Latent Variable Copula Inference for Bundle Pricing from Retail Transaction Data

Benjamin Letham   BLETHAM@MIT.EDU
Operations Research Center, Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, MA 02139 USA

Wei Sun   SUNW@US.IBM.COM
Anshul Sheopuri   SHEOPURI@US.IBM.COM
IBM Thomas J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598 USA

Abstract

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. Simulations and data experiments demonstrate consistency, scalability, and the importance of incorporating correlations in the joint distribution.

1. Introduction

Item bundles, when a collection of items are sold together at a discount, are used across many industries, especially in retail. Both theoretical and empirical work has shown that introducing an appropriately priced bundle can significantly increase profits, with low risk to the retailer (Eppen et al., 1991). Even if a bundle has not been previously offered, useful information about how to price the bundle can be obtained from the sales history of the individual items included in the bundle. Choosing the optimal bundle price relies critically on a knowledge of the price consumers are willing to pay for each item in the bundle, called their valuations, as well as the interplay between the valuations of items in the bundle. A retailer generally does not know the full, joint distribution of valuations. However, the retailer likely does have historical sales transaction data for the in-

dividual items. We introduce a procedure for learning the joint distribution of valuations from individual item sales transaction data, thus allowing for optimal bundle pricing.

The economics literature on bundling has extensively examined the economic efficiency of bundling and how bundling can be used for price discrimination (Adams & 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 & 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 & Martin, 1990).

A number of useful insights can be gained from these simplified models (see, for example, Stremersch & Tellis, 2002). However, our main interest is in learning the consumer response to bundling from data. 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 & 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 & Mahajan, 1999).

Our contribution is an inference procedure for predicting the expected change in profits when a bundle is offered at a particular price. The procedure is developed for sales transaction data, and does not require collecting sales data for the bundle *a priori*, nor does it require direct elicitation of valuations via questionnaires. The procedure is based on inference of a copula model over latent consumer valuations, which allows for arbitrary marginal distributions and does not assume independence. Because the valuations are unobserved, the likelihood function involves integrating over the latent valuations, and standard formulas for copula fitting cannot be directly applied. We show how these computationally intractable integrals can be transformed into distribution function evaluations, thus allowing for efficient estimation. Our simulation studies and data experiments suggest that the inference procedure allows for data-based bundling decisions which can help retailers increase profits.

2. Copula Inference and Bundle Pricing

We suppose that a collection of n items have been selected as a candidate bundle, and our goal is to determine the optimal price and its associated profit if the bundle were to be introduced[1]. We consider the situation where the items have not previously been offered as a bundle, but historical sales transaction data are available for the individual items.

The transaction data that we consider consist of two components: purchase data $\boldsymbol{y}^t$ and price data $\boldsymbol{x}^t$. Specifically, we let $\boldsymbol{y}^t = [y_1^t, \ldots, y_n^t]$ denote the sales data for transaction t, with $y_i^t = 1$ if item i was purchased in transaction t, and 0 otherwise. We assume that the price of each item at the time of each transaction is known, and denote the price of item i at the time of transaction t as x_i^t. Let T denote the total number of transactions.

2.1. Valuations and Consumer Rationality

We suppose that each consumer has a valuation for each item, with v_i^t representing the (unobserved) valuation for item i by the consumer in transaction t. As is done throughout the bundling literature and much of the economics literature, we assume that consumers are rational. Specifically, we model consumers as having infinite budget, and as purchasing the assortment of items that maximizes the total difference between their valuation and the price:

$$\boldsymbol{y}^t \in \operatorname*{argmax}_{\boldsymbol{y} \in \{0,1\}^n} \sum_{i=1}^{n} (v_i^t - x_i^t) y_i. \qquad (1)$$

---

[1]The type of bundle that we consider here is called *mixed bundling*, in which consumers are offered both the bundle and the individual items, with the bundle discounted relative to the sum of the item prices.

The rationality assumption implies that $y_i^t = 1$ if and only if $v_i^t > x_i^t$.[2]

The rationality assumption provides a model for the relationship between valuations v_i^t and transaction data y_i^t and x_i^t. Using this model, we now derive likelihood formulas for inferring a joint distribution of valuations from sales transaction data. Then we show how the valuation distribution can be used to find the optimal bundle price.

2.2. Joint Distribution Models and Copula Inference

The most straightforward approach to model a joint distribution is to assume independence. This type of joint model allows for arbitrary margins, however independence is a potentially unreasonable assumption, especially because correlations are quite important for bundling, as we show in Section 3. Modeling the joint distribution as a multivariate normal allows for correlations via a covariance matrix, however it requires the margins to be normally distributed, which can also be a strong assumption when learning from data. Here we model the joint distribution using a copula model, which is a class of joint distributions that allows for 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, 2010; 2013).

We assume consumers are homogeneous, and model the consumer valuations $\boldsymbol{v}^t$ as independent draws from a joint distribution with distribution function $F(v_1, \ldots, v_n)$. Our goal is to infer this joint distribution. Let $F_i(v_i)$ be the marginal distribution function for item i. Then, a copula $\mathbb{C}(\cdot)$ for $F(\cdot)$ is a distribution function over $[0,1]^n$ with uniform margins such that

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

The copula combines the margins in such a way as to return the joint distribution. A copula allows for the correlation structure to be modeled separately from the marginal distributions, in a specific way which we show below. 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 Trivedi & Zimmer (2005) for a more detailed exposition.

Our approach to estimating $F(\cdot)$ will be to choose parametric forms for the margins $F_i(\cdot)$ and the copula $\mathbb{C}(\cdot)$, and then find the parameters for which $\mathbb{C}(F_1(v_1), \ldots, F_n(v_n))$

---

[2]We model v_i^t as a continuous random variable, and thus do not need to devote attention to the case $v_i^t = x_i^t$.

is closest to $F(v_1, \ldots, v_n)$, 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 $F_i(v_i; \boldsymbol{\theta}_i)$ and the parameterized joint distribution as $F(\boldsymbol{v}; \boldsymbol{\theta}, \boldsymbol{\phi}) = \mathbb{C}(F_1(v_1; \boldsymbol{\theta}_1), \ldots, F_n(v_n; \boldsymbol{\theta}_n); \boldsymbol{\phi})$. We are interested in the maximum likelihood problem

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

where $\ell(\boldsymbol{\theta}, \boldsymbol{\phi})$ is the appropriate log-likelihood function. The main advantage in using a copula model is that the parameters can be separated into those that are specific to one margin ($\boldsymbol{\theta}_i$) and those that are common to all margins ($\boldsymbol{\phi}$). Using a procedure called inference functions for margins (IFM) (Joe & 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:

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

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

This gives computational tractability by significantly reducing the dimensionality of the optimization problem that must be solved. In general, IFM does not yield exactly the maximum likelihood estimate: $(\hat{\boldsymbol{\theta}}_{\text{ML}}, \hat{\boldsymbol{\phi}}_{\text{ML}}) \neq (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$. However, the IFM estimates $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$, like the maximum likelihood estimates, are statistically consistent and asymptotically normal (Joe & Xu, 1996; Xu, 1996).

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. In the next two sections, we use the rationality assumption of (1) to derive tractable likelihood formulas to be used in (2) and (3).

2.3. Margin Likelihood and Demand Models

We first consider the margin maximum likelihood problem in (2). Let $p_i(x_i^t)$ be the probability of purchase for item i at price x_i^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 1. *The demand function and the inverse marginal valuation distribution function are identical, i.e.,*

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

Proof. By the rationality assumption of (1), item i is purchased if and only if $v_i^t > x_i^t$:

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

$\square$

We thus choose the following likelihood model for the observed purchase data:

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

Given data $\{x_i^t, y_i^t\}_{t=1}^T$, the log-likelihood function for each margin is:

$$\ell_i(\boldsymbol{\theta}_i) = \sum_{t=1}^T \left(y_i^t \log(1 - F_i(x_i^t; \boldsymbol{\theta}_i)) \right.$$
$$\left. + (1 - y_i^t) \log(F_i(x_i^t; \boldsymbol{\theta}_i)) \right). \tag{4}$$

If $F_i(\cdot; \boldsymbol{\theta}_i)$ is linear in $\boldsymbol{\theta}_i$, for example when using a linear demand model, then the maximum likelihood problem is a concave maximization. For general demand models, a local maximum can easily be found using standard optimization techniques. In Section 2.7 we discuss some possible choices for the family of $F_i(\cdot; \boldsymbol{\theta}_i)$.

2.4. Copula Inference over Latent Variables

Once the margin parameters $\hat{\boldsymbol{\theta}}_i$ have been estimated by maximizing (4), 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 $\boldsymbol{\phi}$.

$$\ell(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) = \sum_{t=1}^T \log p(\boldsymbol{y}^t | \boldsymbol{x}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi})$$
$$= \sum_{t=1}^T \log \int p(\boldsymbol{y}^t | \boldsymbol{v}^t, \boldsymbol{x}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) p(\boldsymbol{v}^t | \boldsymbol{x}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) d\boldsymbol{v}^t. \tag{5}$$

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 $y_i^t = 0$. We then define the lower and upper limits of integration as,

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

The quantity $p(\boldsymbol{v}^t | \boldsymbol{x}^t, \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) = p(\boldsymbol{v}^t | \hat{\boldsymbol{\theta}}, \boldsymbol{\phi})$ is exactly the copula density function, which we denote as $f(\cdot; \hat{\boldsymbol{\theta}}, \boldsymbol{\phi})$. Continuing the likelihood expression from (5), we have,

$$\ell(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi})$$
$$= \sum_{t=1}^T \log \int_{v_n^{t,\ell}}^{v_n^{t,u}} \cdots \int_{v_1^{t,\ell}}^{v_1^{t,u}} f(v_1^t, \ldots, v_n^t; \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) dv_1^t \ldots dv_n^t. \tag{6}$$

The integral in (6) renders the likelihood formula intractable. To allow for efficient inference, we will use the

following formula for a rectangular integral of a probability density function. This formula is critical to the scalability of our inference procedure as it allows us to replace the multidimensional integral in (6) with distribution function evaluations.

Lemma 1. *Let $f(\cdot)$ be a joint probability density function over continuous random variables $z_1, \ldots, z_n$ with the corresponding joint distribution function $F(\cdot)$. Then,*

$$\int_{z_n^\ell}^{z_n^u} \cdots \int_{z_1^\ell}^{z_1^u} f(z_1, \ldots, z_n) dz_1 \ldots dz_n$$
$$= \sum_{k=0}^{n} (-1)^k \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} F(\tilde{\boldsymbol{z}}(I)),$$

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

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

$$\int_{z_n^\ell}^{z_n^u} \cdots \int_{z_1^\ell}^{z_1^u} f(z_1, \ldots, z_n) dz_1 \ldots dz_n$$
$$= \mathbb{P}\left(B \cap (\cap_{i=1}^{n} A_i^{\mathsf{c}})\right)$$
$$= \mathbb{P}\left(B \cap (\cup_{i=1}^{n} A_i)^{\mathsf{c}}\right)$$
$$= \mathbb{P}(B) - \mathbb{P}\left(B \cap (\cup_{i=1}^{n} A_i)\right)$$
$$= \mathbb{P}(B) - \mathbb{P}\left(\cup_{i=1}^{n} (B \cap A_i)\right)$$
$$= \mathbb{P}(B) - \sum_{k=1}^{n} (-1)^{k-1} \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} \mathbb{P}(B \cap A_I)$$

by the inclusion-exclusion formula, with $A_I = \cap_{i \in I} A_i$. Substituting $\mathbb{P}(B) = F(z_1^u, \ldots, z_n^u)$ and $\mathbb{P}(B \cap A_I) = F(\tilde{\boldsymbol{z}}(I))$ 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 (6):

$$\ell(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}) = \sum_{t=1}^{T} \log \sum_{k=0}^{n} (-1)^k \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} F(\tilde{\boldsymbol{v}}^t(I); \hat{\boldsymbol{\theta}}, \boldsymbol{\phi}),$$
(7)

where as before

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

For the most simple case of two items in a bundle, the inner

expression in (7) evaluates to

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

2.5. Consistency and Scalability

Combining (4) and (7) yields the complete inference procedure, which we give in the following proposition.

Proposition 2. *The inference procedure*

$$\hat{\boldsymbol{\theta}}_i \in \underset{\boldsymbol{\theta}_i}{\arg\max} \sum_{t=1}^{T} \left(y_i^t \log(1 - F_i(x_i^t; \boldsymbol{\theta}_i)) \right.$$
$$\left. + (1 - y_i^t) \log(F_i(x_i^t; \boldsymbol{\theta}_i)) \right)$$
$$\hat{\boldsymbol{\phi}} \in \underset{\boldsymbol{\phi}}{\arg\max} \sum_{t=1}^{T} \log \sum_{k=0}^{n} (-1)^k \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=k}} F(\tilde{\boldsymbol{v}}^t(I); \hat{\boldsymbol{\theta}}, \boldsymbol{\phi})$$

is statistically consistent.

Because the inference is exactly the IFM procedure, it follows from Joe & Xu (1996) that it is statistically consistent.

The computation is exponential in the size of the bundle n, however in retail practice bundle offers generally do not contain a large number of items. Importantly, the computation is linear in the number of transactions T, which allows inference to be performed even on very large transaction databases. The main computational step is evaluating the copula distribution function in (7). For many copula models, such as the Gaussian copula which we describe in Section 2.7, efficient techniques are available for distribution function evaluation.

2.6. Computing the Optimal Bundle Price

Given the joint valuation distribution, the expected profit per consumer as a function of item and bundle prices can be computed. For notational convenience, here we give the result for $n = 2$. Consumers are rational, in that they choose the option (item 1 only, item 2 only, bundle, or no purchase) that maximizes their surplus $v_i - x_i$. For this result, we assume that the valuation for the bundle is the sum of the component valuations $v_B = v_1 + v_2$, although this could easily be relaxed to other bundle valuation models such as those in Venkatesh & Kamakura (2003). Note that inferring the joint valuation distribution does not require any assumption on how valuations combine, rather

this assumption is only used to compute the expected profit of bundling. We denote the cost of item i as c_i and assume that the bundle cost is the sum of the component costs.

Proposition 3. *For joint valuation density function $f(\cdot)$ and joint valuation distribution function $F(\cdot)$, the expected profit per consumer obtained when items 1, 2, and the bundle are priced at x_1, x_2, and x_B respectively is*

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

The proof relies on Lemma 1 and is given in the Appendix. Similar results, albeit notationally complex, can be obtained for $n > 2$. The inference procedure from Proposition 2 is used to estimate the valuation distribution function, which allows the expression in Proposition 3 to be evaluated. Maximizing the expected profit with respect to x_B yields the optimal bundle price, or maximizing over x_B and the item prices simultaneously yields a complete pricing strategy. The formula in Proposition 3 is not concave in general, but a local maximum can be found using standard numerical optimization techniques.

2.7. Distributional Assumptions

The likelihood formulas in (4) and (7) hold for arbitrary margins $F_i(\cdot; \boldsymbol{\theta}_i)$ and an arbitrary copula model $\mathbb{C}(\cdot; \boldsymbol{\phi})$. To apply these formulas to data requires choosing the distributional form of the margins and the copula family.

The connection between marginal valuation distributions and demand models given in Proposition 1 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, two common choices for demand models are the linear demand model and the normal-cdf demand model. The linear demand model is

$$
p(x_i; \beta_i, \eta_i) = \min(1, \max(0, \beta_i - \eta_i x_i)),
$$

and the corresponding valuation distribution is uniform:

$$
v_i \sim \mathrm{Unif}\left(\frac{\beta - 1}{\eta}, \frac{\beta}{\eta}\right).
$$

When the demand model is the normal distribution function

$$
p(x_i; \mu_i, \sigma_i^2) = 1 - \Phi(x_i; \mu_i, \sigma_i^2),
$$

the corresponding marginal valuation distribution is the normal distribution:

$$
v_i \sim \mathcal{N}(\mu_i, \sigma_i^2).
$$

Remark. Additional covariates like competitors' prices or the prices of substitutable and complimentary products are sometimes used in demand modeling, for instance in a choice model. Seasonality effects are also often handled using covariates. Models with covariates can also be transformed into valuation distributions using Proposition 1.

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:

$$
\mathbb{C}(F_1(v_1), \ldots, F_n(v_n); \boldsymbol{\phi}) = \Phi(F_1(v_1), \ldots, F_n(v_n); \boldsymbol{\phi}),
$$

where $\Phi(\cdot; \boldsymbol{\phi})$ represents the multivariate normal distribution function with correlation matrix $\boldsymbol{\phi}$. 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 & Zimmer (2005).

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 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
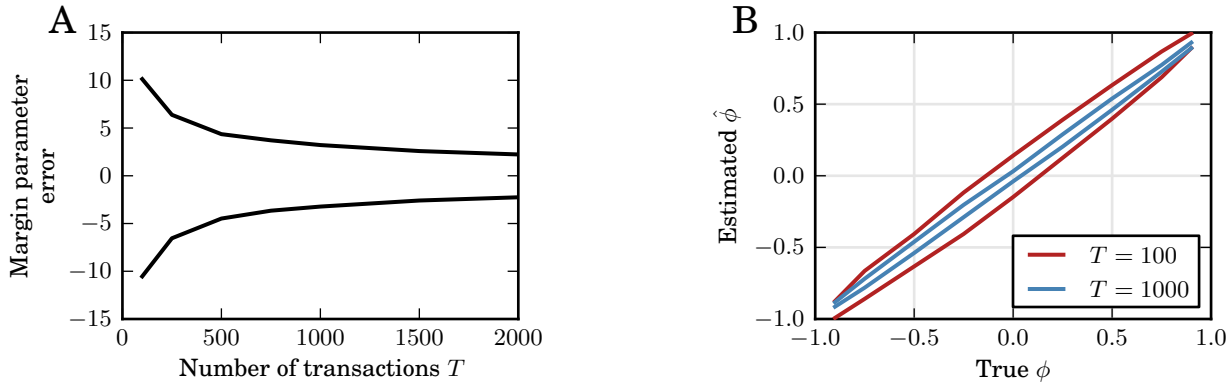
Figure 1. Convergence of both (A) margin parameters and (B) the correlation coefficient to their true values as the number of simulated transactions T is increased. In (A), the lines indicate the first and third quartiles of the margin parameter errors across all simulations with the same number of transactions 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.
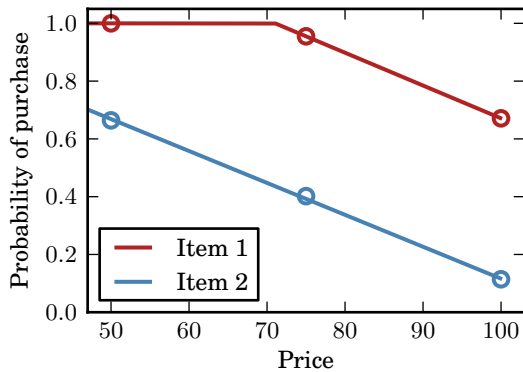

Figure 2. Demand models for each of the two items for one of the simulated datasets. The circles give the empirical purchase probabilities measured from the data, and the lines show the fitted margin distribution function.
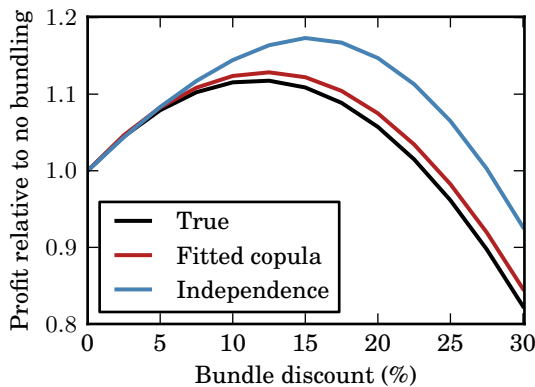
Figure 3. Change in relative profits from introducing the bundle at a particular discount relative to the sum of item prices, as estimated from the true distribution, the fitted copula model, and a distribution using the fit margins but assuming independence.

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 (1) to produce binary purchase data.

We applied the inference procedure in Proposition 2 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 the fitted margins for this particular simulated dataset. The estimated correlation coefficient, found by maximizing (7), was 0.48. To illustrate the potential profitability of bundling, in Figure 3 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 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.
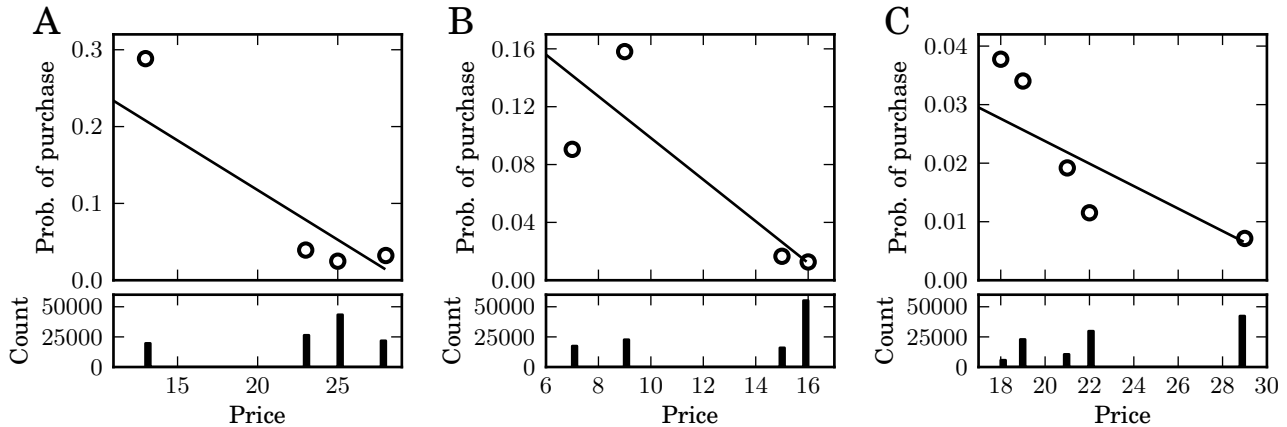
Figure 4. Fitted marginal distributions for items (A) 38, (B) 14, and (C) 08 from the Ta-Feng retail transaction dataset. The underset histogram shows the number of transactions for which the item was offered at each price. For each price at which the item was offered, the circles indicate the purchase probability at that price as measured from the data. The line gives the model fit.

4. Data Experiments

We provide further evaluation and illustration of the inference procedure by applying it to actual retail transaction data. We use the publicly available Ta-Feng dataset, which contains four months of transaction level data from a Taiwanese warehouse club, totaling about 120,000 transactions and 24,000 items (Hsu et al., 2004). After some data pre-processing which is described in the Appendix, we selected the three items with the highest support and considered the four possible bundles that could be obtained from these items (three pairs and one bundle of three). Throughout this section we refer to the three items as item 38, item 14, and item 08 - the full EAN-13 for the items is given in the Appendix. Note that in these experiments the inference procedure scales to a much larger dataset than those used in the simulation studies.

As in Section 3, we model the joint valuation distribution using linear demand models (uniform marginal valuation distributions) and a Gaussian copula. In Figure 4 we show the demand models fit by maximizing (4) for each item. The off-diagonal elements of the correlation matrix ϕ corresponding to pairs 38-14, 38-08, and 14-08 were jointly estimated as 0.085, 0.133, and 0.172 respectively.

We evaluated the predictive performance of the copula model using 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 was done separately for each pair of items (38-14, 38-08, and 14-08) and for the collection of all three items (38-14-08), and the results are compared to the model using the same fitted margins but assuming independence. Figure 5 shows that for all 10 folds and for all bundles, the copula model had higher predictive likelihoods than the corresponding inde-
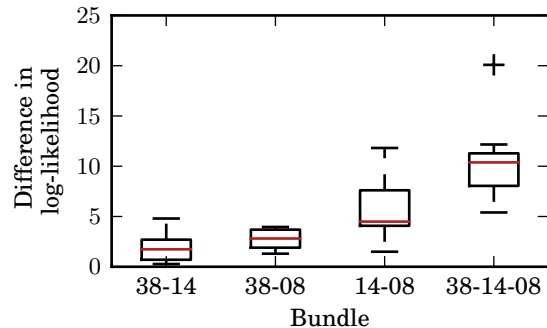

Figure 5. Copula predictive log-likelihood minus the independence model log-likelihood, across 10 folds of cross-validation for each of the four bundles.

pendence model.

To illustrate the results, we report the relative expected profit under various bundle scenarios in Figure 6. For these results we took the item prices as the mode of the price distribution in the data, and since the item costs are unknown, we set them to a 35% markup. In a similar way as Figure 3, Figure 6 shows that introducing a discounted bundle can increase profits, and that assuming independence can lead to very different predictions. This further highlights the importance of including correlations in the valuation distribution model.

5. Discussion and Conclusions

We used copula modeling in the context of an important business analytics problem, and in the process have developed new methodological results on learning a copula distribution over latent variables. Business analytics is a budding application area in machine learning, and our work
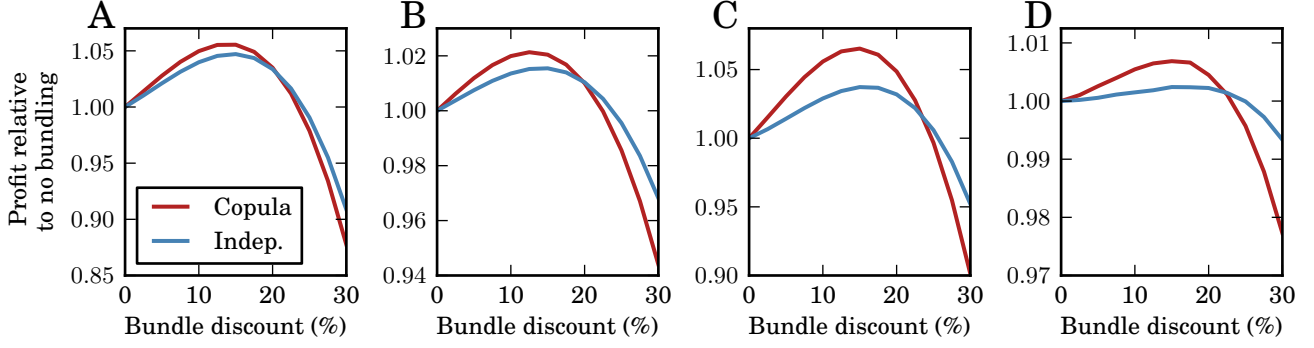
Figure 6. Change in relative profits by introducing bundles (A) 38-14, (B) 38-08, (C) 14-08, and (D) 38-14-08 as a function of the level of bundle discount, estimated from the Ta-Feng dataset. In red is the prediction obtained from the fitted copula model, and in blue is the prediction obtained using the same fitted margins, but assuming independence.

provides foundational results for inferring consumer valuations. 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.

6. Appendix

Here we give the proof of Proposition 3, and describe the data pre-processing done with the Ta-Feng dataset.

6.1. Proof of Proposition 3

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

$$\mathbb{E}\left[\text{profit}\right] = (x_1 - c_1)\mathbb{P}(\text{Purchase item 1 only})$$
$$+ (x_2 - c_2)\mathbb{P}(\text{Purchase item 2 only})$$
$$+ (x_B - c_1 - c_2)\mathbb{P}(\text{Purchase the bundle}).$$

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,

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

using Lemma 1. $\qquad\square$

6.2. Data Pre-processing

Each entry in the Ta-Feng dataset corresponds to the sale of a single item within a transaction. To form the complete transaction of (potentially) multiple items, we grouped all sales that occurred on the same day with the same user ID. For simplicity, we assumed that for each day there was a single price for each item. If there were multiple prices at which an item was sold on a given day, we took that day's price as the median of the observed prices. If an item was not sold on a particular day, then we took that day's price as the price of the preceding day. To further smooth the prices, we allowed only prices that covered at least 5% of transactions, and any price that did not meet that support threshold was rounded to the nearest price that did. After removing items that did not have at least three prices in the data, the three items with the highest support were (EAN-13) 4714981010038, 4711271000014, and 4710583996008. In Section 4 we refer to these items by their last two digits.

References

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

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

Elidan, Gal. Copula bayesian networks. In *Proceedings of NIPS*, 2010.

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

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

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

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

Hsu, Chun-Nan, Chung, Hao-Hsiang, and Huang, Han-Shen. Mining skewed and sparse transaction data for personalized shopping recommendation. *Machine Learning*, 57(1-2):35–59, 2004.

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

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

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

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

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

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

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

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

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

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

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

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

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