Estimating Correlated Valuations for Data-Driven Bundle Pricing with Copula Inference

Ben Lethem, Wei Sun, Anshul Sheopuri

Submitted to *Manufacturing & Service Operations Management*
manuscript

Estimating Correlated Valuations for Data-Driven Bundle Pricing with Copula Inference

(Authors' names blinded for peer review)

Bundle discounts are used by retailers in many industries. Optimal bundle pricing requires learning the joint distribution of consumer valuations for the items in the bundle, that is, how much they are willing to pay for each of the items. We suppose that a retailer has sales transaction data, and the corresponding consumer valuations are latent variables. We develop a statistically consistent and computationally tractable inference procedure for fitting a copula model over correlated valuations, using only sales transaction data for the individual items. We use data from a telecommunication service provider to show how the method can be extended to bundles with intra-category competition. Simulations and data experiments demonstrate consistency, scalability, and the importance of incorporating correlations in the joint distribution.

Key words: Bundle pricing, copula, inference, discrete choice model, mixed logit model

1. Introduction

Item bundles, when a collection of items are sold together at a discount, are widely used in a variety of industries, ranging from technology products (e.g., smartphones with data plans) to fast foods (e.g., McDonalds' Value Meal) to information goods (e.g., album purchases on iTune). Both theoretical and empirical work has shown that introducing an appropriately priced bundle can significantly increase profits (Eppen et al. 1991). The potential downside of bundling is that the bundle offer can cannibalize demand from customers who would have anyway purchased all of the items without the discount. A poorly priced bundle can thus result in a loss of profits, making it critical to be able to predict at what price will it be profitable. The answer to this question relies critically on a knowledge of product valuations, which refer to the prices that consumers are willing to pay, as well as the interplay between the valuations of items in the bundle.

The economics literature on bundling has extensively examined the economic efficiency of bundling and how bundling can be used for price discrimination (Adams and Yellen 1976, Schmalensee 1982, McAfee et al. 1989). These foundational studies have been extended in many directions. Several papers have focused on analytical solutions for the optimal bundle price and other quantities of interest (Venkatesh and Kamakura 2003, McCardle et al. 2007, Eckalbar 2010). These analytical results were obtained for the special case of uniformly distributed valuations, with the distributions for items in the bundle either independent or perfectly correlated. Schmalensee (1984) obtained some analytical results and insights by assuming the joint distribution to be bivariate normal. Other results have been obtained for a finite collection of deterministic valuations (Hanson and Martin 1990).

A number of useful insights can be gained from these simplified models (see, for example, Stremersch and Tellis 2002). When working with data, such strong assumptions about the joint distribution, particularly independence, are no longer appropriate. Jedidi et al. (2003) eschew independence assumptions and use methodology based in utility theory to measure valuations. Their measurement procedure requires offering the bundle at various prices to elicit the demand function for the bundle. Based on their empirical results, they report that "models that assume statistical independence are likely to be misspecified." Venkatesh and Mahajan (1993) also study bundle pricing without distributional assumptions for valuations, by mailing out questionnaires that directly asked consumers for their valuations. Conjoint analysis has also been used to estimate the valuation distribution from questionnaire data in the context of bundling (Goldberg et al. 1984, Wuebeker and Mahajan 1999).

Bundles generally come in two forms: pure bundling and mixed bundling. In pure bundling, the bundle items would no longer be offered individually, whereas in mixed bundling the items continue to be offered individually alongside the bundle offer. In this paper, we consider mixed bundling. The primary goal of this work is to address the following question: *For a collection of items as a possible bundle, given historical transaction data for those items, how can we determine the bundle price that would maximize expected profit?*

We develop an inference procedure to estimate the joint valuation distribution across products in a bundle for predicting the impact of different pricing strategies. Our contributions are four-fold. Firstly, the procedure is data-driven and only requires transaction data. Thus, it circumvents the need of using experiments or questionnaires as what has been proposed in the literature. In addition, it does not require collecting sales data for the bundle *a priori*. In other words, we are able to learn useful information about how to price a bundle even if the bundle has not been previously offered. Our procedure is highly flexible. It is not limited to the distribution families that have been previously considered in the literature (e.g., uniform or normal). In fact, arbitrary marginal distributions can be incorporated. Moreover, the procedure does not require restrictive dependence structure of the joint distribution such as independence or perfect correlation. Our procedure is based on inference of a copula model over latent consumer valuations. For the base scenario where there is a single collection of items, any combination of which might be purchased, we show that the estimates obtained from this estimation procedure are statistically consistent.

Secondly, we show that the inference procedure is also computationally tractable, which is crucial for practical implementation. Because the valuations are unobserved, the likelihood function involves integrating over the latent valuations, and standard formulas for copula fitting cannot be directly applied. In our key result, we show that using the inclusion-exclusion theorem, these computationally intractable integrals can be transformed into distribution function evaluations, thus allowing for efficient estimation. Our numeric studies with simulated data demonstrate that the estimated parameters converge quickly to the true value even with small datasets, suggesting that our method is not limited to retailers with large collections of transaction history.

Thirdly, we extend our method from the base scenario to a setting where a bundle is made of compoents from several categories and each component could face competition from substitutable products in the same category. The existence of both *intra-category substitution* and *cross-category correlation* dramatically increase the analytical and computational complexity of bundle pricing. We show how the valuation estimation can be done in a choice modeling framework. Powerful choice models, like the mixed logit model, are used to describe the valuations within each category, and then we use our copula framework to estimate correlations across categories.

Lastly, we show how the method can have a real impact in practice with a case study in the telecommunications industry. Using sales transactions from a telecommunications service provider, we investigate how the provider can price their phone-plan bundles. In particular, to capture the complex substitution patterns in the smartphone market which is influenced by product cannibalization, technology obsolescence and heterogeneous customer preferences, we implement a mixed logit model with random coefficients to model phone sales. Our results highlight the importance of modeling the joint dependence across product valuations and show that there is room for a significant increase in profits over current practice.

The paper is organized as follows. We begin with the base scenario where a single collection of items is offered (i.e., intra-category competition is absent) in Section 2. We will present the key steps in the inference procedure which utilizes copula estimation. In Section 3, we demonstrate the procedure with a series of simulated studies for the base scenario. In Section 4, we will show how to adapt the inference procedure for a scenario where each bundle component experiences competition from other substitutable produces in the same category. This is the scenario of our case study for a telecommunication service provider, which is shown in Section 5. We conclude the paper in Section 6.

2. Estimating the Profits of Bundling

In this section, we focus on the base scenario, where there is a single collection of items, any combination of which might be purchased. There are two components to the transaction data that our model is based on: purchase data $\boldsymbol{y}^t$ and price data $\boldsymbol{x}^t$. In addition, there may be a collection of covariates $\boldsymbol{z}^t$. The purchase data $\boldsymbol{y}^t = [y_1^t, \dots, y_n^t]$ contain the sales for transaction t, with $y_i^t = 1$ if item i was purchased in transaction t, and 0 otherwise. Similarly for the price data, we let x_i^t denote the price of item i at the time of transaction t. The covariates $\boldsymbol{z}^t$ contain other data that we may have available at the time of transaction t. Competitors' prices and season are examples of covariates that we might include in our model of how the data $\boldsymbol{y}^t$ are generated. We suppose that we have a database of T total transactions.

2.1. Valuations and Consumer Rationality

The relation between prices $\boldsymbol{x}^t$ and purchases $\boldsymbol{y}^t$ is modeled using a valuation, or willingness-to-pay, for each item and for each transaction. Let v_i^t represent the valuation for item i by the consumer in transaction t. We do not observe v_i^t directly, but we gain insight into it by assuming, as is done throughout the bundling literature, that consumers are rational. Specifically, we model consumers as having infinite budget and purchasing an item if and only if it is priced below their valuation: $y_i^t = 1$ if and only if $v_i^t > x_i^t$. This assumption defines a relationship between the valuations v_i^t and the observed data y_i^t and x_i^t that we can use to estimate the distribution of v_i^t. We allow the distribution of v_i^t to also depend on the covariates $\boldsymbol{z}^t$, should there be any.

We assume the valuations for transaction t come from some joint distribution which we denote as $\boldsymbol{v}^t \sim F(v_1, \dots, v_n \mid \boldsymbol{z}^t)$, independently for $t = 1, \dots, T$. We denote the marginal valuation distribution for v_i^t as $F_i(v_i \mid \boldsymbol{z}^t)$. We attempt to estimate the distribution $F(v_1, \dots, v_n \mid \boldsymbol{z}^t)$, as this leads directly to the estimate of bundle profits as we now show.

2.2. Computing the Optimal Bundle Price

Given the joint valuation distribution, the expected profit per consumer as a function of item and bundle prices, and possibly covariates, can be computed. For notational convenience here we give the result only for $n = 2$, but similar results can be had for $n > 2$. There are four possible actions for each consumer: purchase item 1 only, purchase item 2 only, purchase the bundle, or purchase none of the items. The earlier assumption of rational consumers means that they will choose the action that maximizes their surplus $v_i - x_i$. For the result that follows, we assume that the valuation for the bundle is the sum of the component valuations: $v_B = v_1 + v_2$. This assumption could easily be relaxed to other bundle valuation models, for example, Venkatesh and Kamakura (2003) provide alternative models for how valuations combine. We let c_i be the cost of item i and assume that the bundle cost is the sum of the component costs: $c_B = c_1 + c_2$.

PROPOSITION 1. *Let $f(\cdot)$ be the joint valuation density function, $F(\cdot)$ the corresponding distribution function, and $\boldsymbol{z}$ the covariate values. The expected profit per consumer when items 1, 2, and the bundle are priced at x_1, x_2, and x_B respectively is:*

$$\begin{aligned}
\mathbb{E}\left[\textit{profit}\right] = {} & (x_1 - c_1)(F_2(x_B - x_1 \mid \boldsymbol{z}) - F(x_1, x_B - x_1 \mid \boldsymbol{z})) \\
& + (x_2 - c_2)(F_1(x_B - x_2 \mid \boldsymbol{z}) - F(x_B - x_2, x_2 \mid \boldsymbol{z})) \\
& + (x_B - c_1 - c_2)\Bigg(1 - F_1(x_B - x_2 \mid \boldsymbol{z}) - F_2(x_B - x_1 \mid \boldsymbol{z}) + F(x_B - x_2, x_B - x_1 \mid \boldsymbol{z}) \\
& \qquad - \int_{x_B - x_2}^{x_1} \int_{x_B - x_1}^{x_B - v_1} f(v_1, v_2 \mid \boldsymbol{z}) dv_2 dv_1 \Bigg).
\end{aligned}$$

The proof is given in the Appendix. This expression can be maximized with respect to x_B to obtain the optimal bundle price subject to current individual item pricing, or the maximization could be done over all three prices to obtain a complete pricing strategy. The formula is not concave in general, but a local maximum can be found for instance using concave maximization with random restarting. We now turn our attention to the problem of estimating the valuation distribution.

2.3. Copulas

We provide a parametric estimation procedure for the joint valuation distribution, and must begin by selecting an appropriate form. Perhaps the most straightforward approach to modeling a joint distribution is to assume that the valuations for each item are independent, so that $F(v_1, \ldots, v_n \mid \boldsymbol{z}) = \prod_{i=1}^n F_i(v_i \mid \boldsymbol{z})$. This model allows for arbitrary marginal distributions, however independence is likely an unreasonable assumption. In fact, we show in our simulations in Section 3 that modeling correlations between items is important for correctly predicting bundle profits. A commonly used joint distribution model that would allow for correlations between items is the multivariate normal distribution. This model allows for correlations between items according to the covariance matrix, however it requires the margins to be normally distributed. We show in Section 2.6 that the margin distributions correspond directly to demand models, and such a limitation on the demand model would not be reasonable.

Copula models are a class of joint distributions that allow both correlation structures and arbitrary margins. Copula models are widely used in statistics and finance, and are becoming increasingly utilized for machine learning due to their flexibility and computational properties (see, for example, Elidan 2013). A copula $\mathbb{C}(\cdot \mid \boldsymbol{z})$ for distribution $F(\cdot \mid \boldsymbol{z})$ is a distribution function over $[0,1]^n$ with uniform margins such that

$$F(v_1,\ldots,v_n \mid \boldsymbol{z}) = \mathbb{C}(F_1(v_1 \mid \boldsymbol{z}),\ldots,F_n(v_n \mid \boldsymbol{z}) \mid \boldsymbol{z}).$$

The copula is thus the function that combines the margins in such a way as to return the joint distribution. The field of copula modeling is based on a representation theorem by Sklar (1973) which shows that every distribution has a copula, and if the margins are continuous, the copula is unique. The copula representation for a joint distribution has a number of interesting properties that are helpful for efficient inference (see, for example, Trivedi and Zimmer 2005). Here we make use of the fact that a copula allows for the correlation structure to be modeled, and thus estimated, separately from the marginal distributions.

2.4. Copula Estimation

Our approach to estimating $F(\cdot \mid \boldsymbol{z})$ will be to choose parametric forms for the margins $F_i(\cdot \mid \boldsymbol{z})$ and the copula $\mathbb{C}(\cdot \mid \boldsymbol{z})$, and then find the parameters for which $\mathbb{C}(F_1(v_1 \mid \boldsymbol{z}),\ldots,F_n(v_n \mid \boldsymbol{z}) \mid \boldsymbol{z})$ is closest to $F(v_1,\ldots,v_n \mid \boldsymbol{z})$, in a likelihood sense. Specifically, suppose each margin is a distribution function with parameters $\boldsymbol{\theta}_i$, and the copula distribution belongs to a family with parameters $\boldsymbol{\phi}$. We denote the parameterized margins as $F_i(v_i;\boldsymbol{\theta}_i \mid \boldsymbol{z})$ and the parameterized joint distribution as

$$F(\boldsymbol{v};\boldsymbol{\theta},\boldsymbol{\phi} \mid \boldsymbol{z}) = \mathbb{C}(F_1(v_1;\boldsymbol{\theta}_1 \mid \boldsymbol{z}),\ldots,F_n(v_n;\boldsymbol{\theta}_n \mid \boldsymbol{z});\boldsymbol{\phi} \mid \boldsymbol{z}). \tag{1}$$

Equation 1 reveals the main advantage in using a copula model, which is that the parameters can be separated into those that are specific to one variable ($\boldsymbol{\theta}_i$) and those that are common to all variables ($\boldsymbol{\phi}$). We wish to use the available transaction data to estimate the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, for instance by solving the maximum likelihood problem

$$\left(\hat{\boldsymbol{\theta}}_{\mathrm{ML}},\hat{\boldsymbol{\phi}}_{\mathrm{ML}}\right) \in \underset{\boldsymbol{\theta},\boldsymbol{\phi}}{\operatorname{argmax}}\ \ell(\boldsymbol{\theta},\boldsymbol{\phi}),$$

where $\ell(\boldsymbol{\theta},\boldsymbol{\phi})$ is the appropriate log-likelihood function. This would require optimizing over all of the parameters simultaneously. Using a procedure called inference functions for margins (IFM) (Joe and Xu 1996), the optimization can be performed in two steps. First each margin is fit independently, and then the margin estimates are used to fit the correlation structure:

$$\hat{\boldsymbol{\theta}}_i \in \underset{\boldsymbol{\theta}_i}{\operatorname{argmax}}\ \ell_i(\boldsymbol{\theta}_i), \quad i = 1,\ldots,n \tag{2}$$

$$\hat{\boldsymbol{\phi}} \in \underset{\boldsymbol{\phi}}{\operatorname{argmax}}\ \ell(\hat{\boldsymbol{\theta}},\boldsymbol{\phi}), \tag{3}$$

where $\ell_i(\boldsymbol{\theta}_i)$ is the log-likelihood function just for item i. This procedure significantly reduces the dimensionality of the optimization problem that must be solved and makes estimation tractable on a larger scale. In general, IFM does not yield exactly the maximum likelihood estimate: $(\hat{\boldsymbol{\theta}}_{\mathrm{ML}},\hat{\boldsymbol{\phi}}_{\mathrm{ML}}) \neq (\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}})$. However, the IFM estimates $(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}})$, like the maximum likelihood estimates, are statistically consistent and asymptotically normal (Joe and Xu 1996, Xu 1996).

2.5. Estimating the Joint Valuation Distribution

The inference problem that we face here differs from a typical copula modeling problem because the distribution of interest is that over valuations, which are unobserved, latent variables. We now provide our procedure for estimating the valuation joint distribution. First, let

$$v_i^{t,\ell} = \begin{cases} -\infty & \text{if } y_i^t = 0, \\ x_i^t & \text{if } y_i^t = 1, \end{cases} \quad \text{and} \quad v_i^{t,u} = \begin{cases} x_i^t & \text{if } y_i^t = 0, \\ +\infty & \text{if } y_i^t = 1. \end{cases} \tag{4}$$

Given the observed data, these are are known quantities. Further, for any $I \in 2^{\{1,\dots,n\}}$, define

$$\tilde{v}_i^t(I) = \begin{cases} v_i^{t,\ell} & \text{if } i \in I, \\ v_i^{t,u} & \text{otherwise.} \end{cases} \tag{5}$$

The estimate for the valuation joint distribution follows in Theorem 1.

THEOREM 1. *Let $F_i(v_i; \boldsymbol{\theta}_i \mid \boldsymbol{z})$ be the parameterized marginal valuation distribution for item i, and let $F(\boldsymbol{v}; \boldsymbol{\theta}, \boldsymbol{\phi} \mid \boldsymbol{z})$ be the parameterized joint valuation distribution from (1). The estimates*

$$\hat{\boldsymbol{\theta}}_i \in \operatorname*{argmax}_{\boldsymbol{\theta}_i} \sum_{t=1}^{T} \big(y_i^t \log(1 - F_i(x_i^t; \boldsymbol{\theta}_i \mid \boldsymbol{z})) + (1 - y_i^t) \log(F_i(x_i^t; \boldsymbol{\theta}_i \mid \boldsymbol{z})) \big), \quad i = 1, \dots, n$$

$$\hat{\boldsymbol{\phi}} \in \operatorname*{argmax}_{\boldsymbol{\phi}} \sum_{t=1}^{T} \log \sum_{k=0}^{n} (-1)^k \sum_{\substack{I \subseteq \{1,\dots,n\} \\ |I| = k}} F(\tilde{\boldsymbol{v}}^t(I); \hat{\boldsymbol{\theta}}, \boldsymbol{\phi} \mid \boldsymbol{z})$$

are statistically consistent.

In Sections 2.6 and 2.7 we provide the derivation of this result, for the margin and copula parameters respectively. We show that the procedure in Theorem 1 is exactly the IFM estimation procedure given in (2) and (3), so consistency immediately follows.

2.6. Margin Likelihood and Demand Models

We first consider the margin maximum likelihood problem in (2). Let $p_i(x_i^t \mid \boldsymbol{z}^t)$ be the probability of purchase for item i at price x_i^t given covariates $\boldsymbol{z}^t$, that is, the demand model for item i. The following proposition shows an equivalence between the marginal valuation distribution function and demand models.

PROPOSITION 2. *The inverse marginal valuation distribution function for item i is identical to the demand function for item i, i.e.,*

$$p_i(x_i^t \mid \boldsymbol{z}^t) = 1 - F_i(x_i^t; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t).$$

Proof. By the rationality assumption of Section 2.1, item i is purchased if and only if $v_i^t > x_i^t$:

$$p_i(x_i^t \mid \boldsymbol{z}^t) = \mathbb{P}(v_i^t > x_i^t \mid \boldsymbol{z}^t) = 1 - F_i(x_i^t; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t).$$

□

It immediately follows that the likelihood model for the observed purchase data is

$$y_i^t \sim \text{Bernoulli}(1 - F_i(x_i^t; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t)).$$

The log-likelihood function for each margin, considering all T transactions, is then

$$\ell_i(\boldsymbol{\theta}_i) = \sum_{t=1}^{T} \left(y_i^t \log(1 - F_i(x_i^t; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t)) + (1 - y_i^t)\log(F_i(x_i^t; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t))\right). \tag{6}$$

Using this result in (2) yields the first part of Theorem 1. If $F_i(\cdot; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t)$ is linear in $\boldsymbol{\theta}_i$, for example when using a linear demand model, then the maximum likelihood problem is a concave maximization. In Section 2.8 we discuss some possible choices for the family of $F_i(\cdot; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t)$.

2.7. Estimating the Copula Parameters

Once the margin parameters $\hat{\boldsymbol{\theta}}_i$ have been estimated by maximizing (6), these estimates are used, together with the data, to obtain an estimate of the copula parameters $\boldsymbol{\phi}$. We now derive an expression for the log-likelihood of $\boldsymbol{\phi}$ to complete the IFM procedure in (3). We begin by including the latent valuations in the joint likelihood function:

$$\begin{aligned}
\ell(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) &= \sum_{t=1}^{T} \log p(\boldsymbol{y}^t \mid \boldsymbol{x}^t, \boldsymbol{z}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) \\
&= \sum_{t=1}^{T} \log \int p(\boldsymbol{y}^t \mid \boldsymbol{v}^t, \boldsymbol{x}^t, \boldsymbol{z}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) p(\boldsymbol{v}^t \mid \boldsymbol{x}^t, \boldsymbol{z}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) d\boldsymbol{v}^t.
\end{aligned} \tag{7}$$

Given $\boldsymbol{v}^t$ and $\boldsymbol{x}^t$, $\boldsymbol{y}^t$ is deterministic, with $y_i^t = 1$ if $v_i^t > x_i^t$ and 0 otherwise. Thus the integral over $\boldsymbol{v}^t$ can be limited to all $\boldsymbol{v}^t$ that are consistent with $\boldsymbol{y}^t$ and $\boldsymbol{x}^t$, meaning the integral is over $v_i^t > x_i^t$ for i such that $y_i^t = 1$, and over $v_i^t \leq x_i^t$ for i such that $y_i^t = 0$. Thus the lower and upper limits of integration are respectively $v_i^{t,\ell}$ and $v_i^{t,u}$, from (4). Also, the valuations $\boldsymbol{v}^t$ are independent of the prices $\boldsymbol{x}^t$, given covariates $\boldsymbol{z}^t$ and model parameters $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\phi}$. Thus the quantity $p(\boldsymbol{v}^t \mid \boldsymbol{x}^t, \boldsymbol{z}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi})$ is exactly the copula density function, which we denote as $f(\cdot; \hat{\boldsymbol{\theta}}, \boldsymbol{\phi} \mid \boldsymbol{z}^t)$. Continuing the likelihood expression from (7), we have then,

$$\ell(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) = \sum_{t=1}^{T} \log \int_{v_n^{t,\ell}}^{v_n^{t,u}} \cdots \int_{v_1^{t,\ell}}^{v_1^{t,u}} f(v_1^t, \ldots, v_n^t; \hat{\boldsymbol{\theta}}, \boldsymbol{\phi} \mid \boldsymbol{z}^t) dv_1^t \ldots dv_n^t. \tag{8}$$

In Lemma 1 we now show how to evaluate this integral. This formula is critical to the scalability of our inference procedure as it allows us to replace the multidimensional integral in (8) with distribution function evaluations.

LEMMA 1. *Let $f(\cdot)$ be a joint probability density function over continuous random variables $q_1, \ldots, q_n$ with the corresponding joint distribution function $F(\cdot)$. Then,*

$$\int_{q_n^\ell}^{q_n^u} \cdots \int_{q_1^\ell}^{q_1^u} f(q_1, \ldots, q_n) dq_1 \ldots dq_n = \sum_{k=0}^{n} (-1)^k \sum_{\substack{I \subseteq \{1, \ldots, n\} \\ |I| = k}} F(\tilde{\boldsymbol{q}}(I)),$$

where

$$\tilde{q}_i(I) = \begin{cases} q_i^\ell & \text{if } i \in I, \\ q_i^u & \text{otherwise}. \end{cases}$$

Proof. Define the probability events $A_i = \{q_i \le q_i^\ell\}$ for each i. Let $B = \cap_{i=1}^n \{q_i \le q_i^u\}$. Then,

$$\begin{aligned} \int_{q_n^\ell}^{q_n^u} \cdots \int_{q_1^\ell}^{q_1^u} f(q_1, \ldots, q_n) dq_1 \ldots dq_n &= \mathbb{P}\left(B \cap (\cap_{i=1}^n A_i^\mathsf{c})\right) \\ &= \mathbb{P}\left(B \cap (\cup_{i=1}^n A_i)^\mathsf{c}\right) \\ &= \mathbb{P}(B) - \mathbb{P}\left(B \cap (\cup_{i=1}^n A_i)\right) \\ &= \mathbb{P}(B) - \mathbb{P}\left(\cup_{i=1}^n (B \cap A_i)\right) \\ &= \mathbb{P}(B) - \sum_{k=1}^n (-1)^{k-1} \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \mathbb{P}(B \cap A_I) \end{aligned}$$

by the inclusion-exclusion formula, with $A_I = \cap_{i \in I} A_i$. Substituting $\mathbb{P}(B) = F(q_1^u, \ldots, q_n^u)$ and $\mathbb{P}(B \cap A_I) = F(\tilde{\boldsymbol{q}}(I))$ as defined above, we obtain the statement of the lemma. $\square$

With Lemma 1, we are now equipped to evaluate the log-likelihood expression in (8):

$$\ell(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) = \sum_{t=1}^T \log \sum_{k=0}^n (-1)^k \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} F(\tilde{\boldsymbol{v}}^t(I); \hat{\boldsymbol{\theta}}, \boldsymbol{\phi} \mid \boldsymbol{z}^t), \tag{9}$$

where $\tilde{v}_i^t(I)$ is as given in (5). Substituting this expression into (3) yields the second part of Theorem 1.

To provide some insight into this result, for the most simple case of two items in a bundle, the inner expression in (9) evaluates to

$$\sum_{k=0}^2 (-1)^k \sum_{\substack{I \subseteq \{1,2\} \\ |I|=k}} F(\tilde{v}_1^t(I), \tilde{v}_2^t(I) \mid \boldsymbol{z}^t) = \begin{cases} F(x_1^t, x_2^t \mid \boldsymbol{z}^t) & \text{if } \boldsymbol{y}^t = [0,0], \\ F_1(x_1^t \mid \boldsymbol{z}^t) - F(x_1^t, x_2^t \mid \boldsymbol{z}^t) & \text{if } \boldsymbol{y}^t = [0,1], \\ F_2(x_2^t \mid \boldsymbol{z}^t) - F(x_1^t, x_2^t \mid \boldsymbol{z}^t) & \text{if } \boldsymbol{y}^t = [1,0], \\ 1 - F_1(x_1^t \mid \boldsymbol{z}^t) - F_2(x_2^t \mid \boldsymbol{z}^t) + F(x_1^t, x_2^t \mid \boldsymbol{z}^t) & \text{if } \boldsymbol{y}^t = [1,1]. \end{cases} \tag{10}$$

The estimation procedure in Theorem 1 works for arbitrary margins $F_i(\cdot; \boldsymbol{\theta}_i \mid \boldsymbol{z}^t)$ and an arbitrary copula model $\mathbb{C}(\cdot; \boldsymbol{\phi} \mid \boldsymbol{z}^t)$. To apply these formulas to data requires choosing the distributional form of the margins and the copula family, which we now discuss.

2.8. Specifying the Margin Distributions

The connection between marginal valuation distributions and demand models given in Proposition 2 shows that the margin distribution can naturally be selected by choosing an appropriate demand model. Many retailers already use demand models for sales forecasting, and these existing models could be directly converted to marginal valuation distributions. For example, consider the linear demand model and the normal-cdf demand model. The linear demand model is $p(x_i;\boldsymbol{\theta}_i) = \min(1, \max(0, \theta_{i,1} - \theta_{i,2}x_i))$, and the corresponding valuation distribution is uniform:

$$v_i \sim \text{Unif}\left(\frac{\theta_{i,1}-1}{\theta_{i,2}}, \frac{\theta_{i,1}}{\theta_{i,2}}\right).$$

When the demand model is the normal distribution function $p(x_i;\boldsymbol{\theta}_i) = 1 - \Phi(x_i;\theta_{i,1},\theta_{i,2})$, the corresponding marginal valuation distribution is the normal distribution: $v_i \sim \mathcal{N}(\theta_{i,1},\theta_{i,2})$. For more complex demand models, including those with covariates, the corresponding valuation distribution may not be analytic. However, Proposition 2 allows for the distribution function to be evaluated, and this is sufficient for doing the estimation in Theorem 1.

2.9. Specifying the Copula

There is a large selection of copula models, which differ primarily in the types of correlation they can express. One of the most popular copula models, and that which we use in our simulations and data experiments here, is the Gaussian copula:

$$\mathbb{C}(u_1,\dots,u_n;\boldsymbol{\phi}) = \Phi(\Phi^{-1}(u_1),\dots,\Phi^{-1}(u_n);\boldsymbol{\phi}),$$

where $\Phi(\cdot;\boldsymbol{\phi})$ is the multivariate normal with correlation matrix $\boldsymbol{\phi}$, and $\Phi(\cdot)$ the standard normal. The Gaussian copula is in essence an extension of the multivariate normal distribution, in that it extends the multivariate normal correlation structure to arbitrary margins, as opposed to constraining the margins to be normally distributed. If a correlation matrix structure is not appropriate to model the dependencies in a particular application, then alternative copula models are available - see Trivedi and Zimmer (2005).

The Gaussian copula does not include the covariates $\boldsymbol{z}$ in the correlation structure, rather they would be included only in the margin demand models. In essence, this approach assumes that while the valuation for each item depends on the covariates, the correlations between valuations do not. If we wished for the correlations themselves to depend on the covariates, this could be done by parameterizing $\boldsymbol{\phi}$ to be some function of $\boldsymbol{z}$ and then solving the maximum likelihood problem over those function parameters.

3. Simulation Studies

We demonstrate the inference procedure using a series of simulation studies. We first use simulations to show empirically how the estimated parameters converge to their true values as T grows. We then use a simulated dataset to illustrate the importance of including correlations in the model.

We generated purchase data for a pair of items using uniform marginal valuation distributions and a Gaussian copula, which for two items is characterized by the correlation coefficient ϕ. The correlation coefficient ϕ was taken from $\{-0.9, -0.75, -0.5, -0.25, 0, 0.25, 0.5, 0.75, 0.9\}$ and the number of transactions T was taken from $\{100, 250, 500, 750, 1000, 1500, 2000\}$. For each combination of ϕ and T, 500 datasets

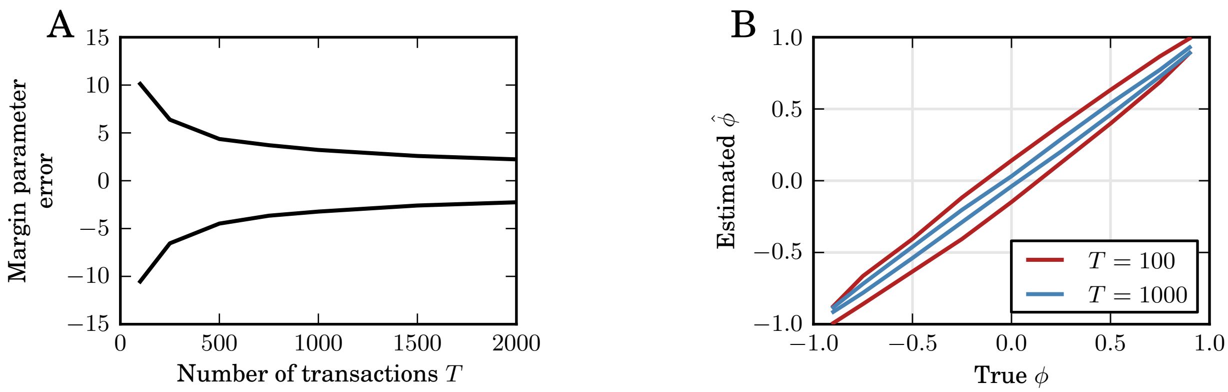
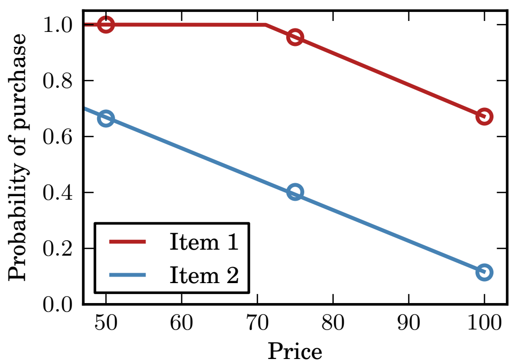


Figure 1 **Convergence of both (A) margin parameters and (B) the correlation coefficient to their true values as the number of simulated transactions T is increased. In (A), the lines indicate the first and third quartiles of the margin parameter errors across all simulations with the same number of transactions T. In (B), each pair of lines shows the first and third quartiles of the estimated correlation coefficient $\hat{\phi}$ across all simulations with the corresponding values of ϕ and T.**

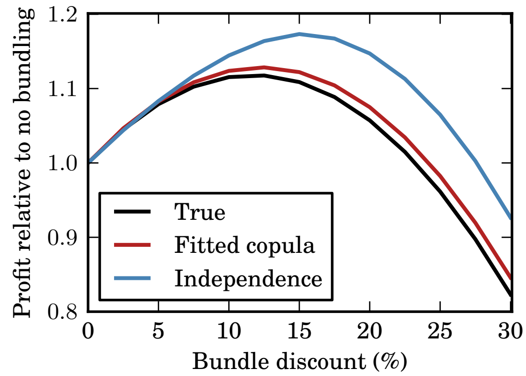
were generated, for a total of 31,500 simulated datasets. For each dataset, the margin parameters $v_{\min}$ and $v_{\max}$ for each of the two uniform valuation distributions were chosen independently at random, to allow the simulations to capture a large range of margin distributions. The parameter $v_{\min}$ was chosen from a uniform distribution over $[-25, 75]$ and $v_{\max}$ chosen from a uniform distribution over $[100, 200]$. For all simulations, the transactions were spread uniformly across three price points, with the prices of the two items taken to be 100 for one third of transactions, 75 for one third, and 50 for the remaining third. In each simulation, the copula defined by the combination of the correlation coefficient and the margin parameters was used to generate T sets of valuations for the items. These valuations were combined with the prices under the rationality assumption of Section 2.1 to produce binary purchase data.

We applied the inference procedure in Theorem 1 to the transaction data, with the goal of recovering the true, generating copula model. Figure 1 shows that as the number of transactions grows, both the margin estimates and the correlation coefficient estimates converge to their true values. This holds for the full range of possible values of the correlation coefficient. In these simulations, only a few thousand samples were required to recover the true distribution with high accuracy, suggesting that these techniques are not limited to retailers with very large datasets.

To further illustrate the simulation results, we selected at random a simulated dataset with $T = 2000$ transactions and $\phi = 0.5$. We show in Figure 2(a) the fitted margins for this particular simulated dataset. The estimated correlation coefficient, found by maximizing (9), was 0.48. To illustrate the potential profitability of bundling, in Figure 2(b) we held the item prices at 100 and set the cost per item to the retailer to a 50% markup, meaning, sales price 50% higher than the retailer's cost. We show for a range of bundle discounts the profit relative to the profit obtained in the absence of a bundle discount. The estimated distribution



(a) Demand models for each of the two items. The circles give the empirical purchase probabilities measured from the data, and the lines show the fitted margin distribution function.



(b) Change in relative profits from introducing the bundle at a particular discount relative to the sum of item prices, as estimated from the true distribution, the fitted copula model, and a distribution using the fit margins but assuming independence.

Figure 2 A simulation experiment which demonstrates the importance of modeling valuation dependence

is very close to the true distribution, and both reveal that offering a bundle discount of about 12% will increase profits by about 10%. Using the same estimated margins but assuming independence to obtain a joint distribution yields very different results. This example highlights the importance of accounting for correlations in valuations when estimating the response to bundle discounts.

4. Estimation for Bundles with Intra-Category Substitution

In the previous section, we have proposed an inference procedure for the base scenario, where only a single collection of items is offered. In many practical settings, bundles are made up of products from different categories (e.g., phone and data plans), and there are competing alternatives from within each category (e.g., different phone models, data plans with different usage amounts). In such a scenario, each individual component experiences substitution effect that comes from similar products in the same category. At the same time, its sales could also be affected by products from different categories due to their correlated valuations.

A popular method to model demand of competing products is discrete choice modeling: a typical individual chooses an option out of a set with a finite number of alternatives. The individual is assumed to choose the one that maximizes his utility over the set of alternatives. Suppose there are N categories and each category l consists of n_l competing products. Under mixed bundling, there are a total of $\prod_{l=1}^{N}(n_l+1)$ possible options that a customer can select. It is not surprising to see that discrete choice models are unable to cope with the dimension of the choice set for most realistic applications with bundles.

Alternatively, we can describe a joint distribution over individual products using a Gaussian copula. To do so, we need to estimate only $\sum_{l=1}^{N} n_l$ marginal valuation distributions and a covariance matrix of size $\sum_{l=1}^{N} n_l$ that describes the pairwise dependence of products within and across the categories. In this section, we propose a procedure that adapts the inference algorithm discussed in Section 2 and leverages discrete choice models which readily capture the substitution effect within a category.

4.1. Valuations for Competing Products Within a Category

Discrete choice modeling is a popular method for estimating demand in markets where there are differentiated products and customers make trade-offs by choosing one of the competing products. In a choice situation, a representative consumer, labeled t, faces a choice among n alternatives. She obtains a certain amount of utility from each alternative i and selects the alternative with the maximum utility. Utility from each alternative i can be decomposed into a deterministic part and an error component: $\mu_i^t = \nu_i^t + \varepsilon_i^t$. The deterministic part of the utility is often specified as a linear combination of features. The error term ε_i^t, depending on different specifications on its probability density, gives rises to different discrete choice models.

The multinomial logit model (MNL), arguably the most popular model for having tractable closed-form choice probabilities, assumes that ε_i^t is independently, identically distributed extreme value. More recently, McFadden and Train (2000) propose mixed logit model (ML) and show that under mild conditions, any discrete choice model derived from random utility maximization generates choice probabilities that can be approximated, arbitrarily closely, by a ML model. It obviates the limitations of standard logit by allowing for random taste variations, unrestricted substitution patterns and correlations in unobserved factors over time.

Valuation refers to the maximum price that a customer is willing to pay for a product, which is also known as the reservation price of a product. A consumer compares his valuation for each product with its purchase price and chooses the product that offers the highest surplus. Surplus maximization, from the reservation price approach, and utility maximization, from the discrete choice model, are analytically equivalent if price enters the utility function linearly (Ratchford 1979, Srinivasan 1982).

Given a set of features, we can express the utility as $\mu_i^t = \bar{\mu}_i^t + \alpha x_i^t + \varepsilon_i^t$, where $\bar{\mu}_i^t$ is the utility excluding the price impact and α is the price coefficient. By its definition, valuation is the measure of preference in dollar terms that *does not* take price into account (Kalish and Nelson 1991, Kohli and Mahajan 1991). Thus, it can be expressed as the following:

$$v_i^t = \frac{1}{|\alpha|}(\mu_i^t - \alpha x_i^t + \varepsilon_i^t) = \frac{1}{|\alpha|}(\bar{\mu}_i^t + \varepsilon_i^t). \tag{11}$$

With the standard MNL model where ε_i^t are distributed iid extreme value, valuations follow the same distribution, with a scaled variance $\pi^2/(6\alpha^2)$, as the variance of the extreme value term ε_i^t is $\pi^2/6$. Since the valuations are uncorrelated across the alternatives, the covariance matrix for product valuations is a diagonal matrix. Moreover, since the error components ε_i^t are identically distributed, the diagonal elements on the covariance are also identical.

With the ML model, the utility can be specified as $\mu_i^t = \bar{\mu}_i^t + \alpha x_i^t + u_t' z_i^t + \varepsilon_i^t$, where u_t is a random vector with zero mean and covariance W. We denote the unobserved portion of the utility as $\eta_i^t = u_t' z_i^t + \varepsilon_i^t$. For the MNL model, u_t is zero, so that there is no correlation in utility over alternatives which gives rise to the IIA property and its restrictive substitution patterns. With nonzero error components, utility is correlated over alternatives: $\text{Cov}(\eta_i^t, \eta_j^t) = E(u_t' z_i^t + \varepsilon_i^t)(u_t' z_j^t + \varepsilon_i^t) = (z_i^t)' W z_j^t$. Utility is correlated over alternative even when, as in most specifications, the error components are independent, i.e, W is diagonal.

The valuation derived from a ML model can be expressed as

$$v_i^t = \frac{1}{|\alpha|}(\bar{\mu}_i^t + u_t' z_i^t + \varepsilon_i^t). \tag{12}$$

Therefore, it follows a distribution which is a sum of $u_t' z_i^t$ and ε_i^t. When $u_t \sim N(0, W)$ with diagonal W, $u_t' z_i^t$ is a sum of independent normal random variables of zero mean, which is still normal with zero mean. The covariance across alternatives is given by $\text{Cov}(v_i^t, v_j^t) = \frac{1}{\alpha^2}(z_i^t)' W z_j^t$, for $i \neq j$, and $\text{Var}(v_i^t) = \frac{1}{\alpha^2}((z_i^t)' W z_j^t + \frac{\pi^2}{6})$. Unlike the diagonal matrix for the MNL model, the covariance matrix derived from the ML model can be a full matrix, which captures the pairwise dependence across competing products in a category.

4.2. Pairwise Copula Inference for Products Across Categories

From the discrete choice model estimated for each category, we can estimate the marginal valuations for individual products. With this information, we can then infer the correlation of components that are from different categories. To make it more concrete, let us assume there are two categories: m products from category A and n products from category B. This gives rise to mn unique bundles. For each bundle of item i from category A and item j from category B, we use a modified inference procedure to determine the pairwise correlation ϕ_{ij}, while holding the valuations of other products fixed.

We propose a heuristic to obtain the pairwise correlation. The key modification to the algorithm comes from the fact that individual products in a bundle compete with other substitutable products in their respective categories. To see this, consider a product category, for a given price vector x, the mean surplus associated with every alternative i for consumer t is given by $s_i^t = E[v_i^t - x_i^t] = E[v_i^t] - x_i^t$, where v_i^t is valuation of product i from Equation 11 or 12 depending on the choice model. Denote the highest average surplus in the category *excluding* product i as $s_{-i^*}^t = \max_{k \neq i} s_k^t$.

We only focus on one product in one category at a time and we fix the valuations of all other alternatives from the same category at their mean values. Thus, the reference $s_{-i^*}^t$ is a constant and the only uncertainty comes from the valuation of product i. If $y_i^t = 1$, i.e., product i is selected by customer t among the alternatives, it implies that $v_i^t - x_i^t > s_{-i^*}^t$ and 0 otherwise. To obtain the probability of the appropriate event, one needs to integrate over $v_i^t > x_i^t + s_{-i^*}^t$ for $y_i^t = 1$, and $v_i^t \leq x_i^t + s_{-i^*}^t$ for $y_i^t = 0$. Note that Lemma 1 continues to hold, so the multidimensional integral can be deposed into a series of rectangular cumulative density valuations that can be done efficiently, as the only difference is that the limit of the integral is changed from x_i^t to $x_i^t + s_{-i^*}^t$.

We use an example to illustrate the modified inference algorithm. For simplicity, we will drop the index t denoting consumers. Consider a bundle made up of product i from category A and j from category B. We will use superscript to denote the category. Let $F_i^A(\cdot)$ and $F_j^B(\cdot)$ to denote the marginal distribution for product i and j from their respective category and let $F_{ij}(\cdot)$ to represent the joint bivariate distribution. x^A and x^B represents the price vectors and where $y = (y_i^A, y_j^B)$ where $y_i^{(\cdot)} = 1$ when product i from that category is purchased. The logliklihood expression derived in Equation 10, which maps each transaction iento one of the four possibilities becomes the following,

$$\begin{cases} F_{ij}(\bar{x}_i^A, \bar{x}_j^B) & \text{if } y = [0,0], \\ F_i^A(\bar{x}_i^A) - F_{ij}(\bar{x}_i^A, \bar{x}_j^B) & \text{if } y = [0,1], \\ F_j^B(\bar{x}_j^B) - F_{ij}(\bar{x}_i^A, \bar{x}_j^B) & \text{if } y = [1,0], \\ 1 - F_i^A(\bar{x}_i^A) - F_j^B(\bar{x}_j^B) + F_{ij}(\bar{x}_i^A, \bar{x}_j^B) & \text{if } y = [1,1], \end{cases} \tag{13}$$

where $\bar{x}_i^A = x_i^A + s_{-i^*}^A$ and $\bar{x}_j^B = x_j^B + s_{-j^*}^B$. The key difference between Equation 10 and 13 is that the reference level for surplus is shifted from 0 to $s_{-i^*}^A$ and $s_{-j^*}^B$ respectively, which represent the mean surplus associated with choosing the best alternative in the remaining category excluding i and j.

With mn unique bundles constructed from two categories, we need to run this inference procedure for every bundle to estimate mn correlation coefficients ϕ_{ij} which describe their dependence.

4.3. Monte-Carlo Pricing Optimization

To specify the joint distribution of mn products using a Gaussian copula, one needs the marginal distribution and the covariance matrix of size $(m+n) \times (m+n)$ which describes the dependence structure. When demand is estimated with discrete choice models, marginal valuation distributions can be easily identified. Our approach breaks the covariance matrix into smaller block matrices (which corresponds to categories) and estimate them separately.

Suppose the covariance matrix has first m elements correspond to products from category A and the next n elements represent products from the other category. The diagonal block matrices of size $m \times m$ and $n \times n$ represent the covariance within the categories. The off-diagonal block matrix of size $m \times n$ is obtained by using the inference procedure described in Section 4.2. When the standard MNL is used to model demand for both categories, these sub-matrices which describe the correlation within the category are diagonal with two values which indicate variances in the two categories. If a ML model is used instead, the full block matrix is expected. Note that since the covariance matrix is obtained via joining sub-matrices, it is possible that the matrix fails to satisfy properties such as being positive semi-definite. When such situation arises, there are techniques available to obtain the "nearest" covariance matrix which satisfies the desirable properties (e.g., Higham 2002, Borsdorf et al. 2010).

After we obtain a joint distribution for individual products, except for very small problem instances, it can be very complex to characterize the expected profit function generated by all bundles and their individual products. One can turn to Monte-Carlo methods. For example, one can draw K samples from this joint distribution where each sample represents a customer's valuations towards m and n products. Surplus for each product is the difference between the realized valuations and the prices which are decision variables. Constraints are set such that a customer selects at most one product from each category and chooses the products or bundle that gives her the maximum surplus. The expected profit is simply the average profit generated by K customers. Such studies have been done for conjoint analysis, except the valuation are solicited directly from experiments instead of estimated from sales data.

5. Case Study: Bundling in the Telecommunication Industry

To empirically investigate the bundle pricing problem, we obtained a data set from a large anonymous telecommunication service provider, hereafter called "Provider X". The dataset includes a sample of individual transactions between April 1, 2012, and November 30, 2013. All transactions involves signing a two-year service contract at a fixed rate, in exchange for a combination of services including voice, text messages and/or data. When signing up a service contract, a customer has the option of purchasing a phone at a discounted price. A service contract and a phone purchase is referred to be a *bundle* in this analysis.

Among all contracts offered by Provider X, the default option comes with "unlimited talk and texts". The key differentiator lies with the data plan included in the contract. The primary goal of the study is to help Provider X to have a better understanding on how sales of phones are related to sales of data plans and how bundle prices impact its business.

In the following sections, we begin by estimating two econometric models that model the demand for phones and data plans respectively. Next, we infer pairwise correlations between consumers' valuations on individual phones and data plans. Lastly, we perform counterfactual experiments to generate insights on pricing strategies for some popular bundles in the dataset.

5.1. Demand Modeling for Smartphones

The smartphone industry is a fast-paced, dynamic environment, which is characterized by quickly evolving technologies and designs, short product lifecycle and rapid imitation of technological advancements. During the study period, over 400 unique mobile phones across 15 brands are recorded. Two brands, namely, Apple and Samsung dominate the sales. Their monthly sales are shown in Figure 3. The combined volume of the two brands consistently exceeds 80% of the total sales. Within the two brands, Apple had a greater market share, approximately doubling Samsung's volume. Figure 3 reveals significant correlations between the two brands, implying intense competition between them.

All of the top ten best selling phones during the study period come from these two brands. Figure 4 shows their monthly sales. The graph prominently depicts the strong substitution effect within a brand driven

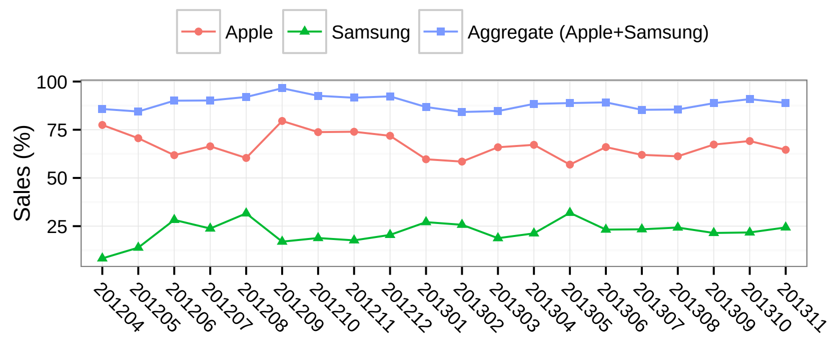


Figure 3 **Monthly sales of the two major mobile phone brands and their aggregate sales as a percentage of the total mobile phone sales.**

by product cannibalization, i.e., newer models compete and hurt the sales of older models from the same brand. For instance, when the latest version of Apple's iPhone was released in September of 2012 and 2013 respectively, the sales of the previous models decreased sharply.

With the availability of many closely substitutable choices, each consumer will weigh the improved performance of a new product against the benefits of older products (e.g., lower prices) and then make a choice as to which one to purchase. There is the possibility that a customer who prefers a product with a high quality level will choose to purchase a product with low quality level because of its lower price. Understanding the underlying decision making process is crucial to intelligent pricing decisions. Next, we will present a discrete choice model which intends to shed some light on consumers' choices on phone purchase by incorporating heterogeneous consumer preferences.

5.1.1. Econometric Model We use a random-coefficient logit model to estimate the distribution of consumer preferences towards different phone characteristics. In our model, the utility for consumer t from choosing phone i can be represented as

$$\mu_i^t = \beta_t' X_i^t + \gamma' Y_i^t + \varepsilon_i^t, \tag{14}$$

where X_i^t and Y_i^t are vectors of observable characteristics that relate to phone i and consumer t, and ε_i^t is an unobserved random term that is distributed iid extreme value, independent of other terms. The coefficients β_t are random and vary over consumers in the population with density $f(\beta)$, parameterized by its mean b and covariance W. The coefficients γ represent a vector of fixed coefficients with respect to observable characteristics Y_i^t.

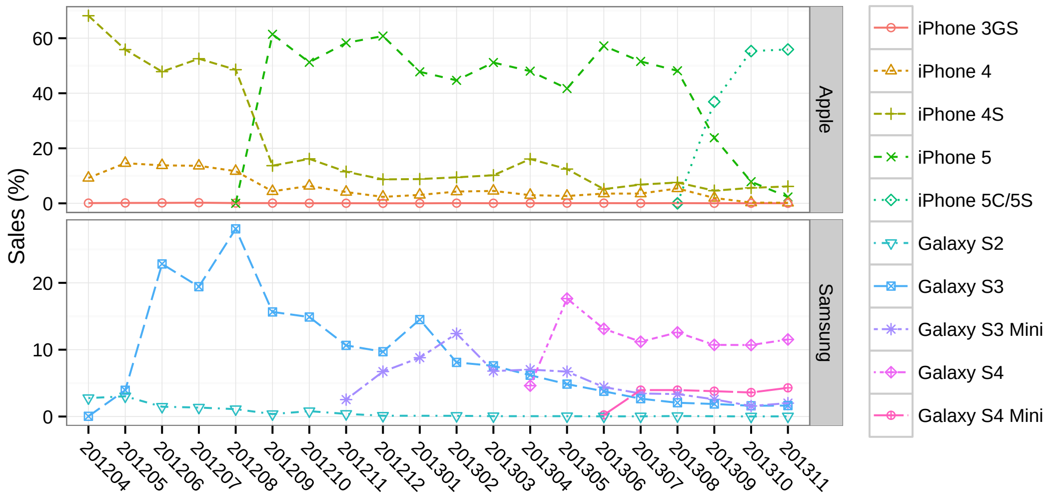


Figure 4 **Monthly volume market share of the best selling phones from Apple and Samsung.**

Since ε_i^t is iid extreme value, the choice probabilities would be a standard logit if the value of β_t is given: i.e., $L_i^t(\beta_t) = \frac{e^{\beta_t' X_i^t}}{\sum_k e^{\beta_t' X_i^t}}$. Since β_t is unknown, the unconditional choice probability is the integral of $L_i^t(\beta_t)$ over all possible variables of β_t weighted by the density of $f(\beta)$:

$$P_i^t = \int L_i^t(\beta_t) f(\beta) d\beta = \int \frac{e^{\beta' X_i^t}}{\sum_k e^{\beta' X_i^t}} f(\beta) d\beta. \tag{15}$$

We assume that the random coefficients are independently normally distributed in the population[1]. That is, $\beta_t \sim N(b, W)$ with diagonal W. As shown in Section 4.1, even with this specification, utility is correlated over alternatives.

5.1.2. Choice Set There are over 400 unique SKUs in the data set. The large number of available alternatives create computational problems to estimate the discrete choice models (e.g., Allenby and Rossi 1991, Erdem and Keane 1996). A common approach is to model the decision at the brand level. While it is plausible, such a model will not be able to explain the cannibalization effect within a brand which is prominently featured in Figure 4. An alternative approach that deals with large choice set focuses on a subset of the SKUs that are frequently purchased (e.g., Fader et al. 1992, Siddarth et al. 1995). However, this approach can only tease out the impact of time-varying attributes (e.g, price and discount) and time-invariant attributes regarding phone characteristics (e.g., screen size, camera resolution, memory etc.) will be absorbed into the SKU-specific constant during estimation.

Taking these factors into consideration, we aggregate the SKUs of the most popular brands (i.e., Apple and Samsung) based on the "product line" and "generation" and consider the remaining SKUs at the major

[1] For simplicity, we assume that standard deviations in the vector W are not correlated.

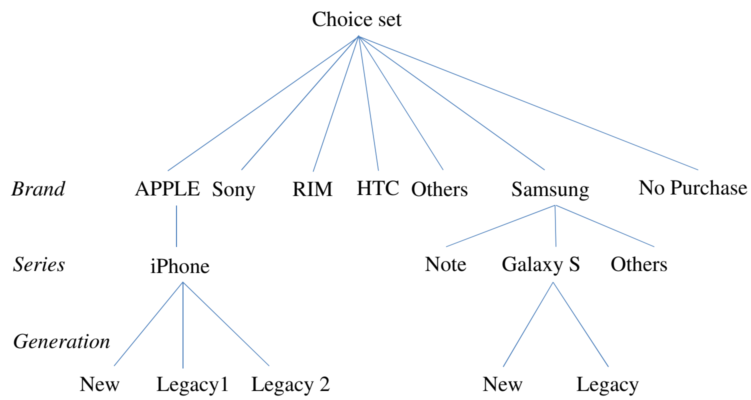


Figure 5 Choice set for the demand model for phone.

brand level. To be precise, consider Apple which only has one phone line, iPhone. We use *iPhone New*, *iPhone Legacy 1* and *iPhone Legacy 2* to represent to the newest model, its predecessor, and the model prior to the predecessor. For example, just before the launch of iPhone 5S/5C in September 2013, *iPhone New*, *iPhone Legacy 1* and *iPhone Legacy 2* represent iPhone 5, 4S and 4 respectively. Meanwhile, after the launch, the three types refer to iPhone 5C/5S, 5 and 4S.

Contrary to Apple's single series, Samsung has introduced a variety of the devices that cover a wide range of price points. Its flagship series, S-series (e.g., Galaxy S2, S3 and S4), is the most popular line. For this series, we will use *Samsung S New* and *Samsung S Legacy* to denote the latest model and its predecessor respectively. Samsung also introduces a series of high-end smartphones and tablets, known as *Samsung Note* series. The remaining Samsung SKUs are grouped under *Samsung Others*, which primarily consist of phones from the S Mini series. It is a lower-range model of the flagship S series, with a similar but noticeably inferior hardware design and software features to its high-end counterpart.

Besides Apple and Samsung phones, phones from *HTC*, *Sony* and *RIM* (manufacturer of BlackBerry) are considered at their respective brand level. SKUs from all other brands are aggregated under a composite alternative called *Others*. Besides alternatives which indicate phone purchases, we also include *No purchase* as an alternative, which primarily occurs with renewal contracts when the customer does not replace his phone. Thus, in total, we model twelve alternatives in each choice decision. The choice set is depicted in Figure 5, where each leaf node represents an alternative.

5.1.3. Attributes We capture an exhaustive list of phone characteristics in the choice modeling and the primary source for the information is based on the database maintained at www.phonearena.com. Table 1 shows the summary statistics of the phone characteristics used in our model. The observed phone characteristics include the following:

- Phone body: Volume and weight
- Screen: Size and resolution

- Platform: Operating system
- Processing capability: Clock speed for its CPU (central processing unit) and the number of independent processing units
- Memory and storage: RAM, internal storage, and external storage
- Camera: Primary (rear) and secondary (front) camera resolution
- Battery: Capacity in terms of continuous talk time
- Age: Number of months since its release date

We measure a phone's size in inch^3 (called *Volume*) and in grams (called *Weight*). Two factors are taken into account when we compare the touchscreen of a phone: 1) the size, and 2) the resolution. The size is measured diagonally in inches (called *Screen size*). The resolution typically refers to the number of pixels displayed on the screen. Since screen sizes vary across phones, we measure the resolution by pixels per inch (called *Screen resolution*), which is defined as the resolution in pixels along the screen diagonal normalized by the screen size. Take Samsung S4 as an example, which has a 5-inch screen with 1080 $\times$1920 pixels. Its resolution is computed as $\sqrt{1080^2 + 1920^2}/5$, approximately 441 PPI.

Mobile operating systems combine the features of a personal computer operating system with other features, including a touchscreen, cellular, Bluetooth, Wi-Fi, GPS mobile navigation, camera, speech recognition, music player etc. While a number of mobile operating systems has been introduced, the market has gone through consolidations with only a few dominant players remaining[2]. We create a categorical variable called *OS*, which is 1 if the phone runs on Android OS, 2 for iOS (which is Apple's proprietary operating system), and 0 for others (e.g., Symbian, BlackBerry, Windows Mobile etc.).

The processing power of a phone is measured by the clock speed of its CPU (called *CPU Speed*) in gigahertz (GHz) and the number of independent processing units (called *Core*). Older phones only have a single-core processor. Beginning in 2012, many smartphone manufacturers began to offer devices with dual-core processors, and not long after that, quad-core processors. In theory, the more processors a phone has, the faster it is, as they can divide a task for parallel processing.

Like CPU processing power, memory and storage capacity are also critical to the performance. We measure the internal memory a smartphone by the size of its RAM in gigabytes (GB) (called *RAM*). In general, the larger the capacity the more information can be stored and accessed quickly. Internal storage is used to store the operating system and critical programs. We measure this feature in gigabytes (called *Internal Storage*). Besides internal storage which is built-in and non-removable, certain phones also include a microSD card slot which enables external storage. The capacity usually increases in a factor of 8GBs, i.e., 16GBs, 32GBs and so on. We use a variable (called *External Storage*) to indicate the maximum allowable microSD expansion. For phones which do not have a microSD card slot (e.g., iPhone), this variable is set to 0.

[2] http://www.emarketer.com/Article/Android-Apple-Continue-Consolidate-US-Smartphone-Market/1010196#sthash.C0ZF8TtP.dpuf

Table 1 Summary Statistics

	Total		Apple		Samsung	
	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev
Volume (in^3)	3.63	1.31	3.49	0.22	4.93	0.65
Weight (100 g)	1.12	0.37	1.18	0.11	1.30	0.19
Screen size (in)	3.79	1.22	3.90	0.20	4.68	0.49
Screen resolution (100 PPI)	3.02	1.02	3.27	0.02	3.21	0.84
OS: Andriod =1	0.29	0.45	0	0	1	0
OS: iOs = 2	0.68	0.47	1	0	0	0
CPU Speed (GHz)	1.29	0.46	1.39	0.26	1.39	0.27
Core	2.08	1.10	1.80	0.40	3.38	0.92
RAM (GBs)	0.98	0.50	0.90	0.20	1.40	0.53
Internal Storage (8 GBs)	1.94	1.10	2.22	1.02	1.67	0.51
External Storage (8 GBs)	1.84	3.20	0	0	6.97	1.75
Primary Camera (MP)	7.45	2.92	7.84	0.67	8.74	3.05
Secondar Camera (MP)	1.08	0.63	1.02	0.36	1.51	0.73
Talk Time (Hours)	8.67	3.93	8.04	0.49	12.50	3.93
Age (Months)$^{(L)}$	1.85	0.87	2.14	0.68	1.8	0.74
Age Square (Months)$^{(L)}$	3.42	1.84	4.00	1.51	3.26	1.66
Phone Price (US$100)	4.69	1.55	5.28	0.39	4.77	1.09

Note. The sample period is from April 1, 2012 to November 30, 2013. $^{(L)}$Logarithm of the variable.

As smartphones become ubiquitous, they are quickly replacing traditional point-and-shoot cameras as the go-to devices for taking photos. Many smartphones are equipped with two cameras, a primary camera located at the back of a phone and a camera with lower resolution on the front of the device for video calling. We measure the resolution of the primary and the secondary camera (called *Primary Camera* and *Secondary Camera* respectively) in megapixels (MP).

In terms of battery capacity, the actual usage depends on many other features of the phone such as screen size, processors etc. Thus, we measure the battery life a phone in terms of talk time in hours, which is the longest time that a single battery charge will last when a user is constantly talking on the phone on a 3G network (called *Talk Time*).

We measure the age of a phone based on the number of months since it was released (called *Age*). It is conceivable that there exists a relationship between the age of a phone and demand due to increased awareness effects. We also include the squared number of months to test for nonlinear effects (called *Age Square*).

In addition to phone characteristics as features, we also incorporate transaction level attributes. We include the price paid for the phone in US dollars (called *Phone Price*) and an indicator variable to represent the type of contract (called *Contract Type*). It represents a renewal contract when *Contract Type* is 1, and a new contract when this variable is 0.

5.1.4. Estimation Results We first estimated a standard MNL model without random coefficients. We used the Lagrange multiplier test on this model (McFadden and Train 2000) which indicated that there were significant random components for eight of the variables. These variables (which are denoted as $X_i^{\tau t}$ in Equation 14) enter the non-stochastic portion of the utility in the mixed logit model. Both models are estimated using simulated maximum likelihood method. The results are in shown Table 2, where Column 1 gives the estimated parameters and standard errors for the MNL model, and Column 2 represents the mixed logit with the same specification.

In the first panel in Table 2, estimates of the mean utility levels for each feature are presented. The signs of the estimated coefficients in the two models completely agree with each other. Most coefficients have expected signs. However, variable *Screen size* and *RAM* have negative coefficients. It is in fact not surprising, noting that the most popular phone in the data set, the iPhones, has smaller screens and lower RAM relative to other models in the market (see Table 1). *Age* and *Age Square* have positive and negative coefficients respectively, implying that consumer's utility of a phone follows an inverted-U shape with respect to time.

In the second panel, with the mixed logit model, the distribution of parameters across consumers captures the effects of heterogeneity around the mean utility level for each phone characteristic due to the unobserved features. The estimated standard deviations of coefficients are highly significant for six of the variables (weight, screen size, screen resolution, primary camera resolution, battery power and the age of the phone), indicating that those parameters do indeed vary in the population. The standard deviations for camera resolution and screen size coefficients are quite large relative to the estimated means, indicating a wide range of positive and negative preferences for these features. For example, the distribution of the coefficient of *Primary Camera* obtains an estimated mean of 0.246 and estimated standard deviation of 0.184, such that 91% of the distribution is above zero and 9% below. This implies that 91% of the population prefer having a higher resolution camera while the remaining 9% do not care about this feature. The distribution of the coefficient of *Screen size* implies that having a larger screen on a phone provides positive inducement for 15% of the population and is deemed as a negative factor by the rest.

The third panel in Table 2 shows the alternative specific constants with respect to *No purchase*, which is the reference level for the choice set. The constants measure the mean intrinsic values of different alternatives that are not captured by the model specifications. The rankings of the estimated values are consistent in the MNL and the mixed logit model. For instance, *Samsung Note* and *HTC* has the highest and the lowest estimates among all the alternatives. One could argue that for the alternative specific constants reflect the additional value that consumers place on a "premium" product from a reputable brand versus that for a less well-known company, excluding the phone specifications that are currently captured. Within Apple and Samsung's S-series, we see that the estimates for the older models are lower than that of the latest models.

The likelihood ratio statistic for mixed logit versus MNL is 194.62 with eight degrees of freedom, which is highly significant, indicating that the explanatory power of the mixed logit model is considerably greater than with standard logit. The main advantage of a mixed logit model is that it avoids the independence of irrelevance (IIA) property of the standard MNL model, as illustrated in the following example. If we assume

Table 2 Main Estimation Results

	Coefficient (std. err.)	Coefficient (std. err.)
Variable	Multinomial Logit	Mixed Logit
	Mean effects	
Phone Price	−3.748 (0.089)***	−4.404 (0.135)***
Volume	−4.019 (0.195)*	−5.958 (0.333)***
Weight	0.033 (0.005)**	0.079 (0.008)**
Screen size	−0.809 (0.101)***	−0.835 (0.112)***
Screen resolution	1.559 (0.091)***	1.851 (0.104)***
OS: Andriod = 1	1.415 (0.088)***	1.689 (0.249)***
OS: iOS = 2	1.462 (0.087)***	1.738 (0.250)***
CPU Speed	0.432 (0.074)***	0.640 (0.093)***
Core	2.609 (0.072)***	3.210 (0.111)***
RAM	−2.937 (0.141)***	−3.479 (0.181)**
Internal storage	1.874 (0.040)***	2.259 (0.067)***
External storage	0.270 (0.016)***	0.169 (0.021)***
Primary camera	0.221 (0.013)***	0.246 (0.017)***
Secondary camera	−0.324 (0.053)*	−0.092 (0.061)
Talk time	0.300 (0.014)***	0.384 (0.018)***
Age$^{(L)}$	4.518 (0.280)***	5.778 (0.399)***
Age Square$^{(L)}$	−1.926 (0.123)***	−2.383 (0.172)***
	Distribution of parameters across consumers	
Weight		0.011 (0.002)***
Screen size		0.806 (0.077)***
Screen resolution		0.675 (0.066)***
Primary camera		0.184 (0.013)***
Core		0.006 (0.057)
Internal storage		0.025 (0.074)
Talk time		0.106 (0.013)***
Age$^{(L)}$		0.335 (0.050)***
	Alternative specific constants	
iPhone New	15.796 (0.587)***	23.412 (1.011)***
iPhone Legacy	14.668 (0.586)***	22.010 (1.003)***
iPhone Legacy 2	13.795 (0.590)***	20.859 (1.001)***
Samsung Note	19.619 (0.979)***	28.290 (1.553)***
Samsung S New	11.879 (0.725)***	20.106 (1.213)***
Samsung S Legacy	11.039 (0.714)***	19.354 (1.208)***
Samsung Others	14.356 (0.738)***	23.590 (1.256)***
Sony	8.737 (0.672)**	16.324 (1.115)**
HTC	7.236 (0.664)**	13.732 (1.066)**
RIM	14.686 (0.727)***	23.097 (1.263)***
Others	13.659 (0.725)*	20.820 (1.183)**
Log-Likelihood	−76,075.160	−75,977.850

Number of observations = 44,648 *p<0.1; **p<0.05; ***p<0.01

Note. The reference level for the choice set is *No purchase*. The reference level for OS is 0 = Others. Coefficients for fixed effects of *Contract Type* are omitted due to brevity. $^{(L)}$Logarithm of the variable.

Table 3 **Market share changes with respect a 10% discount on *Sony* phones**

	Original	MNL Predicted	MNL Change (%)	ML Predicted	ML Change (%)
Sony	2.36	12.26		12.19	
iPhone New	49.50	44.57	-9.96	44.47	-10.17
iPhone Legacy 1	10.06	9.05	-10.02	9.03	-10.17
iPhone Legacy 2	3.92	3.53	-9.95	3.64	-7.19
Samsung Note	1.87	1.69	-10.02	1.62	-13.45
Samsung S New	11.32	10.13	-10.48	9.48	-16.28
Samsung S Legacy	2.09	1.86	-10.96	1.86	-10.99
Samsung Others	6.26	5.62	-10.29	5.77	-7.87
HTC	2.14	1.89	-11.63	1.92	-10.26
RIM	1.40	1.26	-10.29	1.26	-10.10
Others	1.04	0.93	-10.49	0.93	-11.19
No Purchase	8.03	7.21	-10.21	7.83	-2.38

a 10% discount has been awareded to *Sony* phones, we can use the estimated MNL and the mixed logit models to predict the market impact. Table 3 depicts the original market share of all the alternatives and the corresponding changes predicted by the two models. In fact, the predicted demand for Sony phones given by both models are rather similar (i.e., 12.26% versus 12.19%). For the MNL model, the price decrease for Sony phones results in a reduction of approximately 10% in sales of all other alternatives. The proportional substitution pattern is expected because of the IIA property. On the other hand, the prediction of the mixed logit model highlights disproportional substitution pattern. For instance, the phone that experiences the highest demand decrease is *Samsung S New*. This particular series and *Sony*'s most poplar products share many similar characteristics, such as Android OS, large screens and high resolution cameras etc. One can even argue that the two brands are among the largest TV and monitor manufacturers. Meanwhile, the impact experienced by Apple is much smaller, presumably because the two brands use different OS, reducing the competition between them. Among all the alternatives, the least affected alternative is *No purchase*, which is also quite intuitive. The fact that the mixed logit model is capable of capturing more flexible and realistic substitution patterns makes it a superior choice than the standard MNL model in our study.

5.2. Demand Modeling for Data Plans

To model data plans, we consider two alternatives: we use *Base* plan to denote the default option that comes with the "unlimited minutes and text" contracts, and use *Upgraded* plan as a composite alternative which represents the plans with more data than the default option. In the data set, there are some transactions with data lower than *Base* plan. These transactions are known as the "grandfathered" plans, which have different terms from the more recent contracts. Since they merely make up about 0.6% of total sales, we remove them from the data set.

Table 4 shows the summary statistics for the data plans. To protect confidential data, we normalize the revenue with respect that of *Base* plan. Table 4 shows that the vast majority of consumers select *Base* plan.

Table 4 Summary statistics for data plans

	Sales (%)	Relative Revenue	Coef. Variation on Revenue
Base plan	96.91	1	0.125
Upgraded plan	3.09	1.24	0.250

Upgraded plan, which has a markup of 24% over the default plan has a very tiny market share of 3%. One of the our research goals is to provide pricing recommendations which lead to more consumers switching to the more expensive plan.

We construct a discrete choice model to fit the demand of data plans. Utility of purchasing a particular plan is a linear function of *Price* (US\$100), *Quota* (GB) which indicates the data usage included in the plan, and *Contract Type*. Table 5 gives the estimate of the MNL model using transaction data. Although the Lagrange multiplier test found significant error components for *Price*, we were unable to estimate any mixed logit model with log-likelihood significantly better than the MNL model in Table 5.

Table 5 Estimation result on data plans

	Coefficient (std. err.)
Upgraded plan: Constant	-5.747 (0.425)***
Price	-1.459 (0.379)***
Quota	1.931 (0.261)**
Upgraded plan: Contract Type (1=Renewal)	0.527 (0.150)***
Log Likelihood	-8818.1
Number of observations = 38,747	*p<0.1; **p<0.05; ***p<0.01

Note. The reference level for the choice set is *Base Plan*.

The estimated coefficients for *Price* and *Quota* bear expected signs. They suggest that an average customer is willing to pay about \$132 for every 1GB increase in the data plan over the entire contract period, or \$5.5 more per month. The estimated constant for *Upgraded* plan measures the unconditional utility of the plan is -5.75, while that for *Base* plan is set at 0. The results should not be surprising since very few consumers select the *Upgraded* plan. *Contract Type* (which is a binary indicator and is 1 for a renewal contract) is an attribute of a consumer and does not vary over the alternatives. It enters the model when we assume that the effect of the variable differs across alternatives. Table 5 shows that compared to the *Base* plan, the differential impact of being a renewal contract on the utility of *Upgraded* plan to is 0.527. In other words, being a renewal contract yields higher utility of the *Upgraded* plan than the *Base* plan.

With the estimated coefficients, we can derive the valuations for different data plans with Equation 11. For new contracts (as opposed to renewals), when *Quota* is set at the median values of the respective data plans, the valuations for *Upgraded* plan is higher than that of *Base* plan. However, it is worth mentioning that that the average valuation for *Base* plan is 6.2% *above* the median contract cost, whereas the valuation for *Upgrade* plan is 3.2% *below* the cost. In other words, on average, an average consumer derives positive surplus from the *Base* plan and negative surplus from the *Upgraded* plan.

5.3. Inter-Category Correlation Estimation

With the discrete choice models fitted for the phone and the data plan categories respectively, we derive the valuations for every alternative by appropriately scaling the utility which excludes the price impact. With 11 phones in the phone category and 2 data plans, there are 22 unique bundles. For every bundle, we estimate the pairwise correlation coefficient between the particular phone and the specific data plan using the procedure described in Section 4.2. The results are shown in Table 6.

Table 6 Correlation between phone model and data plan

Phone Model	*Upgraded Plan*	*Base Plan*
iPhone New	0.751	−0.804
iPhone Legacy 1	0.569	−0.517
iPhone Legacy 2	0.414	−0.188
Samsung Note	0.680	−0.372
Samsung S New	0.488	0.203
Samsung S Legacy	0.191	−0.454
Samsung Other	−0.279	−0.065
HTC	0.834	−0.710
RIM	−0.096	0.103
Sony	0.325	−0.174
Others	−0.107	0.203

The results show that high correlations exist between phones and data plans. Except *Samsung Other, RIM* and a composite alternative labeled as *Others* (which includes phones that do not belong to the major brands that are explicitly listed in the model), the valuations of phones are positively correlated with the valuations of *Upgraded* plan and are negatively correlated with *Base* plan.

For Apple's iPhones, we observe that for both plans, the magnitude of correlation decreases as the age of the phone increases, i.e., more positively correlated with *Upgraded* plan and less negatively correlated with *Base* with age. One explanation is that a consumer who values the latest phone model which is equipped with more functionalities than its predecessors is likely to value a higher data plan that allows him to take advantage of those new features.

Comparing the two main brands, Apple's iPhones on average have a higher correlation with *Upgraded* plan than the Samsung phones. The observation seems to agree with empirical evidences (Forbes 2014, BusinessInsider 2014) which suggest that an average iOS user consumes more data than his Android counterpart. One could argue that Samsung has introduced many models that cover a wide price range. Models

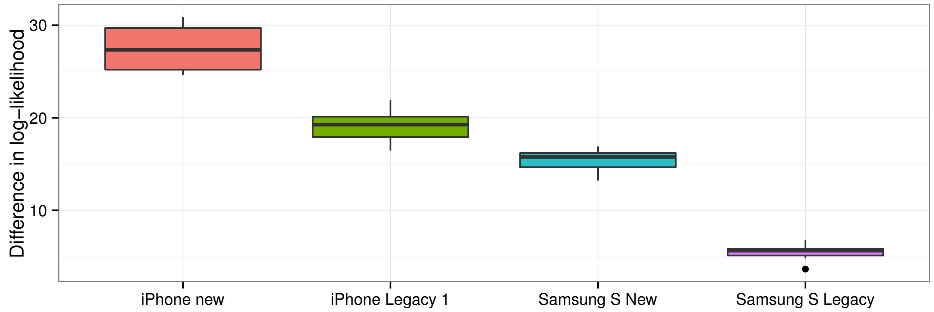


Figure 6 Copula predictive log-likelihood minus the independence model log-likelihood, across 10 folds of cross-validation for each of the four bundles (i.e., between a phone and an *Upgraded* data plan).

which are cheaper and have mediocre hardware capabilities (e.g., Samsung Galaxy S Mini Series, aggregated under *Samsung Other*, is a product line with inferior specifications to its flagship S series) might provide a sub-par user experience that could also affect user engagement. Moreover, consumers who purchase such a device might just be attracted by the low price point, even though they have little interest in more than voice and text.

It is interesting to note that among all the alternatives in the phone category, HTC phones have the highest correlation with *Upgraded* plan and the one of the most negative correlation with *Base* plan. In our data, HTC phones make up merely about 2% of the market share in terms of sales volume. Its best selling models are considered as the top-range Android devices with more superior features than most competing phones in the market. Our results suggest that its users tend to be tech-savvy and demand more data.

To evaluate the predictive performance of the copula model, we use 10-fold cross validation, by fitting the model to 9 folds of the data and then evaluating the (predictive) log-likelihood on the remaining fold. This is done separately for each bundle and the results are compared to the model using the same fitted margins but assuming independence. Figure 6 shows that the result for four popular bundles, which are made up different phone models and *Upgraded* plan. The results show that the copula model has higher predictive likelihoods than the corresponding independence model.

We like to point out that since the analysis is performed on sales data where every consumer purchased "unlimited talk and text" as part of the contract, all the results we have derived (e.g, purchase probabilities of phones and data plans, correlations between the valuations) are *conditioned* on this fact. It is very plausible that the magnitude of the *unconditional* correlation between phones and data plans is much smaller than what we have estimated in Table 6.

5.4. Counterfactual Experiments

A key advantage of structural modeling is that it allows for policy evaluation. To generate insights for bundle pricing, we conduct the following counterfactual experiments. Specifically, the experiments demonstrate

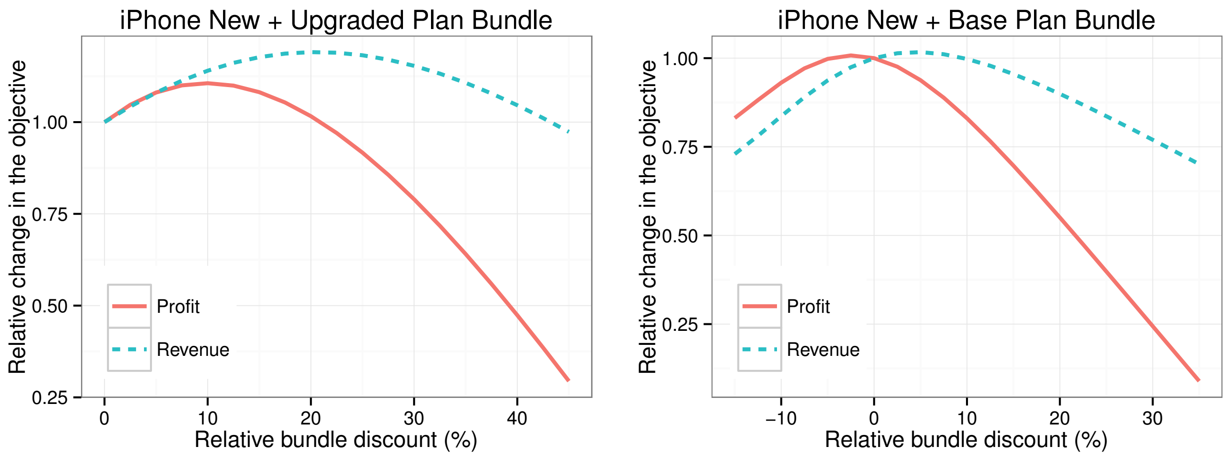


Figure 7 **Relative change in the profit and the revenue for two popular bundles (i.e., latest iPhones with *Upgraded* data plan and *Base* plan respectively) as a function of the relative bundle discount (normalizing with respect to the current bundle discounts).**

the pricing strategies that focus on individual bundles (i.e., ignoring inter-category substitution) as well as collective profit generated by competing bundles.

5.4.1. Individual Bundle Optimization In this experiment, we focus on the some popular bundles and determine the optimal bundle price that maximizes the profit or revenue of an individual bundle. We emphasize that objective function (Proposition 1) includes sales from the bundle as well as the individual products. For example, for the bundle with *iPhone New* and *Base* plan, we are maximizing the expected profit (revenue) generated by this particular bundle, as well as sales from *iPhone New* and *Base* plan. If a customer purchased a different bundle such as *iPhone New* with *Upgraded* plan, the profit (revenue) from the purchase *iPhone New* is captured in the objective function.

We consider the most popular phone in the dataset, *iPhone New*, and study two bundles constructed with the available data plans. We report the experiment results in Figure 7. To protect the client's confidentiality, we normalize the bundle discount with respect to what is currently offered on the particular bundle. Thus, a positive relative bundle discount imply further discount on a bundle, while a negative value suggests to reduce the current discount. We also compute the relative profit (revenue) change by normalizing it with the expected profit (revenue) at the current discount level. Thus, when the relative bundle discount is 0, the relative change is always 1 in the figures.

From the profit-maximizing perspective, Figure 7 implies that there is a lot of room to grow for *Upgraded* plan bundle, whereas the current discount for *Base* plan is "almost" optimal. To be precise, with *Upgraded* plan, an additional 10% discount will result in a 10.6% increase in the expected profit. On the other hand, for *Base* plan bundle, the optimal strategy requires Provider X to impose a markup, i.e., reduce the current discount by 2.5%, which yields an increase of 0.7% in the expected profit. One explanation is that we have shown in Section 5.2 that, the average surplus on *Upgraded* plan is negative at the current price. Thus, very few consumers choose *Upgraded* plan. In order to increase the profit of this bundle, a higher discount which increases the surplus must be given to encourage more consumers to upgrade from the basic plan.

Figure 7 also shows the revenue change with respect to bundle discount, which depicts very different operational recommendations from the profit-driven objective. For example, for *Base* plan bundle, the model recommends a 2.5% *markup* on the existing discount to increase profit, whereas a 7.5% *markdown* to maximize revenue. The reason lies with the asymmetric cost of the bundle components. To calculate the profit margin of phones, we use the information on wholesale phone prices given by Provider X. As we do not have any cost information on servicing the plans, we set it to zero. In other words, entire revenue obtained from the data plan is counted towards the profit. While such accounting is not precise, it is well-known that telecommunication carriers make the bulk of their profits from selling service contracts to consumers. Comparing the profit function to the revenue function in Figure 7, we see that the former has a much steeper slope with respect to discounts. It suggests that a small change in the discount can lead to a much substantial impact to profit, highlighting the importance of accurately determining bundle pricing strategies.

5.4.2. Optimization with intra-category substituion In this experiment, we consider the problem of optimizing aggregate profit generated by several bundles. Because of the substitution effect within each category and the dependence across categories, changing a particular bundle price affects the sales of all other bundles as well as the sales of the individual products.

We consider the best selling phones from the two major brands, namely, *iPhone New*, *iPhone Legacy 1*, *Samsung S New* and *Samsung S Legacy*. Given two different data plans, there are eight unique bundles in the experiment. At the current prices, the products studied here make up over 70% of the total sales. Thus, understanding the price response of this set of products is crucial to the bottom line of the business. For simplicity, we only vary the bundle discount of a particular bundle, i.e., *iPhone new* and the *Upgraded* plan and keep the prices of remaining products fixed. In other words, we are optimizing the particular bundle price with the goal to maximize the aggregate profits generated by all the products. This is a different setting from the experiment in Figure 7, where we only maximize the profit generated by that particular bundle and its components.

To perform the optimization, we use Monto-Carlo simulation procedure highlighted in Section 4.3. We generate 10,000 samples, with each sample corresponds to a consumer's individual valuations for six products (i.e., 4 phones and 2 data plans). The objective function is to maximize the average empirical profit generated from these customers. The result is shown in Figure 8. While all the prices for other products are fixed as they were in the data set, the model suggests that the optimal strategy is to increase the existing discount on *iPhone new* and *Upgraded* plan by 7.5%, which leads to an increase of 3.3% in the expected aggregate profit derived from all the products. Compared to Figure 7 where we only maximize the profit

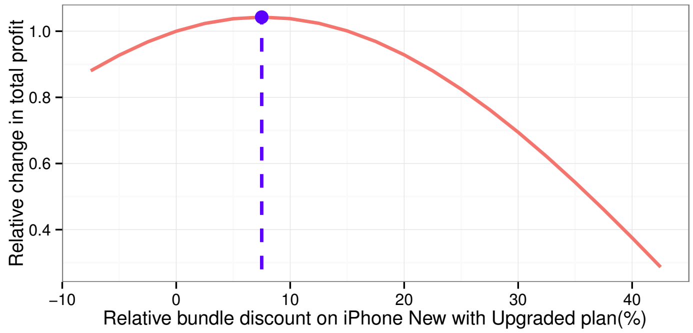


Figure 8 Relative profit change of eight bundles and the individual components with respect to the bundle discount on *iPhone new* and the *Upgraded* plan.

from one bundle (and its individual components), we see that the strategy recommended by this model is less aggressive (7.5% versus 10%) and the predicted profit uplift (3.3% versus 10.6%) is more modest.

The main intuition is that while increasing bundle discount increases sales of this particular bundle, it also makes it less profitable. With the presence of other closely substitutable products, even if a customer does not purchase this particular bundle, she might still purchase other products. Therefore, it is undesirable to apply steep discounts on this bundle while it is possible to extract higher profit from other alternatives. Moreover, because of the availability of other alternatives, a new price strategy on a particular bundle only affects the decisions of a small subset of customers. The subsequent change to the total profit is smaller than the experiment without intra-category substitution shown in Figure 7. We like to point out in the current setup, the only variable is the discount of one bundle. Thus, the result indicates the incremental profit uplift by fine-tuning one lever. The profit improvement is expected to be even larger when one has the full freedom to optimize the prices of all the bundles and the individual products.

6. Discussion and Conclusions

Our work provides foundational results for inferring consumer valuations from data that data that are available to most retailers: sales transaction data. The ability to predict the effect of introducing a bundle at a particular price using only historical sales data is a major advancement in data-driven pricing, and the copula model at the core of the inference here is flexible enough to be useful in real applications. Because the copula allows for arbitrary margins, if a retailer has already developed demand models for a particular item, the demand model can be used directly to obtain the marginal valuation distribution. The likelihood formulas that we derived in this paper provide a theoretically and computationally sound framework for copula learning over latent valuations.

References

Adams, William James, Janet L. Yellen. 1976. Commodity bundling and the burden of monopoly. *The Quarterly Journal of Economics* **90**(3) 475–498.

Allenby, Greg M, Peter E Rossi. 1991. There is no aggregation bias: Why macro logit models work. *Journal of Business & Economic Statistics* **9**(1) 1–14.

Borsdorf, Rüdiger, Nicholas J Higham, Marcos Raydan. 2010. Computing a nearest correlation matrix with factor structure. *SIAM Journal on Matrix Analysis and Applications* **31**(5) 2603–2622.

BusinessInsider. 2014. Google's dirty secret: Android phones are basically used as dumbphones. URL http://www.businessinsider.com/apple-android-market-share-ecommerce-2014-1/.

Eckalbar, John C. 2010. Closed-form solutions to bundling problems. *Journal of Economics and Management Strategy* **19**(2) 513–544.

Elidan, Gal. 2013. Copulas in machine learning. *Copulae in Mathematical and Quantitative Finance*. Lecture Notes in Statistics, 39–60.

Eppen, Gary D., Ward A. Hanson, Martin R. Kipp. 1991. Bundling - new products, new markets, low risk. *Sloan Management Review* **32**(4) 7–14.

Erdem, Tülin, Michael P Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science* **15**(1) 1–20.

Fader, Peter S, James M Lattin, John DC Little. 1992. Estimating nonlinear parameters in the multinomial logit model. *Marketing Science* **11**(4) 372–385.

Forbes. 2014. ios users seven times more active than android users, suggests net applications. URL http://www.forbes.com/sites/ewanspence/2014/08/03/ios-users-seven-times-more\-active-than-android-users-suggests-net-applications/.

Goldberg, Stephen M., Paul E. Green, Yoram Wind. 1984. Conjoint analysis of price premiums for hotel amenities. *Journal of Business* **57** S111–S132.

Hanson, Ward, R. Kipp Martin. 1990. Optimal bundle pricing. *Management Science* **36**(2) 155–174.

Higham, Nicholas J. 2002. Computing the nearest correlation matrixa problem from finance. *IMA journal of Numerical Analysis* **22**(3) 329–343.

Jedidi, Kamel, Sharan Jagpal, Puneet Manchanda. 2003. Measuring heterogeneous reservation prices for product bundles. *Marketing Science* **22**(1) 107–130.

Joe, Harry, James J. Xu. 1996. The estimation method of inference functions for margins for multivariate models. Tech. Rep. 166, Department of Statistics, University of British Columbia.

Kalish, Shlomo, Paul Nelson. 1991. A comparison of ranking, rating and reservation price measurement in conjoint analysis. *Marketing Letters* **2**(4) 327–335.

Kohli, Rajeev, Vijay Mahajan. 1991. A reservation-price model for optimal pricing of multiattribute products in conjoint analysis. *Journal of Marketing Research* 347–354.

McAfee, Preston R., John McMillan, Michael D. Whinston. 1989. Multiproduct monopoly, commodity bundling, and correlation of values. *The Quarterly Journal of Economics* **104**(2) 371–383.

McCardle, Kevin F., Kumar Rajaram, Christopher S. Tang. 2007. Bundling retail products: models and analysis. *European Journal of Operational Research* **177** 1197–1217.

McFadden, Daniel, Kenneth Train. 2000. Mixed mnl models for discrete response. *Journal of applied Econometrics* **15**(5) 447–470.

Ratchford, Brian T. 1979. Operationalizing economic models of demand for product characteristics. *Journal of Consumer Research* 76–85.

Schmalensee, Richard. 1982. Commodity bundling by single-product monopolies. *Journal of Law and Economics* **25**(1) 67–71.

Schmalensee, Richard. 1984. Gaussian demand and commodity bundling. *Journal of Business* **57**(1) S211–S230.

Siddarth, S, Randolph E Bucklin, Donald G Morrison. 1995. Making the cut: Modeling and analyzing choice set restriction in scanner panel data. *Journal of Marketing Research* 255–266.

Sklar, Abe. 1973. Random variables, joint distributions, and copulas. *Kybernetica* **9**(6) 449–460.

Srinivasan, Vinkat. 1982. Comments on the role of price in individual utility judgments. .

Stremersch, Stefan, Gerard J. Tellis. 2002. Strategic bundling of products and prices: a new synthesis for marketing. *Journal of Marketing* **66**(1) 55–72.

Trivedi, Pravin K., David M. Zimmer. 2005. Copula modeling: an introduction for practitioners. *Foundations and Trends in Econometrics* **1**(1) 1–111.

Venkatesh, R., Wagner Kamakura. 2003. Optimal bundling and pricing under a monopoly: constrasting complements and substitutes from independently valued products. *Journal of Business* **76**(2).

Venkatesh, R., Vijay Mahajan. 1993. A probabilistic approach to pricing a bundle of products or services. *Journal of Marketing Research* **30** 494–508.

Wuebeker, Georg, Vijay Mahajan. 1999. A conjoint analysis-based procedure to measure reservation price and to optimally price product bundles. Ralph Fuerderer, Andreas Hermann, Georg Wuebeker, eds., *Optimal Bundling: Marketing Strategies for Improving Economic Performance*. Springer-Verlag, 157–174.

Xu, James J. 1996. Statistical modelling and inference for multivariate and longitudinal discrete response. Ph.D. thesis, Department of Statistics, University of British Columbia.

Appendix A: Proof of Proposition 1

The profit can be decomposed into that obtained from each of the purchase options.

$$\begin{aligned}\mathbb{E}\,[\text{profit}] &= (x_1 - c_1)\mathbb{P}(\text{Purchase item 1 only}) + (x_2 - c_2)\mathbb{P}(\text{Purchase item 2 only}) \\ &\quad + (x_B - c_1 - c_2)\mathbb{P}(\text{Purchase the bundle}).\end{aligned}$$

The options no purchase, purchasing item 1 only, purchasing item 2 only, and purchasing the bundle give the consumer surplus 0, $v_1 - x_1$, $v_2 - x_2$, and $v_1 + v_2 - x_B$ respectively. Let us consider the consumers that purchase only item 1. By the rationality assumption, $v_1 - x_1 \geq 0$, $v_1 - x_1 \geq v_2 - x_2$, and $v_1 - x_1 \geq v_1 + v_2 - x_B$. Thus,

$$\mathbb{P}(\text{Purchase item 1 only}) = \mathbb{P}\left(\{v_1 \geq x_1\} \cap \{v_2 \leq x_B - x_1\}\right) = F_2(x_B - x_1) - F(x_1, x_B - x_1),$$

by Lemma 1. A similar derivation applies to item 2. For the bundle,

$$
\begin{aligned}
&\mathbb{P}(\text{Purchase the bundle}) = \mathbb{P}\left(\{v_1 \geq x_B - x_2\} \cap \{v_2 \geq x_B - x_1\} \cap \{v_1 + v_2 \geq x_B\}\right) \\
=&\mathbb{P}\left(\{v_1 \geq x_B - x_2\} \cap \{v_2 \geq x_B - x_1\}\right) - \mathbb{P}\left(\{v_1 \geq x_B - x_2\} \cap \{v_2 \geq x_B - x_1\} \cap \{v_1 + v_2 \leq x_B\}\right) \\
=&1 - F_1(x_B - x_2) - F_2(x_B - x_1) + F(x_B - x_2, x_B - x_1) - \int_{x_B - x_2}^{x_1} \int_{x_B - x_1}^{x_B - v_1} f(v_1, v_2) dv_2 dv_1,
\end{aligned}
$$

using Lemma 1. Q.E.D.