\pi_k^1$ is monotonically decreasing in $t$. However, as $k$ approaches the first deal threshold, $M_1 = 20$, additional sign ups do not necessarily increase $\pi_k^1$, because they increase the probability of the second deal materializing, and hence reduce the probability that only the first deal materializes. Similarly, as can be seen for $k = 11$, $\pi_k^1$ is also no longer monotonic for intermediate $k$ values, since earlier in the time window, as the time progresses without any additional sign ups, the probability of the second deal threshold being crossed decreases and the overall probability of ending up with only the first deal increases. Finally, when $k \geq 20$, the number of sign ups already exceeds the first deal threshold, so any passing time without new arrivals will reduce the probability that the second deal materializes, and consequently $\pi_k^1(t)$ is monotonically increasing in $t$.

Panel (c) of Figure 3 shows the time evolution of equilibrium consumer sign-up probability $H_k$ for various existing sign-up levels, $k$. As can be seen from the figure, $H_k(t)$ curves are clustered in two groups, namely for $k < M_1 - 1 = 19$ and $k \geq 19$. For a given $k < 19$, $H_k(t)$ is decreasing in $t$ since having the same number of sign ups $k$ at a larger $t$ means that the probability of any of the two deals materializing is lower, and hence, signing up is less attractive for consumers, i.e., $H_k(t)$ decreases with $t$. For $k < 20$, as $t$ approaches 1, the probability that any of the deals will materialize vanishes, and almost no customer other than those with utilities greater than the base price, i.e., $u > p_0$, signs up. Hence, in the figure, the customer sign-up probability converges to $1 - F(p_0) = 1 - 0.7 = 0.3$. For $k \geq 19$, each arriving customer knows that the first deal has already materialized or will materialize if he chooses to sign up. Therefore, for $k \geq 19$, for any $t \in [0,1]$, any arriving customer with $u \geq p_1$ will sign up, and hence $H_k(t) \geq 1 - F(p_1) = 1 - 0.6 = 0.4$. For this larger $k$ value cluster, $H_k(t)$ is also decreasing in $t$ but converges to 0.4 as $t$ approaches 1.

Utilizing our theoretical analysis we have developed thus far, and data on customer sign ups during an event, we can structurally estimate the problem parameters such as consumer arrival rates and utility distributions for a group buying event. In the next section, we will utilize that structural estimation to empirically explore the effects of group buying on consumer sign-up behavior, retailer pricing and profits.

4 Empirical Analysis

In this section we aim to first estimate and verify the goodness of fit for the theoretical model we studied in Section 3, and use it to empirically demonstrate and measure demand increase associated with group buying events and the profit gains they provide to the retailer. Our data comes from the sales of a major
Chinese appliance retailer in 2013. The data includes 266 group buying events and 2715 cases of products sold through traditional single-pricing on Taobao’s retailing website Tmall.com. We give the detailed description of the data for group buying and single-price sales in Sections 4.1.1 and 4.2.1, respectively.

The outline of our strategy for empirical analysis is as follows: Utilizing our theoretical analysis and the continuous time consumer equilibrium expressions derived, and given group buying sign-up data from Taobao, we first perform a structural maximum likelihood estimation to jointly estimate (i) the consumer arrival rate (\( \lambda_g \)), and (ii) the consumer utility distribution function (\( F \)) and its parameters for each group buying event. Based on this estimation framework, we then test the model fit and assumptions, and check for the predictive power of our model. Next, utilizing our estimates for group buying events and an extended data set that includes observations on traditional single-price sales, we estimate the increase in demand brought about by group buying, and based on this estimate, perform a counterfactual analysis to determine the profit gains from employing group buying over a single-price strategy. Finally, using the estimation results, we empirically demonstrate recommendable pricing patterns that help improve retailer profits.

4.1 Estimation of Group Buying Parameters

4.1.1 Data Description for Group Buying Events

Our data includes detailed information from 266 group buying events held on November 11 (proclaimed as “Singles’ Day” in China) 2013. In each group buying event, a unique product is sold through group buying via the mechanism we had described in Section 3. For all events, the first and second sign-up thresholds are \( M_1 = 20 \) and \( M_2 = 50 \) respectively, and the required customer deposit for signing up during the event is \( d = 99 \) CNY. The data for each event includes the product identifier, the three prices, \( p_0 \), \( p_1 \), and \( p_2 \), time length of each event (11 or 12 hours), time for each sign up during the event (hour, minute, and second), and the number of sign ups who choose to stay after the event. In total, we have 41,496 sign-up data observations. In 217 of the 266 events in the data set, the second deal threshold is reached (i.e., \( N \geq 50 \)), in 42 events, the first deal threshold is reached but not the second (\( 20 \leq N < 50 \)), and in the remaining 7 events no deal threshold is reached (\( N < 20 \)). The products sold in the events belong to six major categories: Refrigerators (44 events), Air Conditioners (39), Television Sets (63), Water Heaters (32), Gas Stoves (27), and Washing Machines (61). Descriptive statistics for group buying event data is given in Table B.1 in the Online Supplement.
4.1.2 Estimation of Model Parameters

We start with the estimation of model parameters. For each event and at any given time point \( t \in [0, T] \), given the base price and the discount chosen by the retailer \((p_0, \delta)\), and the number of arrivals, \( k \geq 0 \), up to that point, the instantaneous sign-up rate for the next customer is

\[
\lambda_{k+1}(t) = \lambda g H_k(t),
\]

where \( H_k(t) \) is as defined in Proposition 1, and since the consumer arrival process is Poisson, conditional on the number of existing sign ups \( k \), the sign-up rate for the \( k + 1 \)st customer is independent of the history of the process on \([0, t]\). Therefore, for each \( t \) and \( k \), the next sign-up follows a process that is distributionally equivalent to the first sign-up of a non-homogenous Poisson process with instantaneous arrival rate as given in (8), and at time \( t \), with \( k \) existing sign ups, the appearance time for the \( k + 1 \)st sign-up is exponentially distributed with density

\[
\varphi_k(t, s|H) \triangleq \int_s^t \lambda_k(\tau) d\tau \cdot e^{-\int_s^t \lambda_k(\tau) d\tau} \cdot e^{-\int_s^t \lambda g H_{k-1}(\tau) d\tau}.
\]

In order to estimate the consumer utility distribution, we will determine the best parameter fit for a variety of family of distributions, namely Uniform, Normal and Log-Normal, Beta, Gamma, Weibull and Gumbel. Let \( \xi \) be the parameter vector for the consumer utility distribution for a given type of distribution. For instance, for Normal distribution \( \xi \) will be \((\mu, \sigma)\), i.e., the mean and standard deviation of the distribution. For each distribution type, we will find the best fitting parameter vector \( \theta = (\lambda, \xi) \) through Maximum Likelihood Estimation. For a given event, let \((t_1, t_2, ..., t_N)\) be the sign-up times observed in the data. In order to perform the estimation, for any candidate parameter vector \( \theta \), we first calculate the consumer equilibrium outcome \((\pi^1, \pi^2, H)\) utilizing Proposition 1. Denote the equilibrium consumer sign-up probability function sequence \( H : \mathbb{N} \times [0, T] \to [0, 1] \), derived for the parameter vector \( \theta \) as \( H^\theta \). For instance, again for Normally distributed consumer utility, we have \( \theta = (\lambda, \mu, \sigma) \) and for \( k \geq 0 \) and \( t \in [0, T] \),

\[
H_k^\theta(t) = Pr\{u \geq \bar{u}_{k,t}\} = \frac{1}{2} \left( 1 - \text{erf} \left( \frac{\bar{u}_{k,t} - \mu}{\sigma \sqrt{2}} \right) \right),
\]

where \( \bar{u}_{k,t} \) is as given in Proposition 1 for the parameter vector \( \theta \). Then we can write the likelihood function as

\[
\mathcal{L}(\theta; t_1, t_2, ..., t_N) = \prod_{k=1}^{N} \varphi_{k-1}(t_{k-1}, t_k|H^\theta),
\]
Table 1: Group Buying Parameter Estimation Results

<table>
<thead>
<tr>
<th>Arrival Rate ($\lambda_g$)</th>
<th>No. of Events</th>
<th>Mean (CNY)</th>
<th>Standard Deviation (CNY)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Min</td>
<td>Average</td>
</tr>
<tr>
<td></td>
<td>266</td>
<td>1.86</td>
<td>16.75</td>
</tr>
<tr>
<td>Utility Distribution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta</td>
<td>169</td>
<td>1002.42</td>
<td>3616.79</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>49</td>
<td>892.35</td>
<td>3392.41</td>
</tr>
<tr>
<td>Normal</td>
<td>48</td>
<td>1311.9</td>
<td>3587.62</td>
</tr>
<tr>
<td>McFadden Pseudo $R^2$</td>
<td></td>
<td>0.35</td>
<td>0.64</td>
</tr>
</tbody>
</table>

where $t_0 = 0$. The maximum likelihood estimate for the given event and distribution type then is

$$
\theta^* = \arg\max_{\theta} \mathcal{L}(\theta; t_1, t_2, ..., t_N) = \arg\max_{\theta=(\lambda_g, \xi)} \prod_{k=1}^{N} \left\{ \int_{t_{k-1}}^{t_k} \lambda_g H_{k-1}^{\theta}(\tau) d\tau \cdot e^{-\int_{t_{k-1}}^{t_k} \lambda_g H_{k-1}^{\theta}(\tau) d\tau} \right\}. \quad (12)
$$

For each distribution type, at each iteration of the estimation, for the corresponding parameter vector $\theta$, the equilibrium sign-up probability function sequence $H^{\theta}$ is calculated by solving the dynamic equation system given in (6)-(7), for $0 \leq k \leq N$. Subsequently, the objective (12) is calculated. Iterations continue until convergence to the optimal parameter vector $\theta = (\lambda_g, \xi)$. Repeating the process for all the distribution types we listed above, and choosing the one that yields the highest likelihood score, we can determine the best fitting one among the considered distributions with its parameters and the corresponding best estimate for the arrival rate for each event.

Table 1 presents the results of the estimation. Category-based breakdown of these estimation results are given in Section C in the Online Supplement. As can be seen from Table 1, for approximately two thirds of the 266 events in our data set, i.e., for 169 events, Beta distribution is the best fit for the consumer utility distribution. The rest of the events are split approximately evenly between Log-Normal and Normal distributions as the best fit. The distributional estimates indicate that average unit reservation value for customers is about 3,500 CNY (approximately 500 USD) with a standard deviation at nearly one tenth of the mean. We can also observe from the table that the estimated arrival rate of customers vary significantly across events ranging from less than two to more than thirty per hour. Similarly, the estimated means of the consumer utility distributions have significant variability, ranging from about 1,000 to approximately 7,000 CNY. However, standard deviations for estimated consumer utility distributions show less variation.

4.1.3 Model Fit and Predictive Power

We next discuss the goodness of fit and predictive power of our model, demonstrating that it provides a very good approximation for the consumer equilibrium behavior that emerges in group buying events.
This is important, since we will be using customer arrival rate estimates ($\lambda_g$) from our model to measure the impact of employing group buying in increasing customer demand in Section 4.2.

**Testing the Model Fit and Distributional Assumptions**

As Table 1 presents the average McFadden Pseudo-$R^2$ value of the model parameter estimation for the 266 events is 64%, with a minimum of 35%, and can be as high as 89%. Further, 224 out of 266 events (84.21%) have Pseudo-$R^2$ values greater than 50%, indicating that the model fit to the data, in general, is very good, supporting the fit of the model’s assumptions and the projected equilibrium behavior with the underlying process. We can further test the fit of the Poisson inter-arrival time distributional assumption of the model. For this, we need to take into account the modulation of the arrival process with equilibrium consumer sign-up process. Recall from the discussion in Section 4.1 that conditional on the previous consumer sign-up, the time between each pair of sign-ups has a distribution, equivalent to the first sign-up of a non-homogenous Poisson process with instantaneous arrival rate, given in (8), and hence, the interarrival times for sign-ups are exponential with density given in equation (9). Therefore, given the sign-up time vector ($t_1, t_2, \ldots, t_N$) for event $k$, $1 \leq k \leq N$,

$$
\frac{t_k - t_{k-1}}{\int_{t_{k-1}}^{t_k} \lambda_t H_{k-1}(\tau)d\tau} \cdot e^{-\int_{t_{k-1}}^{t_k} \lambda_t H_{k-1}(\tau)d\tau}
$$

has an exponential distribution with mean 1, where $t_0 = 0$, and $H_k(t)$, $k = 0, \ldots, N$ is as defined in Proposition 1. For a given event, we can then test the distributional fitting of the observed data with the Poisson arrival rates implied by the consumer sign-up process from the model, by applying a Kolmogorov-Smirnov distributional fit test to the statistic vector given in (13) for each event. Running this test for each of the 266 group buying events in the data set, we find that 257 of the events (96.6%) pass the test, implying that Poisson inter-arrival time distributional assumption of the model is widely supported.

**Testing the Predictive Power of the Model**

Next, we will test how good our model is in predicting the evolution of the consumers’ reaction to crossing deal thresholds. In particular, one deviation from our model set up can be that some customers may wait for the price reduction thresholds to be crossed and lower prices to be guaranteed before signing up and paying the deposit. If that is the case, we should see a significant increase in sign ups after the second threshold is crossed compared to the number that would be predicted by the pre-threshold crossing sign-up observations. On the other hand, if our model is a good approximation for the consumer sign up behavior, the model estimated sign-up behavior before the second threshold is crossed should have strong power in
predicting the number of consumers, who will sign up after this threshold is crossed.

In order to test this, first let us define two threshold times $\tau_1 = \inf\{t | N^+_t \geq M_1 - 1, 0 \leq t \leq T\}$, and $\tau_2 = \inf\{t | N^+_t \geq M_2 - 1, 0 \leq t \leq T\}$. That is, $\tau_1$ and $\tau_2$ are the time points on the event window $[0, T]$, when the first and second deal sign-up thresholds are crossed with one more sign up, provided that the corresponding thresholds are crossed during the event window. For the purposes of this test, we will focus on the 217 out of the 266 events that are in our data set, where there were sufficient sign ups that the second deal materialized, i.e., where $N \geq M_2$.\(^1\) In each of these events, according to our model, the sign up process for $t > \tau_2$ would be a Poisson process with arrival rate $\lambda_g(1 - F(p_2))$, and hence the expected number of sign ups on $(\tau_2, T]$ is

$$\nu(\lambda, F, \tau_2) = \mathbb{E}[N_T - N_{\tau_2}] = \lambda_g(1 - F(p_2))(T - \tau_2). \quad (14)$$

Therefore, we can test our model’s power of predicting the number of arrivals after $\tau_2$ as follows: First, for each event $i$, $1 \leq i \leq n$, using Maximum Likelihood Estimation as described in Section 4.1.2 on $[0, \tau_{2i}]$, we can estimate $\hat{\lambda}_g$, and $\hat{F}_i$. That is,

$$(\hat{\lambda}_g, \hat{F}) = \arg\max_{\lambda_g, F} L(\lambda_g, F; t_1, t_2, ..., t_{M_2 - 1}), \quad (15)$$

where $L$ is as described in Section 4.1.2. Then, we can calculate $\nu(\hat{\lambda}_g, \hat{F}_i, \tau_{2i})$, $1 \leq i \leq n$ as defined in (14), and comparing to the actually observed $N_{\tau_2} - N_{\tau_2}$, perform a t-test to determine whether the model’s predictions are in line with the data or can be rejected.

Figure 4, graphically illustrates the methodology and comparison. It presents the timeline of a single event from the data set, illustrating the evolution of the sign ups, $N_t$, and the rate of sign ups after $\tau_2$ as projected by our model and estimation, given by

$$Y(t) = N_{\tau_2} + \nu(\hat{\lambda}_g, \hat{F}, \tau_2) \frac{t - \tau_2}{T - \tau_2}. \quad (16)$$

That is, based on the estimation of the arrival rate and the customer utility distribution on the sign-up data on $[0, \tau_2]$, $Y(t)$ is the expected number of sign ups for each $t > \tau_2$ in our model. The estimated hourly customer arrival rate based on the observations on $[0, \tau_2]$ for this event is $\hat{\lambda}_g = 19.40$, and the estimated

---

\(^1\)For this test, we are focusing on the second threshold rather than the first threshold, and hence setting our “training” data set to $[0, \tau_2]$ and not $[0, \tau_1]$, for several reasons. First, for each event, before the first threshold is crossed, there are only 20 observations, which is a limited amount of data to construct a reliable estimation from. Second, the estimation “learns” from our model’s dynamic evolution of consumer equilibrium on the entirety of $[0, \tau_2]$, until the second threshold is crossed. After that, there is no longer any consumer strategic sign-up behavior, as the customers sign up solely based on their valuations of the product. Limiting the estimation to $[0, \tau_1]$ would mean limiting the training data set to a subset of available training information, and throwing away all the observations on $[\tau_1, \tau_2]$, that affect the estimation.
Figure 4: Evolution of cumulative customer sign ups and model projection for a sample event, for which the second deal threshold is reached. The event is for a refrigerator with capacity 451 liters. $N_t$ represents the cumulative number of sign ups, and $Y(t)$ is the model projected linear growth trend for $t > \tau_2$ based on consumer sign ups on $[0, \tau_2]$. For the event, the base price, $p_0$, is 2,999 CNY, and the price discount, $\delta$, is 200 CNY.

Table 2: Test results for the Predictive Power of the Model

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Mean $E[N_T - N_{\tau_2}[t_1, t_2, \ldots, t_{M_2-1}]$</th>
<th>Mean observed $N_T - N_{\tau_2}$</th>
<th>Mean difference</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>82.16</td>
<td>82.60</td>
<td>-0.44</td>
<td>0.6875</td>
</tr>
</tbody>
</table>

customer utility distribution is Normal with mean 3155.64 and standard deviation 257.81. Therefore, the estimated hourly rate of sign ups after $\tau_2$ in this case is $\hat{\lambda}_g(1 - F(p_2)) = 12.79$, and the total estimated number of sign ups on $(\tau_2, T]$ is $\nu(\hat{\lambda}, \hat{F}, \tau_2) = 79.27$. In comparison, the actual number of sign ups on $(\tau_2, T]$ is 77, and as the figure also illustrates, the model’s prediction is very close to the actual observed number of sign ups. Repeating the process for all 217 events in our data set, for which the second deal threshold was reached, we can perform the test for the strength of the model’s predictive power. Table 2 summarizes the test results for the 217 events where the second deal materialized. The mean sign ups predicted by model is 82.16, which is very close to the mean observed number of sign ups, 82.60, and the difference is not statistically significant.

Overall, we can conclude that our model provides a good fit and approximation for the consumer behavior resulting in predictions closely fitting with the actual data. The data suggests that if there is any consumer behavior not captured in our model (e.g., threshold waiting behavior), its effect on the equilibrium outcome is small and statistically insignificant. In the rest of the paper, we will use our model’s parameter estimates to study consumer network effects and profit gains from employing group buying events.
4.2 Demand Effects and Gains from Group Buying

In this section, we measure the retailer’s demand and profit gains from employing group buying. Utilizing our customer arrival rate estimations from Section 4.1 together with data from products sold through traditional single-price mechanisms, we first estimate the magnitude of the demand increase due to network effects in group buying events. Then, through a counterfactual analysis, we estimate the retailer profits had the retailer sold the product via traditional single-price events, and study the combined effects of group buying discounts and increased demand due to customer networking to determine the net gains on retailer profits brought about by employing group buying.

4.2.1 Description for the Extended Data Set

We start by describing the extended data set we use for our analysis of demand effects, which, in addition to the data on 266 Group Buying Events we described in Section 4.1.1 covers 2715 observations of sales numbers for products sold under traditional single-price sales over 6 days. First, our data includes information from 260 products from all six categories that were sold through single-price sales on November 11 2013. In addition, for those 260 products and 231 of the 266 products that were sold through group buying on November 11, we have information for traditional single-price sales for Christmas eve and the four following days (December 24-28) in 2013. For ease of exposition, we call each one of these 2715 observations as a *single-price event*. For each single-price event, we have the duration (12 hours), the price, and the number of units sold during the period.

In addition to the sales information, for each of the 2981 events (group buying and single-price), we have the average review score for the product sold in that event posted on Taobao at the end of the day of the event. The review score is an important factor reflecting how satisfied consumers feel about the products, thereby measuring the consumer perception of the quality of the product at the time of the event. The review scores are on a scale of 0 to 5, with 0 being the lowest and 5 being the highest satisfaction score. Finally, in order to estimate customer utility for products, we also have technical characteristics of all the products in our data sets. Category-specific value determinants we use are capacity for Refrigerators; capacity and energy level for Air Conditioners; screen size and display resolution for Television Sets; capacity and power for Water Heaters; power and the number of panels for Gas Stoves; and capacity and energy level for Washing Machines. We obtain the corresponding data for all 491 products in our data set from the product design department of the retailer and through online public sources. Descriptive statistics for the extended data set is given in Tables B.1 and B.2 in the Online Supplement.
4.2.2 Quantifying the Increase in Customer Demand for Group Buying

Utilizing the extended data set, we can now estimate the determinants of the consumer arrival rate for purchasing a product in both group buying and single-price events. For this, we employ a logarithmic doubly stochastic arrival process model, or a Generalized Linear Model (GLM) with log-link function (Nelder and Baker 1972), which we will estimate. Specifically, the arrival process for a given event \( i \) is Poisson with rate \( \lambda_i \), where \( \lambda_i \) is also stochastic, and for the corresponding regressor variable vector \( \mathbf{X}_i \in \mathbb{R}^n \),

\[
E[\lambda_i|\mathbf{X}_i] = e^{\beta'\mathbf{X}_i},
\]

(17)

where \( \beta \in \mathbb{R}^n \) is a vector of coefficients. For our estimation, we employ a doubly Poisson arrival process, meaning that the distribution of \( \lambda_i \) conditional on \( \mathbf{X}_i \) is generalized Poisson with mean as given in (17) (Cameron and Trivedi 1998).\(^2\) That is, defining \( h_i \) as the p.d.f. of \( \lambda_i \),

\[
h_i(x) = \frac{e^{(x\beta'\mathbf{X}_i - e^{\beta'\mathbf{X}_i})}}{\int_0^\infty z^{x-1}e^{-z}dz}, \quad x > 0.
\]

(18)

The determinants of the magnitude of the potential consumer demand for an event include various factors, such as the characteristics of the product that is sold during the event, e.g., product category and quality, in addition to the date of the event and whether group buying was employed in the event, and if so, the discount offered. We include these factors in our regressor vector, and in order to separate the date and category effects, we also include date and category interaction terms. In particular, taking the logarithm of both sides of (17), we have

\[
\log(E[\lambda_i|\mathbf{X}_i]) = \beta'\mathbf{X}_i = \beta_1\frac{\delta_i}{p_{si}^*} + \sum_{j=1}^6 I_{ij}(\beta_{2j} + \beta_{3j}R_i + \beta_{4j}T_i) + \sum_{t=1}^5 \beta_{5jt}D_{it} + \beta_{6j}D_{i1}T_i.
\]

(19)

In (19), \( \delta_i/p_{si}^* \) is the deal discount for a given event, normalized by the estimated optimal single-price corresponding to an event, \( p_{si}^* \) (please see below for further details on the derivation of \( p_{si}^* \)). \( I_{ij} \) is a dummy variable that indicates whether product sold in event \( i \) was in category \( j \). \( R_i \) is the average consumer review score for the product sold in event \( i \). \( T_i \) (or Treatment indicator) is a dummy variable that indicates if the product in event \( i \) is one of those that were put on for sale group buying on November 11, 2013. Note that if a product was put up for a group buying event on Singles’ Day 2013, if later on that product is sold in a single-price event \( i \), \( T_i \) will still be one, i.e., this variable controls for the effect of being put on a Group Buying event for the potential demand for the product. Finally, \( D_{it} \) is the day indicator for the event. In

\(^2\)To check the robustness of the Poisson distribution of the arrival rates, we also performed the estimation under other distributional assumptions such as Beta, Log-normal, and Normal. In addition to being the natural conjugate for the log-link function (Coxe et al. 2009) Poisson distribution gives the best fit for the model among these distributional alternatives.
particular, if event $i$ was held on Day $t$, then $D_{it} = 1$, otherwise $D_{it} = 0$, where $t = 1$ for Single’s Day (November 11), and $t = 2, \ldots, 5$ correspond to the four days after the Christmas Eve, December 25-28, 2013, respectively, i.e., Christmas Eve is the base for the date category.

For category $j = 1, \ldots, 6$, $\beta_{4j}$ measures the difference in logarithm of the expected potential demand between the products that were chosen to be sold by group buying events and those that were never included in group buying events, and $\beta_{5j}$ measure the date effects, reflecting the increase in logarithm of the potential demand for day $t \in \{1, \ldots, 5\}$ compared to Christmas Eve. Importantly, the coefficients $\beta_{6j}$, $j = 1, \ldots, 6$, capture the consumer network effects associated with group buying events. Specifically, for an event $i$, and category $j$, $I_{ij}D_{i1}T_i = 1$ if and only if product sold in the event is in category $j$ and was one of the products that were included in a group buying event on November 11 (i.e., $T_i = 1$), and the date of the event was November 11, i.e., if and only if the event was a group buying event (note that if the same product was sold on any other day in the data set, $T_i$ would still be one but $D_{i1}$ would be zero, making the whole term zero). As such, the coefficient $\beta_{6j}$ captures the increase in the logarithm of the expected arrival rate for an event in category $j$ due to the event being a group buying event.

The GLM estimation proceeds with a Maximum Likelihood Estimation of (19) by fitting it to the observed $\lambda_i$ values under Generalized Poisson Distribution. Note that $\lambda_i$ in (19) captures the magnitude of the potential consumer demand. Consumers who are arriving with rate $\lambda_i$ make purchasing decisions based on the pricing of the product. For the 266 group buying events, the $\lambda_{gi}$ calculated in Section 4.1.2 correspond to this potential demand arrival rate. For the events that were priced through traditional single-pricing, we have to calculate the corresponding rates from the observed sign up numbers and other information we have.

**Estimation of Consumer Arrival Rates and Utility Distributions for Single-Price Events**

As we mentioned above, for each single-price event $i$ in the data set, we have the number of purchases, $N_{oi}$, for that event over a 12 hour period (i.e., $T = 12$). Note that given the c.d.f., $F_i$ of the consumer utility distribution for the product in event $i$, and the single-price for the event $p_{ci}$, the consumer purchase process in event $i$ is Poisson with rate $\lambda_i(1 - F_i(p_{ci}))$. Using this information, for event $i$, the Maximum Likelihood Estimate for the customer arrival rate $\lambda_i$ for a single-price event is

$$\lambda_i = \frac{N_{oi}}{T(1 - F_i(p_{ci}))}. \quad (20)$$

Therefore, in order to estimate the customer arrival rate for a product sold through a single-price event, we need to first estimate the customer utility distribution for that product. If the product sold in a single-price event

---

3Under the Poisson sign-up rate $\Lambda$, the probability of $N$ sign ups on $[0, T]$ is $((\Lambda T)^N e^{-\Lambda T})/N!$, which is unimodal in $\Lambda$, and the first order condition yields the global maximum $\Lambda = N/T$. 

20
event on December 24-28 is the same as one of those products sold under group-buying, we already have
the consumer utility distribution estimate for the product calculated in Section 4.1.2. For those products
that do not appear in any of the 266 group buying events in our data set, we use the existing utility
distribution estimates for each product category and product characteristics to estimate the parameters of
the utility distribution. Note that as presented in Table 1, best distribution estimates for all 266 products
are Beta, Log-normal, or Normal distributions. All three of these distribution types are two parameter
distributions and can be uniquely identified by their mean and standard deviations. Therefore, using the
266 estimated utility distributions as our training data, we first estimate the determinants of the mean
and standard deviation of consumer utility in each category.

Depending on the product category, we use a specific set of factors that affect the value of a product
for consumers in that category we listed in Section 4.2.1. For example, the Television Sets are generally
homogeneous and their value for the customers are mainly determined by the screen size and display
resolution. Specifically, for each category $j = 1, \ldots, 6$, using the products in that category for which the
estimates of the mean, $\{\hat{\mu}_{gij}\}$, and standard deviation, $\{\hat{\sigma}_{gij}\}$, of consumer utility distribution exist from
our analysis in Section 4.1.2, and given the category-specific factor matrix, denoted as $Z_{gj}$, we run the set
of regressions

$$\log(\hat{\mu}_{gij}) = \alpha_0 + \alpha_1 Z_{gij} + \epsilon_{ij}^\mu,$$

$$\log(\hat{\sigma}_{gij}) = \gamma_0 + \gamma_1 Z_{gij} + \epsilon_{ij}^\sigma,$$

where $Z_{gij}$ is the row of $Z_{gj}$ that corresponds to product $i$ in category $j$, and $\epsilon_{ij}^\mu$ and $\epsilon_{ij}^\sigma$ are the affiliated
error terms. We choose a logarithmic regression structure as it is commonly used in estimating consumer
utility, and since our robustness checks with other regression structures show that the fit of the logarithmic
regression is the best. The detailed estimation results are given in Table D.1 in the Online Supplement.

The fit of the model is very good for the distribution mean regressions, as the F-Ratio for each model is
highly significant, with p-values for the F-test less than 0.01%, indicating the validity of the model, and the
adjusted $R^2$ values have a minimum of 0.40 and reach as high as 0.92. As could be expected, the fit of the
model for the estimation of standard deviation is somewhat worse, but still good with adjusted $R^2$ values
ranging from 0.21 to 0.71, and again the F-ratio for each model is highly significant, with corresponding
p-values all less than 1%.

Taking the regression results estimated from the training data-set, we can then estimate the consumer
utility distribution for the products that were only sold through single-price events. Specifically, for product
$i$ of category $j$ in this group, denote the factor vector by $Z_{cij}$. The estimated mean and standard deviation,
\( \hat{\mu}_{cij} \) and \( \hat{\sigma}_{cij} \) for the utility distribution for this product can then be calculated as

\[
\hat{\mu}_{cij} = e^{\alpha_0 + \alpha_1 Z_{cij}} \quad \text{and} \quad \hat{\sigma}_{cij} = e^{\gamma_0 + \gamma_1 Z_{cij}}.
\]  

(23)

For the consumer utility distribution for these products, we use the Beta distribution with mean and standard deviation as specified in (23), since in our utility estimations in Section 4.1.2, as given in Table 1, for approximately two-thirds of product utility distributions, a Beta distribution is the best fit.\(^4\) For the product in single-pricing event \( i \), calculating \( 1 - F_i(p_{ci}) \) using these estimated consumer utility distributions and the product price in that single-price event, \( p_{ci} \), we can then estimate the consumer arrival rate for event \( i \), using (20).

**Determining the optimal single price for each product**

As we mentioned above, we aim to control for the size of the deal discount when determining the network effects, as we expect larger deal discounts to have greater effect on average in boosting demand through network effects. However, each product is on a different price scale, and in order to have a fair comparison among deal discounts, we need to normalize the deal discounts across products to obtain the relative magnitude of the deal discounts to the product’s “value” or a benchmark price. For this we will use the estimated optimal single price for each product.

For a given product with consumer arrival rate \( \lambda \), and the marginal production cost \( c \), the optimal traditional single-price solution is

\[
p^*_s = \text{argmax}_{p \geq 0} (p - c)\lambda T(1 - F(p)) = \text{argmax}_{p \geq 0} (p - c)(1 - F(p)).
\]  

(24)

That is, \( p^*_s \) is the expected profit maximizing price for the retailer, if she simply sets a single price, \( p \), with each arriving customer with reservation price \( u \) purchasing the good when \( u \geq p \). For convenience in expression, we will refer to \( p^*_s \) as the *single price optimum* in the rest of the paper. Denote the corresponding optimal profit \( (p^*_s - c)\lambda T(1 - F(p^*_s)) \) by \( \Pi^*_s \).

For product \( i \), denote the marginal unit production cost as \( c_i \). Then utilizing the customer utility distribution estimated in Section 4.1, \( F_i \), and by (24), we can calculate the retailer single-price profit maximizer, \( p^*_{si} \), as

\[
p^*_{si} = \text{argmax}_{p \geq 0} (p - c_i)(1 - F_i(p)).
\]

For this, we need to have an estimate of the unit production cost, \( c_i \). In the data set, we do not have the unit production cost at the product level. However, for each of the six categories of products in the data set, we have the average margin, based on the group buying event final

\(^4\)To perform a robustness check on this assumption, we have also performed our regression analysis with Log-Normal and Normal Distributions. The results show that our analysis is again very robust to this assumption, with all our results being preserved and with only negligible changes in regression coefficients.
price.\textsuperscript{5} Thus, utilizing these percentage margins and the group buying event final price for each product, we can calculate the estimated production cost, $c_i$ for each product, and subsequently, $p^{*}_{si}$ as above.

**Estimation Results and Measuring Demand Increase from Group Buying**

Finally, by using the estimated consumer arrival rates for all events, we can perform the GLM estimation with the specification given in (19). The estimation results are given in Table 3. Note that there are 55 independent variables on the right hand side of the regression equation (19). For conciseness, the date effects for December 25-28 are omitted in the table in order to focus the variables most relevant to the determination of the Group Buying effects. For most date-category combinations for December 25-28, the estimated regression coefficients are not significant, indicating no detectable consumer demand difference between those dates and Christmas Eve.

The regression fit is very good with Adjusted $R^2$ value 0.6524. Further the $p$-value of the Pearson chi-square test on the null that the underlying model is correct is 0.2678, implying that the model fit stands (Consul and Famoye 1992). As can be seen from the table, the consumer review score has a significant and positive effect on the estimated consumer arrival rate for all categories, other than Washing Machines, for which the effect is not significant. The product being one of those selected to be sold by group buying on November 11th (i.e., treatment) has a mixed effect on the demand depending on the category. This is because in some cases the retailer sells products that are not very popular through group buying to boost their sales, i.e., being in the treatment group may be an indicator of inherent low demand. In other cases, the retailer chooses more popular products to include in group buying to promote its own brand. In either case, this variable controls for and separates product popularity from group buying effects we are aiming to extract. As can also be seen from the table, the positive effect of being sold on Single’s Day in increasing consumer demand is strong and highly significant across all categories. In particular, compared to Christmas Eve, the estimated average demand increases from approximately $8.82\% (=e^{0.0845} - 1)$, for Television Sets, to $25.72\% (=e^{0.2289} - 1)$, for Refrigerators.

Having controlled for the above factors, we can now observe the net effects of group buying on consumer demand. First, for each category, the group buying effects are positive and highly significant, providing evidence that group buying events increase the consumer interest in a product by generating a special promotional effect, or giving incentives to customers to spread the information about the event and recruit other potential customers. According to the regression results, the additional constant demand boost with employment of group buying ranges from around 7% to more than 15%. In addition, the coefficient of the normalized group buying deal discount, $\delta/p^{*}_{si}$ is positive and significant at the 5% level, indicating that the higher the deal discount relative to the product’s optimal monopoly price, the higher the increase in

\textsuperscript{5}The corresponding margin for each category is given in Table 5 below.
the expected consumer demand. This is because, in general, the higher the deal discount, the more the customers would have incentives to spend effort to act as voluntary sales agents and recruit other customers to join the event.

We further perform a robustness check of the double Poisson process assumption by calculating heteroskedasticity adjusted robust standard errors and coefficient estimates (Cameron and Trivedi 2009). The results are given in Table 4. As can be seen from the table, all constant group buying effects as well as the deal discount size effect are still significant, providing support for the robustness of our findings under the model assumptions. Finally, by using the discount-to-monopoly price ratio \(\frac{\delta}{p^*}\) for each event, we can calculate the total estimated demand boosting effect from group buying. As can be seen from Table 4, the average normalized deal discount in each category ranges from 9.32% to 11.88% of the estimated

<table>
<thead>
<tr>
<th>Table 3: Results for the Demand Effects Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td><strong>(\frac{\delta}{p^*})</strong></td>
</tr>
<tr>
<td>Refrigerators</td>
</tr>
<tr>
<td>Air Conditioners</td>
</tr>
<tr>
<td>Television Sets</td>
</tr>
<tr>
<td>Water Heaters</td>
</tr>
<tr>
<td>Gas Stoves</td>
</tr>
<tr>
<td>Washing Machines</td>
</tr>
<tr>
<td>Review Score: Refrigerators</td>
</tr>
<tr>
<td>Review Score: Air Conditioners</td>
</tr>
<tr>
<td>Review Score: Television Sets</td>
</tr>
<tr>
<td>Review Score: Water Heaters</td>
</tr>
<tr>
<td>Review Score: Gas Stoves</td>
</tr>
<tr>
<td>Review Score: Washing Machines</td>
</tr>
<tr>
<td>Treatment: Refrigerators</td>
</tr>
<tr>
<td>Treatment: Air Conditioners</td>
</tr>
<tr>
<td>Treatment: Television Sets</td>
</tr>
<tr>
<td>Treatment: Gas Stoves</td>
</tr>
<tr>
<td>Treatment: Washing Machines</td>
</tr>
<tr>
<td>Single’s Day: Refrigerators</td>
</tr>
<tr>
<td>Single’s Day: Air Conditioners</td>
</tr>
<tr>
<td>Single’s Day: Television Sets</td>
</tr>
<tr>
<td>Single’s Day: Water Heaters</td>
</tr>
<tr>
<td>Single’s Day: Gas Stoves</td>
</tr>
<tr>
<td>Single’s Day: Washing Machines</td>
</tr>
<tr>
<td>Group Buying: Refrigerators</td>
</tr>
<tr>
<td>Group Buying: Air Conditioners</td>
</tr>
<tr>
<td>Group Buying: Television Sets</td>
</tr>
<tr>
<td>Group Buying: Water Heaters</td>
</tr>
<tr>
<td>Group Buying: Gas Stoves</td>
</tr>
<tr>
<td>Group Buying: Washing Machines</td>
</tr>
</tbody>
</table>

Observations: 2946, Pearson \(\chi^2\): 2938.7, p-value: 0.2678
Adjusted \(R^2\): 0.6524

\*p < 0.1, \*\*p < 0.05, \*\*\*p < 0.01
optimal monopoly price. For each event $i$ in category $j = 1, \ldots, 6$, calculating $e^{\beta_1j(\delta_i/p_s^*)+\beta_6j} - 1$, we can evaluate the total percentage increase in customer demand through group buying for each event $i$ using the robust estimates. The estimated average demand increase ranges from 9.97% to 19.08% across the six categories, and overall average demand increase due to group buying network effects is 15.40%. In sum, we can conclude that there is evidence of significant network effects in increasing consumer demand through group buying events.

### 4.2.3 Counterfactual Analysis for Profit Gains through Group Buying

Having determined the percentage increase in the consumer arrival rate coming from employment of group buying, we can estimate the single-price event customer arrival rates for the 266 products in the data set to perform a counterfactual analysis to estimate the base customer arrival rates for these products and calculate expected optimal retailer profit if they had been sold on November 11 through single-price events instead of group buying events. Then, we can calculate the realized profits from the group buying events, and estimate the net percentage gains in retailer profits from employing group buying.

For a product $i$ that is in category $j$, the estimated consumer arrival rate if the product were sold through single price instead of group buying, $\lambda_{oi}$ can be calculated by taking out the group buying effects in increasing the consumer arrival rate utilizing the estimated regression equation (19), which implies $\lambda_{oi} = \lambda_{oi}e^{-(\beta_6j+\beta_1j(\delta_i/p_s^*))}$. In order to calculate the profits for the counterfactual scenario for selling a product at a single price, recall that in Section 4.2.2, we obtained the estimated single-price optimum $p_s^*$ for each product. Utilizing this price together with the c.d.f. of the estimated consumer utility distribution for the product, $F_i$, its estimated marginal cost $c_i$ and the estimated single price arrival rate $\lambda_{oi}$, we can obtain the counterfactual single-price mechanism profit $\Pi_{S_i}^* = (p_s^* - c_i)\lambda_{oi}T(1 - F_i(p_s^*))$.

Having calculated the retailer’s counterfactual single-price expected profits, we can next calculate the
Table 5: Counterfactual Analysis Results

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of Events</th>
<th>Avg. Margin</th>
<th>Profit Gain</th>
<th>Avg.(%)</th>
<th>Std.Dev</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refrigerators</td>
<td>44</td>
<td>6.6%</td>
<td>4.42%</td>
<td>0.2378</td>
<td>0.2247</td>
<td>0.2247</td>
</tr>
<tr>
<td>Air Conditioners</td>
<td>39</td>
<td>21.4%</td>
<td>12.42%***</td>
<td>0.2552</td>
<td>0.0043</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Television Sets</td>
<td>63</td>
<td>16.7%</td>
<td>22.71%***</td>
<td>0.2585</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Water Heaters</td>
<td>32</td>
<td>18.5%</td>
<td>5.02%</td>
<td>0.1948</td>
<td>0.1553</td>
<td></td>
</tr>
<tr>
<td>Gas Stoves</td>
<td>27</td>
<td>5.3%</td>
<td>10.28%**</td>
<td>0.2245</td>
<td>0.0250</td>
<td></td>
</tr>
<tr>
<td>Washing Machines</td>
<td>61</td>
<td>27.2%</td>
<td>5.78%</td>
<td>0.2010</td>
<td>0.0284</td>
<td></td>
</tr>
<tr>
<td>All Products</td>
<td>266</td>
<td>17.0%</td>
<td>10.90%***</td>
<td>0.2403</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*: p < 0.1, **: p < 0.05, ***: p < 0.01

estimated realized profits of the retailer from each one of the group buying events in order to determine the net percentage profit increase for each event. The retailer’s realized profit from a group-buying event depends on the total number of customers who signed up during the event window and stayed in after the event, Q. Whether a customer who signed up stays or drops out, in turn, depends on the number of sign ups, N. The firm’s profit equals to the profits from the number of units sold, i.e., \((p - c)Q\), plus the deposits collected from the customers who drop out, i.e., \(d(N - Q)\). Therefore, for product i, the firm’s total realized profit from the group buying event is \(\Pi_{Gi} = (p_{gi} - c_i)Q_i + d(N_i - Q_i)\), where \(p_{gi}\) denotes the realized group-buying price for product i.

The breakdown of the profit gains according to product categories are given in Table 5. As can be seen from the table, estimated profit gains for all categories are statistically significant, except for Refrigerators and Water Heaters which are 4.42% and 5.02% respectively. Television sets have the highest profit gain, 22.71% from implementing Group Buying, and the estimated average profit gains for Air Conditioners and Gas Stoves are also higher than 10%. For the whole sample, the estimated average profit gain is 10.90%. Utilizing a t-test for significance, yields a t-value of 7.40, which is significant at the 1% level (p-value 1.82 \cdot 10^{-12}). Further, applying a Wilcoxon rank sum test for robustness we obtain a z-value of 6.26 with a p-value of 3.96 \cdot 10^{-10}, i.e., it is again significant at 1% level.

Finally, we can project the approximate estimated annual monetary gains from employing group buying events for the retailer in our data set. For the 266 events we analyzed, the average monetary profit per event is 45,460 CNY. The retailer in our data set ran an estimated 5,775 individual group buying events in 2013. Therefore the total annual profits from the group buying events for the retailer is about 262 Million CNY, or 39.10 Million USD. With a 10.90% average gain over the optimum single-price, we can then calculate the estimated total monetary gain for the retailer as 28.56 Million CNY, which is approximately 4.26 Million USD.
4.2.4 The Effect of Retailer Pricing Patterns on Profit Gains

As we have observed from Tables 4 and 5, despite the estimated network effects from group buying is positive and statistically significant, not every product category is estimated to have had significant profit gains from employing the mechanism. For Refrigerators, Water Heaters, and Washing Machines, the estimated profit gains are substantially lower than those for Air Conditioners, Gas Stoves, and Television Sets, and the gains from the former two categories are not statistically significant. What causes these differences among different categories? For this, let us examine the tradeoffs underlying retailer’s decision to select the deal discount, $\delta$. In determining the deal discount, $\delta$, the retailer considers two opposing factors: First, discounts yield losses for the dealer compared to the base price (Cao et al. 2015), and they need to be kept as small as possible. On the other hand, as we have also empirically demonstrated in Section 4.2.2, the larger the deal discounts, the higher the consumer arrival rate, $\lambda_g$, since larger discounts give customers increased incentives to spend effort to disseminate information and recruit other customers (Jing and Xie 2011), and an increase in customer arrival rate on average increases profits.

Given this trade-off, let us first consider the case when $\lambda_o$, the existing customer arrival rate without any networking and promotion by the customers or the base arrival rate is very low. In that case, the probability of crossing even the first threshold is very low and it is very likely that no deal will happen. Therefore, there is very little incentive for customers to spend effort recruiting other customers through networking. As a result, the gains from giving discounts is very limited when the customer arrival rate is low, and she should be setting a low discount level, $\delta$. On the other end of the spectrum, for very high base consumer arrival rates it is very likely that even without any networking by the customers, the second deal threshold will be crossed. Hence, in this case the upside for the customers to spend effort in recruiting others is again very low, and it is again very difficult to incentivize them to network. Therefore, for high base customer arrival rates, it should similarly be in the best interest of the retailer to minimize her losses by setting a very small discount. Hence, for profitable pricing, for very low and very high consumer arrival rates, the discount level, $\delta$, should be set close to zero. On the other hand, for intermediate consumer arrival rate levels, there are likely benefits from setting positive deal discounts. Putting all of this together, an inverse U-shaped pattern of (normalized) deal discounts as a function of the arrival rate would help maximize the profitability from group buying. On the flip side, if the deal discounts are not showing such an inverse U-shaped pattern, then it is likely that the retailer gave too much discount for low and/or high consumer arrival rate products and the retailer profits will be reduced.

To formally demonstrate this role of pricing patterns on profits, we focus on the pricing pattern of each category separately. The deal discounts are set by separate managers who are independently responsible for each category and the pattern for each category reflects the pricing choices of those decision makers.
Figure 5: Clustering of normalized deal discounts based on product categories and the corresponding cluster based fitted regression lines. In all panels, the circles indicate the cluster centroids.

For each event $i$ of the 266 group buying events in the data set, we first start by estimating the base consumer arrival rate, $\lambda_{oi}$, by using the estimated by the regression (19) as discussed in Section 4.2.3. Then for each category, we plot $\delta_i/p_i^*$, as calculated in Section 4.2 versus the estimated $\lambda_{oi}$. In order to detect the shape of the pricing patterns, we then proceed to formally identify the individual segments of the data by clustering for each category. We perform the clustering by employing K-means, K-medoids and GMM methods separately (Bouman et al. 1997; Vassilvitskii 2007), and choosing the method with a lower Euclidean total distance of the centroids to every point in their respective cluster. In order to determine the number of clusters, we evaluate the performance of different numbers of clusters using the Gap criterion (Pujari 2001), which gives us a suggested optimal number of clusters by looking for the highest decrease in error measurement, and by testing it again using the Davies-Bouldin index method (Aggarwal and Reddy 2013). The data plots and clustering results are demonstrated in Figure 5. The clustering process for each category in optimality results in the two clusters seen for each product category – one for small and one for large $\lambda_o$ values (left and right clusters). For a given category, if the deal-discount pricing demonstrates an inverse U-shaped pattern as discussed above, then $\delta/p^*$ values will be increasing for low $\lambda_o$ values (left cluster) and decreasing for high $\lambda_o$ values. To measure these patterns formally, we then group the data points in all the left and right clusters for all categories in two groups, $k = 1$ and $k = 2$ respectively, and for each category $j \in \{1, \ldots, 6\}$ and $k \in \{1, 2\}$ run the regression

$$
\frac{\delta_i}{p_i^*} = \kappa_j + \phi_j^k \lambda_{oi} + \epsilon_{ij}^k.
$$

(25)

In (25), the subscript $i$ indicates the $i^{th}$ event in group $k$ in category $j$ and $\epsilon_{ij}^k$ is the corresponding error term. The full regression results are given in Table E.1 in the Online Supplement. For each category.
Table 6: Effect of pricing patterns on the profit gains

<table>
<thead>
<tr>
<th>Category</th>
<th>Left Clusters</th>
<th>Right Clusters</th>
<th>Profit Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
<td>p-Value</td>
</tr>
<tr>
<td>Television Sets</td>
<td>0.0009***</td>
<td>0.0003</td>
<td>0.0053**</td>
</tr>
<tr>
<td>Air Conditioners</td>
<td>0.0019***</td>
<td>0.0003</td>
<td>&lt;0.0001***</td>
</tr>
<tr>
<td>Gas Stoves</td>
<td>0.0013***</td>
<td>0.0004</td>
<td>0.0065***</td>
</tr>
<tr>
<td>Washing Machines</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.2712</td>
</tr>
<tr>
<td>Water Heaters</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.3248</td>
</tr>
<tr>
<td>Refrigerators</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.4785</td>
</tr>
</tbody>
</table>

*:* \( p < 0.1, \) **:* \( p < 0.05, \) ***:* \( p < 0.01 \)

\( j \in \{1, \ldots, 6\}, \) to visually demonstrate the patterns, we plot the regression line corresponding to that category in Figure 5 on the corresponding cluster.

The pricing patterns estimated from regression (25) and their relationship to profitability is summarized in Table 6. It can be seen from the table that for Television Sets, Air Conditioners and Gas Stoves, the normalized deal discounts are increasing with the consumer arrival rate and decreasing for larger ones, and the slopes are highly statistically significant. That is, these three categories sharply demonstrate the recommended inverse U-shaped pricing pattern. For Washing Machines, Water Heaters, and Refrigerators on the other hand, the slopes are not significant and there is no statistical support for the existence of the recommended price pattern. These observations are also clearly visible in Figure 5. Remarkably, as can be seen from Table 6, there is a clear correspondence between the average estimated profit gains due to group buying and the pricing patterns. In particular, the three categories that demonstrate the recommended inverse U-shaped pattern have much higher profit gains compared to the three categories that failed to employ this pricing pattern. Further, a two-sample t-test confirms the difference in estimated profit gains between the two groups is statistically significant at 1% level with a t-value of 4.1333 and p-value of \( 4.81 \cdot 10^{-5} \). In addition, for a robustness check, we also perform the non-parametric Wilcoxon Rank Sum test for this difference, again finding that it is significant at 1% level with a z-value of 4.0171 and a corresponding p-value of \( 5.89 \cdot 10^{-5} \). These observations demonstrate the profit boosting effect of an inverse U-shaped pricing pattern in group buying, with low deal discounts for least and most popular products, and higher deal discounts for products of intermediate popularity.

5 Concluding Remarks

In this study, we examined consumer behavior and retailer pricing strategy in online group buying events, and demand and profit gains induced by the employment of this retail strategy. Considering consumer sign-up behavior as a continuous time dynamic problem, we presented a game-theoretical model of group buying, and derived the stochastic consumer equilibrium as a solution to a recursive differential equation.
system. After observing existing sign ups, the decision of whether to pay a non-refundable deposit and sign-up depends on a consumer’s belief about the success rate of the deal, i.e., the decisions of subsequent arrivals. Through our equilibrium, we analyzed the evolution of the likelihood of number of sign ups reaching the posted deal thresholds and the consumer propensity of signing up.

Utilizing prices, thresholds, and consumer sign-up times from group buying events hosted on Taobao, we were able to structurally estimate our model’s parameters such as the consumer arrival rates and reservation price distributions through the equilibrium expressions we derived in our theoretical model. The data demonstrates that our model captures the customer sign up behavior in group buying events very well. We measured and provided evidence for the existence of significant demand boosting effects of employing group buying, and estimated that the mechanism improves the retailer profits by more than 10%. Our results also indicate that when the consumer arrival rate is very small or very large, the retailer should set the deal discount amount relatively low, approaching single-pricing, while setting higher discounts for intermediate arrival rates. This suggests that bulk of the gains from group buying events come from not highly popular products nor those with very low consumer interest, but rather for products with medium levels of consumer demand, since in this region, the retailer can harvest the benefits of customers’ networking activities by offering discounts. Managers who employ group buying can improve profitability by utilizing insights from our analysis and following such pricing patterns.

Our study can be considered as a first step into a broader future stream of research that analyzes group buying, and provides insights into this novel method of retailing that is growing in popularity. There are many future directions of research to deepen our understanding of the subject. One such research avenue can be analyzing other formats of implementing the concept, such as increased number of deal thresholds or varying deposits levels. The analysis of such more complicated settings, though, can be challenging, especially in a dynamic, multi-agent environment as we aimed to tackle in this paper. However, the insights provided from our study as well as follow up studies can be helpful in future design and efficient use of group buying mechanisms, and ultimately better harvesting of the value generated from this innovative channel strategy.

References


Appendix for
Consumer Equilibrium, Demand Effects,
and Efficiency in Group Buying

A Mathematical Proofs

Proof of Lemma 1: Consider a customer with utility \( u > p_0 - d \). For such customer \( \max\{u - p_0, -d\} = u - p_0 \), and provided that this customer signs up, he will always stay in after the event, even if neither of the two deals happen. Hence, such a customer arriving at time \( t \in [0, 1] \) will sign up if and only if

\[
V_k(u, t) = (u - p_1)\pi_{k+1}^1(t) + (u - p_2)\pi_{k+1}^2(t) + (u - p_0)(1 - \pi_{k+1}^1(t) - \pi_{k+1}^2(t)) \geq 0,
\]

which holds if and only if

\[
u \geq \bar{u}_{k,t}^1 \triangleq p_0 - \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)).
\]

In a similar manner, a consumer with reservation value \( p_1 - d < u \leq p_0 - d \), will drop out after the event only when neither of the two deals materializes, and signs up at time \( t \) if and only if

\[
V_k(u, t) = (u - p_1)\pi_{k+1}^1(t) + (u - p_2)\pi_{k+1}^2(t) + (-d)(1 - \pi_{k+1}^1(t) - \pi_{k+1}^2(t)) \geq 0
\]

\[
\iff u \geq \bar{u}_{k,t}^2 \triangleq p_0 - d + \frac{d - \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t))}{\pi_{k+1}^1(t) + \pi_{k+1}^2(t)};
\]

and finally, a customer with utility \( u \leq p_1 - d \), will drop out if the second deal does not materialize and signs up if and only if

\[
V_k(u, t) = (-d)\pi_{k+1}^1(t) + (u - p_2)\pi_{k+1}^2(t) + (-d)(1 - \pi_{k+1}^1(t) - \pi_{k+1}^2(t)) \geq 0
\]

\[
\iff u \geq \bar{u}_{k,t}^3 \triangleq p_0 - 2d - d \left(1 - \frac{1}{\pi_{k+1}^1(t)}\right).
\]

Note that by (A.2)-(A.4), \( \bar{u}_{k,t}^1 > p_0 - d \) if and only if \( d > \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) \); \( p_1 - d < \bar{u}_{k,t}^2 \leq p_0 - d \), if and only if \( d\pi_{k+1}^2(t) < d \leq \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) \); and \( \bar{u}_{k,t}^3 \leq p_1 - d \) if and only if \( d \leq \delta\pi_{k+1}^2(t) \). Further, \( \bar{u}_{k,t}^1 \geq p_2 \) for \( i \in \{1, 2, 3\} \).

Consider a customer with utility \( u \), arriving at time \( t \) with \( k \) existing sign ups at the time. First suppose \( d \leq \delta\pi_{k+1}^2(t) \). Then, since \( d \leq \delta\pi_{k+1}^2(t) \leq \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) \), we have \( \bar{u}_{k,t}^1 \leq p_0 - d \), and \( \bar{u}_{k,t}^2 \leq p_1 - d \).

Therefore, in this case, any customer with reservation value \( u \geq \bar{u}_{k,t}^3 \) will sign up. Next, consider the case \( d\pi_{k+1}^2(t) < d \leq \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) \). Then \( \bar{u}_{k,t}^1 \leq p_0 - d \), and \( \bar{u}_{k,t}^3 > p_1 - d \), which means that, all

A.1
customers with \( u \geq p_0 - d \) will sign up and no customer with \( u < p_1 - d \) will sign up, and a customer will sign up if and only if \( u \geq \bar{u}_{k,t}^2 \). Finally, if \( d > \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) \), we have \( \bar{u}_{k,t}^2 \geq p_0 - d \), and \( \bar{u}_{k,t}^3 \geq p_1 - d \), i.e., no consumer with reservation value \( u < p_0 - d \) will sign up. Hence, when \( d > \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) \), with \( k \) existing arrivals at time \( t \), an arriving customer will sign up if and only if \( u \geq \pi_{k,t}^1 \).

By (A.2)-(A.4), it then follows that with \( k \geq 0 \) existing sign ups at time \( t \in [0,T] \) an arriving customer with utility \( u \), will sign up if and only if \( u \geq \bar{u}_{k,t}^2 \) \( \geq p_2 \), where

\[
\bar{u}_{k,t} = \begin{cases}
    p_0 - 2\delta - d \left( 1 - \frac{1}{\pi_{k+1}^2(t)} \right) & \text{if } 0 \leq d \leq \delta \pi_{k+1}^2(t), \\
    p_0 - d - \frac{d - \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t))}{\pi_{k+1}^1(t) + \pi_{k+1}^2(t)} & \text{if } \delta \pi_{k+1}^2(t) < d \leq \delta \pi_{k+1}^1(t) + 2\delta \pi_{k+1}^2(t), \\
    p_0 - \delta(\pi_{k+1}^1(t) + 2\pi_{k+1}^2(t)) & \text{if } d > \delta \pi_{k+1}^1(t) + 2\delta \pi_{k+1}^2(t).
\end{cases} \tag{A.5}
\]

This completes the proof. \( \blacksquare \)

**Proof of Proposition 1:** We start with the derivation of the recursive functional form of \( \pi_k^2(t) \). By (4), we have

\[
\pi_k^2(t) = \int_0^{T-t} (H_k(t+x)\pi_{k+1}^2(t+x) + (1 - H_k(t+x))\pi_k^2(t+x))\lambda g e^{-\lambda_s u} dx = \int_t^T H_k(u)\pi_{k+1}^2(u)\lambda g e^{-\lambda_s (u-t)} du + \int_t^T (1 - H_k(u))\pi_k^2(u)\lambda g e^{-\lambda_s (u-t)} du. \tag{A.6}
\]

Let \( g_k(t) = \int_t^T H_k(u)\pi_{k+1}^2(u)\lambda g e^{-\lambda_s (u-t)} du \). Then,

\[
\pi_k^2(t) = g_k(t) + \int_t^T (1 - H_k(u))\pi_k^2(u)\lambda g e^{-\lambda_s (u-t)} du = g_k(t) + e^{\lambda_s t} \int_T^T (H_k(u) - 1)\pi_k^2(u)\lambda g e^{-\lambda_s u} du. \tag{A.7}
\]

Taking the derivative with respect to \( t \) and by (A.7), we have

\[
\frac{\partial \pi_k^2(t)}{\partial t} = \frac{\partial g_k(t)}{\partial t} + \lambda g \int_T^t (H_k(u) - 1)\pi_k^2(u)\lambda g e^{-\lambda_s u} du + e^{\lambda_s t}(H_k(t) - 1)\pi_k^2(t)\lambda g e^{-\lambda_s t} \\
= \frac{\partial g_k(t)}{\partial t} + \lambda g (\pi_k^2(t) - g_k(t)) + \lambda g (H_k(t) - 1)\pi_k^2(t) \\
= \frac{\partial g_k(t)}{\partial t} - \lambda g g_k(t) + \lambda g H_k(t)\pi_k^2(t). \tag{A.8}
\]

Now,

\[
\frac{\partial g_k(t)}{\partial t} = -\lambda g e^{\lambda_s t} \int_T^t H_k(u)\pi_{k+1}^2(u)\lambda g e^{-\lambda_s u} du - e^{\lambda_s t} H_k(t)\pi_{k+1}^2(t)\lambda g e^{-\lambda_s t} = \lambda g g_k(t) - \lambda g H_k(t)\pi_k^2(t), \tag{A.9}
\]

A.2
which, by substituting into (A.8) yields the recursive differential equation

\[
\frac{\partial \pi_k^2(t)}{\partial t} = \lambda_g H_k(t) \pi_k^2(t) - \lambda_g H_k(t) \pi_{k+1}^2(t).
\]  

(A.10)

By (A.10), we have

\[
e^{-\int_0^t \lambda_g H_k(v) dv} \left( \frac{\partial \pi_k^2(u)}{\partial u} - \lambda_g H_k(u) \pi_k^2(u) \right) = -e^{-\int_0^t \lambda_g H_k(v) dv} \lambda_g H_k(u) \pi_{k+1}^2(u),
\]

which implies

\[
e^{-\int_0^t \lambda_g H_k(v) dv} \pi_k^2(u) \bigg|_t^T = - \int_t^T \! e^{-\int_0^s \lambda_g H_k(v) dv} \lambda_g H_k(u) \pi_{k+1}^2(u) du.
\]

(A.12)

By (A.12) and the boundary condition \(\pi_k^2(T) = 0\) for all \(k \leq M_2 - 1\), it follows that

\[
-e^{-\lambda_g \int_0^t H_k(v) dv} \pi_k^2(t) = -\lambda_g \int_t^T \! e^{-\lambda_g \int_0^s H_k(v) dv} H_k(u) \pi_{k+1}^2(u) du,
\]

(A.13)

from which we have

\[
\pi_k^2(t) = \lambda_g \int_t^T \! e^{-\lambda_g \int_0^s H_k(v) dv} \lambda_g H_k(u) \pi_{k+1}^2(u) du,
\]

(A.14)

as stated in (6). The corresponding equation for \(\pi_k^1(t)\) for \(0 \leq k \leq M_1 - 1\), can also be derived similarly.

Now, since \(H_k(t) = Pr\{V_k(u, t) \geq 0\}\), and by Lemma 1, a customer with utility \(u\) arriving at time \(t\) with \(k\) existing arrivals will sign up if and only if \(u \geq \bar{u}_{k,t}\), where \(\bar{u}_{k,t}\) is as defined in (3), it follows that \(H_k(t) = 1 - F(\bar{u}_{k,t})\). By the boundary conditions, \(\pi_k^1(t) = 0\), \(\pi_k^2(t) = 1\), and by (3), \(H_k(t) = 1 - F(p_2)\) for all \(t \in [0, T]\), \(k \geq M_2\), and hence we have a full unique characterization of \((\pi_k^1, \pi_k^2, H_k)\) for \(k \geq M_2\).

Notice that by (3), for all \(k \geq 0\), \(H_k\) is uniquely determined by \(\pi_k^1\) and \(\pi_k^2\). Hence, for any \(k\), such that \(1 \leq k \leq M_2 - 1\), if we have a full unique characterization of \((\pi_k^1, \pi_k^2, H_k)\), we have a full unique characterization of \(H_{k-1}\). Further, given a full unique characterization of \((\pi_k^1, \pi_k^2, H_{k-1})\), by utilizing (6) we can uniquely obtain \(\pi_{k-1}^2\). Further, for all \(t \in [0, T]\), for \(M_1 \leq k \leq M_2 - 1\), by the boundary conditions, \(\pi_k^1(t) = 1 - \pi_k^2(t)\), and for \(1 \leq k < M_1\), \(\pi_k^1\) can be solved again utilizing (6). Therefore, we can obtain \(\pi_k^1\) and \(\pi_k^2\), which implies that we have a full unique characterization of \((\pi_k^1, \pi_k^2, H_{k-1})\). Hence, by backward induction, we have a full unique characterization of \((\pi_k^1, \pi_k^2, H_k)\) for all \(k \geq 0\).
## B Descriptive Statistics

### Table B.1: Descriptive Statistics for Group Buying Events and Single Price Sales

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of Events</th>
<th>Price (CNY)</th>
<th>Discount (CNY)</th>
<th>Avg. No of Reviews</th>
<th>Mean Review Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Std.Dev</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Group Buying Events</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refrigerators</td>
<td>44</td>
<td>2994.45</td>
<td>2799</td>
<td>1098.30</td>
<td>163.84</td>
</tr>
<tr>
<td>Air Conditioners</td>
<td>39</td>
<td>3087.36</td>
<td>2899</td>
<td>1523.93</td>
<td>159.04</td>
</tr>
<tr>
<td>Television Sets</td>
<td>63</td>
<td>3189.88</td>
<td>2599</td>
<td>1092.26</td>
<td>177.39</td>
</tr>
<tr>
<td>Water Heaters</td>
<td>32</td>
<td>2859.37</td>
<td>2499</td>
<td>793.03</td>
<td>144.22</td>
</tr>
<tr>
<td>Gas Stoves</td>
<td>27</td>
<td>3099.35</td>
<td>2949</td>
<td>1484.52</td>
<td>150.53</td>
</tr>
<tr>
<td>Washing Machines</td>
<td>61</td>
<td>3101.04</td>
<td>2549</td>
<td>1211.67</td>
<td>158.34</td>
</tr>
<tr>
<td><strong>Single Price Sales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refrigerators</td>
<td>252</td>
<td>3095.91</td>
<td>2849</td>
<td>1113.25</td>
<td></td>
</tr>
<tr>
<td>Air Conditioners</td>
<td>234</td>
<td>3236.18</td>
<td>2999</td>
<td>1425.16</td>
<td></td>
</tr>
<tr>
<td>Television Sets</td>
<td>264</td>
<td>3338.44</td>
<td>2899</td>
<td>1496.00</td>
<td></td>
</tr>
<tr>
<td>Water Heaters</td>
<td>234</td>
<td>2960.36</td>
<td>2599</td>
<td>901.77</td>
<td></td>
</tr>
<tr>
<td>Gas Stoves</td>
<td>276</td>
<td>3145.72</td>
<td>2899</td>
<td>1517.48</td>
<td></td>
</tr>
<tr>
<td>Washing Machines</td>
<td>300</td>
<td>3238.98</td>
<td>2599</td>
<td>1393.98</td>
<td></td>
</tr>
</tbody>
</table>

### Table B.2: Descriptive Statistics for the Products

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of Products</th>
<th>Product Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Capacity (L)</strong></td>
</tr>
<tr>
<td>Refrigerators</td>
<td>86</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Air Conditioners</td>
<td>78</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.22</td>
</tr>
<tr>
<td>Television Sets</td>
<td>107</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Water Heaters</td>
<td>71</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>Gas Stoves</td>
<td>73</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Washing Machines</td>
<td>111</td>
<td>Min</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

A.4
## C Category Breakdown of Group Buying Parameter Estimation

### 1-Refrigerators

<table>
<thead>
<tr>
<th>Utility Distribution</th>
<th>No. of Events</th>
<th>Min</th>
<th>Average</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>30</td>
<td>1062.29</td>
<td>3618.66</td>
<td>6399.40</td>
<td>166.64</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>7</td>
<td>1602.56</td>
<td>3273.54</td>
<td>6160.33</td>
<td>255.27</td>
</tr>
<tr>
<td>Normal</td>
<td>7</td>
<td>1679.70</td>
<td>2927.02</td>
<td>5183.90</td>
<td>219.09</td>
</tr>
</tbody>
</table>

Pseudo $R^2$: Min Average Max
0.35 0.61 0.87

### 2-Air Conditioners

<table>
<thead>
<tr>
<th>Utility Distribution</th>
<th>No. of Events</th>
<th>Min</th>
<th>Average</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>24</td>
<td>1125.90</td>
<td>3837.67</td>
<td>6365.67</td>
<td>175.91</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>8</td>
<td>1974.48</td>
<td>3832.79</td>
<td>7231.87</td>
<td>128.91</td>
</tr>
<tr>
<td>Normal</td>
<td>7</td>
<td>1670.62</td>
<td>3648.18</td>
<td>5541.02</td>
<td>458.05</td>
</tr>
</tbody>
</table>

Pseudo $R^2$: Min Average Max
0.39 0.62 0.87

### 3-Television Sets

<table>
<thead>
<tr>
<th>Utility Distribution</th>
<th>No. of Events</th>
<th>Min</th>
<th>Average</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>37</td>
<td>1055.15</td>
<td>3580.55</td>
<td>6286.38</td>
<td>192.95</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>12</td>
<td>892.35</td>
<td>4265.38</td>
<td>7084.22</td>
<td>143.71</td>
</tr>
<tr>
<td>Normal</td>
<td>14</td>
<td>1311.90</td>
<td>3902.11</td>
<td>5835.86</td>
<td>201.49</td>
</tr>
</tbody>
</table>

Pseudo $R^2$: Min Average Max
0.39 0.66 0.86

### 4-Water Heaters

<table>
<thead>
<tr>
<th>Utility Distribution</th>
<th>No. of Events</th>
<th>Min</th>
<th>Average</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>23</td>
<td>1002.42</td>
<td>3931.13</td>
<td>6305.64</td>
<td>169.57</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>3</td>
<td>2665.80</td>
<td>3017.57</td>
<td>3233.54</td>
<td>224.87</td>
</tr>
<tr>
<td>Normal</td>
<td>6</td>
<td>2093.49</td>
<td>3934.29</td>
<td>5390.35</td>
<td>201.49</td>
</tr>
</tbody>
</table>

Pseudo $R^2$: Min Average Max
0.39 0.64 0.87

### 5-Gas Stoves

<table>
<thead>
<tr>
<th>Utility Distribution</th>
<th>No. of Events</th>
<th>Min</th>
<th>Average</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>17</td>
<td>1198.45</td>
<td>3194.38</td>
<td>5788.39</td>
<td>201.85</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>5</td>
<td>1495.04</td>
<td>3022.77</td>
<td>6429.49</td>
<td>224.37</td>
</tr>
<tr>
<td>Normal</td>
<td>5</td>
<td>1397.42</td>
<td>2740.75</td>
<td>3955.14</td>
<td>250.91</td>
</tr>
</tbody>
</table>

Pseudo $R^2$: Min Average Max
0.37 0.70 0.89

### 6-Washing Machines

<table>
<thead>
<tr>
<th>Utility Distribution</th>
<th>No. of Events</th>
<th>Min</th>
<th>Average</th>
<th>Max</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>38</td>
<td>1206.72</td>
<td>3509.82</td>
<td>6511.37</td>
<td>165.16</td>
</tr>
<tr>
<td>Log-Normal</td>
<td>14</td>
<td>906.54</td>
<td>2657.82</td>
<td>6018.87</td>
<td>145.14</td>
</tr>
<tr>
<td>Normal</td>
<td>9</td>
<td>1463.15</td>
<td>3804.47</td>
<td>5520.93</td>
<td>213.72</td>
</tr>
</tbody>
</table>

Pseudo $R^2$: Min Average Max
0.36 0.62 0.85

A.5
## D Consumer Utility Distribution Regression Results

Table D.1: Regressions for Consumer Utility Distributions based on Product Category

<table>
<thead>
<tr>
<th>Product Category</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
</tr>
<tr>
<td><strong>Refrigerators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>4.9965***</td>
<td>0.5947</td>
</tr>
<tr>
<td>Capacity</td>
<td>0.5134***</td>
<td>0.1007</td>
</tr>
<tr>
<td>F-ratio: 26.00, ( R^2 ): 0.40</td>
<td>F-ratio: 10.85, ( R^2 ): 0.21</td>
<td></td>
</tr>
<tr>
<td><strong>Air Conditioners</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>3.8223***</td>
<td>1.4121</td>
</tr>
<tr>
<td>Capacity</td>
<td>1.4121***</td>
<td>0.2569</td>
</tr>
<tr>
<td>Energy Level</td>
<td>-0.2452**</td>
<td>0.0955</td>
</tr>
<tr>
<td>F-ratio: 46.37, ( R^2 ): 0.74</td>
<td>F-ratio: 15.65, ( R^2 ): 0.45</td>
<td></td>
</tr>
<tr>
<td><strong>Television Sets</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-3.3524***</td>
<td>0.4399</td>
</tr>
<tr>
<td>Size</td>
<td>3.0247***</td>
<td>0.1156</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.1274***</td>
<td>0.0369</td>
</tr>
<tr>
<td>F-ratio: 342.20, ( R^2 ): 0.92</td>
<td>F-ratio: 64.63, ( R^2 ): 0.71</td>
<td></td>
</tr>
<tr>
<td><strong>Water Heaters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>5.0525***</td>
<td>0.5810</td>
</tr>
<tr>
<td>Capacity</td>
<td>0.5513***</td>
<td>0.1275</td>
</tr>
<tr>
<td>Power</td>
<td>0.3524***</td>
<td>0.0821</td>
</tr>
<tr>
<td>F-ratio: 15.76, ( R^2 ): 0.52</td>
<td>F-ratio: 6.91, ( R^2 ): 0.30</td>
<td></td>
</tr>
<tr>
<td><strong>Gas Stoves</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>3.6399***</td>
<td>0.6508</td>
</tr>
<tr>
<td>Power</td>
<td>0.8903***</td>
<td>0.1464</td>
</tr>
<tr>
<td>Panel</td>
<td>0.1192*</td>
<td>0.0579</td>
</tr>
<tr>
<td>F-ratio: 23.39, ( R^2 ): 0.67</td>
<td>F-ratio: 13.80, ( R^2 ): 0.54</td>
<td></td>
</tr>
<tr>
<td><strong>Washing Machines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>6.5766***</td>
<td>1.0361</td>
</tr>
<tr>
<td>Capacity</td>
<td>1.0361***</td>
<td>0.0870</td>
</tr>
<tr>
<td>Energy level</td>
<td>-0.1085***</td>
<td>0.0299</td>
</tr>
<tr>
<td>F-ratio: 77.10, ( R^2 ): 0.74</td>
<td>F-ratio: 16.24, ( R^2 ): 0.36</td>
<td></td>
</tr>
</tbody>
</table>

*:* \( p < 0.1 \), **:* \( p < 0.05 \), ***:* \( p < 0.01 \)
E Price Discount Trend Regression Results

Table E.1: Trend regressions for Normalized Deal discounts

<table>
<thead>
<tr>
<th>Category</th>
<th>$\delta/p^*_\kappa$</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-Value</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left Clusters ($k = 1$)</td>
<td></td>
<td></td>
<td></td>
<td>Right Clusters ($k = 2$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refrigerators</td>
<td>0.1438***</td>
<td>0.0250</td>
<td>&lt;0.0001</td>
<td></td>
<td>0.0783</td>
<td>0.0668</td>
<td>0.2433</td>
</tr>
<tr>
<td>Air Conditioners</td>
<td>0.0233</td>
<td>0.0353</td>
<td>0.4593</td>
<td></td>
<td>0.2173***</td>
<td>0.0611</td>
<td>0.0005</td>
</tr>
<tr>
<td>Television Sets</td>
<td>0.0447***</td>
<td>0.0225</td>
<td>0.0496</td>
<td></td>
<td>0.1930***</td>
<td>0.0407</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Water Heaters</td>
<td>0.0856**</td>
<td>0.0289</td>
<td>0.0111</td>
<td></td>
<td>0.2189***</td>
<td>0.0522</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Gas Stoves</td>
<td>0.0329</td>
<td>0.0300</td>
<td>0.2747</td>
<td></td>
<td>0.3671***</td>
<td>0.1023</td>
<td>0.0005</td>
</tr>
<tr>
<td>Washing Machines</td>
<td>0.0376**</td>
<td>0.0227</td>
<td>0.0495</td>
<td></td>
<td>0.1292**</td>
<td>0.0619</td>
<td>0.0474</td>
</tr>
<tr>
<td>Refrigerators</td>
<td>-0.0001</td>
<td>0.0002</td>
<td>0.5332</td>
<td></td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.5842</td>
</tr>
<tr>
<td>Air Conditioners</td>
<td>0.0008***</td>
<td>0.0003</td>
<td>0.0036</td>
<td></td>
<td>-0.0004*</td>
<td>0.0002</td>
<td>0.0934</td>
</tr>
<tr>
<td>Television Sets</td>
<td>0.0005***</td>
<td>0.0002</td>
<td>0.0073</td>
<td></td>
<td>-0.0004***</td>
<td>0.0001</td>
<td>0.0089</td>
</tr>
<tr>
<td>Water Heaters</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.7210</td>
<td></td>
<td>-0.0005***</td>
<td>0.0002</td>
<td>0.0051</td>
</tr>
<tr>
<td>Gas Stoves</td>
<td>0.0006**</td>
<td>0.0002</td>
<td>0.0148</td>
<td></td>
<td>-0.0009***</td>
<td>0.0003</td>
<td>0.0069</td>
</tr>
<tr>
<td>Washing Machines</td>
<td>0.0006***</td>
<td>0.0002</td>
<td>0.0044</td>
<td></td>
<td>-0.0001</td>
<td>0.0002</td>
<td>0.2616</td>
</tr>
</tbody>
</table>

No. of observations: 131, $R^2$: 0.25
F-ratio: 3.60, p-value: 0.0002

No. of observations: 135, $R^2$: 0.23
F-ratio: 3.42, p-value: 0.0004

*: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$